Special Section:
Opportunities and Challenges for the Discipline

An Invitation: Editor's Introduction to "Desired and Feared—What Do We Do Now and Over the Next 50 Years?"

John STUFKEN

With enormous amounts of data being collected and complex processes being studied nowadays, opportunities for statisticians are virtually limitless. Most statisticians are involved in some of these opportunities (and, perhaps, occasionally suffer from being pulled into too many), regardless of whether they make a living in academia, industry, or government. This is truly a golden age for statisticians (and for others with quantitative interests and/or skills).

At a time that many statisticians flourish, so should statistics as a discipline. At some level this is true and there are many success stories for statistics. But just as individual statisticians can suffer from being pulled into too many activities, the discipline also seems to face serious challenges that result, among others, from an abundance of opportunities. Examples of challenges for the discipline can be found in the interesting articles "A Report on the Future of Statistics" (Lindsay, Kettenring, and Siegmund 2004, *Statistical Science*, 19, 387–413) and "What Is Statistics?" (Brown and Kass 2009, *The American Statistician*, 63, 105–123).

Challenges range from the discipline losing its cohesiveness, with individuals becoming more and more (narrowly?) specialized in different interdisciplinary areas to a need for more federal dollars to support statistical training and research. Other challenges include finding the best ways to interact with other disciplines through teaching and research, and how to modify and modernize our programs in order to attract larger numbers of talented students to the discipline. *TAS* has recently published a number of different articles on some of these issues, including, but not limited to, the above referenced article by Brown and Kass and the article entitled "Desired and Feared—What Do We Do Now and Over the Next 50 Years" by Meng in this issue.

As Editor, I consider *TAS* to be an ideal forum for discussion of opportunities and challenges to the discipline. Contributions are welcome at any time and can be in the form of a short article or as a "Letter to the Editor" in response to an article. Ideas for special sections on issues of this nature are highly welcome. At this time I would like to invite contributions to a discussion of the article by Meng and will consider the most interesting of these contributions for publication in a special section of a later *TAS* issue. The chance that a contribution is included in this special section is better if I receive it within two months of the publication date of this article, if it offers an alternative perspective or idea or a substantial addition to the article, and if it is written clearly and concisely. Contributions should be submitted just as any other *TAS* article, with a cover letter that refers to the special section on the article by Meng.

—John Stufken, *TAS* Editor

 DOI: 10.1198/tast.2009.633intro

Desired and Feared—What Do We Do Now and Over the Next 50 Years?

Xiao-Li MENG

An intense debate about Harvard University's General Education Curriculum demonstrates that statistics, as a discipline, is now both desired and feared. With this new status comes a set of enormous challenges. We no longer simply enjoy the privilege of playing in or cleaning up everyone's backyard. We are now being invited into everyone's study or living room, and trusted with the task of being their offspring's first quantitative nanny. Are we up to such a nerve-wracking task, given the insignificant size of our profession relative to the sheer number of our hosts and their progeny? Echoing Brown and Kass's "What Is Statistics?" (2009), this article further suggests ways to prepare our profession to meet the ever-increasing demand, in terms of both quantity and quality. Discussed are (1) the need to supplement our graduate curricula with a *professional development curriculum* (*PDC*); (2) the need to develop more *subject oriented statistics* (*SOS*) *courses* and *happy courses* at the undergraduate level; (3) the need to have the most qualified statisticians—in terms of both teaching and research credentials—to teach introductory statistical courses, especially those for other disciplines; (4) the need to deepen our foundation while expanding our horizon in both teaching and research; and (5) the need to greatly increase the general awareness and avoidance of unprincipled data analysis methods, through our practice and teaching, as a way to combat "incentive bias," a main culprit of false discoveries in science, misleading information in media, and misguided policies in society.

KEY WORDS: Communication skills; General education curriculum; Graduate education; Incentive bias; Statistical education; Undergraduate education.

1. WHAT IS STATISTICS—DESIRED OR FEARED?

In the past few years, the Faculty of Arts and Sciences (FAS) at Harvard undertook a heated and intense debate regarding a new General Education (Gen Ed) curriculum. One of the initial categories of Gen Ed was *Empirical Reasoning*, with the following proposed requirement. Courses in this category must:

a. teach how to gather and assess empirical data, weigh evidence, understand estimates of probabilities, draw inferences from the data available, and also recognize when an issue cannot be settled on the basis of the available evidence;
b. teach the conceptual and theoretical tools used in reasoning and problem solving, such as statistics, probability theory, mathematics, logic, and decision theory;
c. provide exercises in which students apply these tools to concrete problems in an area of general interest to undergraduates; and
d. where practicable, familiarize students with some of the mistakes human beings typically make in reasoning and problem-solving.

Pleasantly surprised by this proposal, I wanted to know which of my statistical colleagues were involved in drafting it. So did my colleagues, as they thought that I must have had a hand in this, representing our department. Given the language, particularly (a), it is not illogical to infer a statistician's involvement.

No statisticians, at least by the current definition, were involved. It was written by several social and natural scientists. Naturally, my colleagues and I were delighted, at least until the FAS faculty meeting in which it was voted on. With the support from social and natural scientists, surely it would pass with flying colors, right? Quite the contrary—it was defeated! Our academic relatives in mathematics, applied mathematics, and computer science (CS) strongly rejected it fearing that part (a) would exclude almost all of their courses. Humanists, who often dominate FAS meetings with eloquent speeches and resounding articulations, apparently were in a similar mood, concerned with the dominance of social and natural sciences in Gen Ed, and, therefore, particularly appreciated our relatives' sentiment.

Following the meeting, I was bombarded by E-mails from our relatives accusing the statistics department of self promotion at others' expense. Some phrases were so strong I could only enjoy them with a glass of blended Bordeaux. No kidding about being intoxicated—when was the last time our math or CS relatives felt threatened by their distant cousin?

The moral of this story, of course, is not about rivalries among disciplines; we, the statisticians, were not even aware of the proposal until its formal circulation. But it reminded me of a quote from an ex-colleague at Chicago: "You know you've really made it when others start to fear you." As cynical as this quote sounds, our professional identity, *Statistics*, is crystal clear in this "fear-filled" incident. Some have been concerned with our losing ground to other disciplines, especially to CS and

Xiao-Li Meng is the Whipple V. N. Jones Professor and Chair of Statistics, Harvard University, Cambridge, MA 02138. He can be reached at *chair@stat.harvard.edu* or *meng@stat.harvard.edu*. He thanks his colleagues, particularly Joe Blitzstein, Rick Cleary, Dave Harrington, Sam Kou, and Carl Morris, as well as his Happy Team (Paul Baines, Yves Chretien, Paul Edlefsen, Reetu Kumra, Kari Lock, Wee Lee Loh, Linjuan Qian, and Casey Wolos), for making many "Harvard experiments" a reality, and for boosting his (V). He is working hard on achieving (I)–(IV), knowing it will be a life-long journey. [Riddle: No number theory is needed to decipher these codes.] He also thanks Joe Blitzstein, Rick Cleary, Yves Chretien, Radu Craiu, Paul Edlefsen, Reetu Kumra, Kari Lock, Carl Morris, Dale Rinkel, and Casey Wolos for constructive comments and for proofreading. Finally, he thanks Emery Brown and Robert Kass for their inspiring article, and the Editor John Stufken for the invitation to prepare a discussion, which ultimately led to the current article.

  DOI: 10.1198/tast.2009.09045

Engineering. Researchers in these fields tend to be fearless—to "leap daringly into the fray" as Brown and Kass (2009) put it—in their quest to invent and (unknowingly) reinvent the wheel. However, in this debate, both proponents and opponents understand clearly that courses for the proposed empirical reasoning category would be predominantly *statistical* courses (though not necessarily offered by the statistics department), not (applied) mathematical or CS or engineering courses. Indeed, the proposers' intention, as I learned later, was to exclude courses such as calculus and programming language. Thus, the fear of our math and CS colleagues was actually well-founded, except that they went after the wrong party with their complaints!

Ultimately, a compromise was reached, with part (a) dropped, parts (b)–(d) very slightly modified, and the category renamed as "Empirical and Mathematical Reasoning" (see *http://isites.harvard.edu/icb/icb.do?keyword=k37826&pageid=icb.page163841*). Some social and natural scientists are unhappy, fearing that the original objectives of the requirement are largely lost because a student may take, for example, a course in number theory to satisfy the Gen Ed requirement. Others are more optimistic, reasoning "Well, let such courses in. How many students would opt for a number theory course instead of a course in statistics or in another more applied field?" This free-market spirit seems to prevail, or perhaps I should say that the need for statistics prevails. Statistics, as a discipline, is clearly identified in this debate, whether desired or feared. Brown and Kass's (2009) inspirational article started with the philosophical question "What Is Statistics?" My motivation for writing this follow-up article is to supplement their suggestions and action list to address an urgent practical question: "*What can and should we do now and in the near future, given that we are in the spotlight?*"

2. WHAT SHOULD BE OUR DEEPEST FEAR?

During a recent discussion at a life science department at Harvard, about half of the faculty indicated they want their students to take one course in statistics, and the other half want students to take one course in calculus and then one course in statistics. What is the implication of this? If students can only be required to take *one* course in mathematical sciences it would be a course in statistics. This sentiment is now shared by many of my science colleagues (though I am acutely aware of the selection bias in what I hear, a benefit of being a statistician!). My personal belief, which I surmise many share, is that the minimal training for a modern scientist should include one course in calculus, one course in CS, and one course in statistics. But the very fact that a good number of scientists are now willing to let their students forgo calculus to make room for statistics is something we all should take a deep breath and reflect upon carefully.

We are, of course, excited by this general recognition. With it, however, comes an exceedingly challenging task. Some of us are concerned, or even have a bit of fear ourselves. John Tukey is often quoted as having said that the best thing about being a statistician is that we get to play in everyone's backyard. But we are now being invited into everyone's study or playroom, to perform a vital role in nurturing and educating their offspring. Are we ready for such a sea change? Messing up a backyard can certainly upset the host, but imagine the consequences of messing up someone's progeny? Are we training enough qualified educators to take on this enormous task? Do we have enough qualified trainers to conduct such training? Do we, as a discipline, even have a clear consensus on what constitute *qualifications* for being the first quantitative trainers of future generations of scientists, engineers, policy makers, etc.?

Perhaps injecting a bit more "fear" could help us to see the urgency. A current general misperception can be summarized as, "Statistics is *easy to teach*, but *hard (and boring) to learn*." As we know, many disciplines teach their own statistics courses, some with well-qualified scholars who indeed can better motivate their students than we can. But then there are many more who themselves have fallen victim to inadequate or misguided statistical training, or who have no training at all, but have been asked to teach statistics simply because they had a quantitative degree of some sort or have analyzed some data.

On the other hand, most of us (see Craiu 2009 and Meng 2009, for example) have frequently had the experience of telling someone, "I teach statistics," only to hear, "Oh, that's the hardest course I have ever taken!", or even, "Sorry, but I really hated my stat course." How could that be? How could teaching statistics require little disciplinary training or credentials, which would imply that statistics is an easy subject to pick up, and yet the majority of students find learning statistics difficult and dreadful? What will it be like if this phenomenon continues when many more students are required to take statistics, possibly as their only quantitative training?

This should be our profession's deepest fear: we could screw up big time because it is no longer just about helping others clean up their backyards, but rather about preparing whole generations of future scientists and policy makers. If we do not offer enough good quality courses, others will do whatever they can, and even more so than in the past because of the greatly increased demand. We will then have much to worry about or even to fear, not because statistical methods are being invented or reinvented by nonstatisticians, but because a discipline's identity, and ultimately, the discipline itself, is greatly diluted and devalued when it allows many unqualified people to serve one of its fundamental missions, that is, to educate future generations about the discipline. So again, what can, and should, we do to minimize the chance of this happening?

3. SUPPLEMENTING GRADUATE CURRICULA WITH PROFESSIONAL DEVELOPMENT CURRICULUM (PDC)

Clearly recognizing the shortage of supply, Brown and Kass (2009) suggest changes to current curricula to train more and better statistical players in everyone's backyard or even front yard. The need for greatly expanded undergraduate statistical education demands further improvement to our current graduate curricula: a supplementary *Professional Development Curriculum* (PDC) for training more and better educators and communicators for our discipline. Good communication skills are also essential for interdisciplinary work, especially those large-scale collaborations emphasized by Brown and Kass (2009).

Currently we have far too few good statistical educators and communicators relative to the task at hand and the coming demand. It would take strong collective effort, led by professional societies such as ASA and IMS, to change the general perception (to a certain degree, an earned perception) that statisticians are not effective educators or communicators. Great efforts are being made. For example, the (past) ASA President Tony Lachenbruch chose "Communicating Statistics and Developing Professionals" as his central theme, and appointed a corresponding Task Force, chaired by Karen Kafadar, which has compiled a list of action items, some of which are exactly what the PDC is designed for (see the President's Invited Column, *Amstat News*, August, 2008). Such ongoing and sustainable effort is critical for preventing the type of perception vividly clear in the following anecdote.

I was invited to give a talk to a group of health science and medical researchers last year, and the host tried to impress the attendees by introducing me as "perhaps the best speaker in statistics." This, of course, would offend many statisticians—"What about me?" But hold the complaint until you hear one medical doctor's immediate interruption: "Oh, that's not hard to be at all!"

Since 2005, we have experimented with such a PDC at Harvard and so far the feedback and reaction from students, colleagues elsewhere, and the FAS administration has been overwhelmingly positive (e.g., our department has received multiple awards and increased general attention; see Meng 2008 and Cassidy 2009). Indeed, much of our PDC was requested by our students. This includes *Stat 303*, *The Art and Practice of Teaching Statistics*, a year-long required course for all first year Ph.D. students, or G1s (at Harvard, *n*th year graduate students are known as the Gns), aimed at helping the students develop into better Teaching Fellows and general speakers; and *Stat 399, Problem Solving in Statistics*, designed for students, mostly G2s, who are preparing for their Ph.D. qualifying examinations, which emphasizes deep, broad, and creative statistical thinking instead of technical problems that correspond to an identifiable textbook chapter. All our ladder faculty members have participated in Stat 399, which serves the further purpose of improving student-faculty communication.

We have also just test-ran *Stat 366: Research Cultivation and Culmination Workshop*, focusing on walking through the entire process of developing a research idea into a publication with an emphasis on effective scientific writing and communication, including how to read and respond to referees' comments. This new workshop course is aimed at G3s, who need to prepare for their qualifying papers, biannual postqualifying presentations, and ultimately their Ph.D. theses. The next installment will be a workshop for G4s and beyond on preparing for their job applications, interviews, and first jobs. For departments that are not as interdisciplinary-oriented as ours, a course on statistical consultation should also be considered as a part of a PDC, or of the regular curriculum, as already exists at a good number of universities.

The central mission of the PDC is the development of future statisticians who will need stronger communication skills, both oral and written, and a higher level of versatility in thinking and in connecting the dots, in order to be successful at the forefront of scientific research and education, not just in the "backyard," where most current generations reside. It is this changing of a statistician's role in scientific arenas and societal endeavors that makes the lack of systematic training in this regard another set of "deep deficiencies requiring immediate attention," to echo Brown and Kass (2009).

4. DEVELOPING MORE SUBJECT ORIENTED STATISTICAL (SOS) COURSES

In addition to better training for graduate students, another essential task is to offer as many high quality undergraduate introductory courses as possible; or as Brown and Kass (2009) put it, "the first college-level exposure to statistics matter." Many excellent courses already exist, with tremendous on-going efforts, such as those made by CAUSE (*http://www.causeweb.org/*). But Brown and Kass (2009) call for more courses with somewhat different structures than the current ones. There are two broad types of such courses that I believe we should further develop whenever possible. The first type is primarily for students who have invested in their majors—I label these as *subject oriented statistics* (SOS) courses. The second type is for general audiences, especially those who need to be inspired to sit though a statistics course; for reasons that will be clear in Section 5, I label these as *Happy Courses*.

By "SOS course" I do not mean a traditional introductory course with more examples taken from a specific field, say, economics. What I mean is a *statistical* course that is designed with direct input from experts from a broad field or fields by determining what they want, or more importantly, what they need; "wanting" and "needing" can be quite different when the disciplinary experts themselves do not know enough about modern statistical concepts or thinking to ask for the right methods or even pose the right questions. SOS courses are, however, not compromised in educating students about the unifying theme of statistics as a fundamental discipline in scientific inquiry. Indeed, an SOS course can be more effective in conveying the general statistical principles and statistical thinking precisely because it places them in a context about which the students want to learn, especially when it is taught with tailored delineation.

For instance, economics students studying a time series may need more help to understand where "replications" come from when there is only one long time series, while for psychology students who study experimental design the notion of replication is easier to grasp. As another example, for engineering students designing experiments, we teach them the efficiency-robustness trade-off by studying how to reach a comprise between learning more factors and learning a few well, given a fixed resource. For life science students building Markovian models, the same trade-off may become striking a balance between increasing goodness of fit to the current data versus reducing predictive errors for future outcome.

Undoubtedly, designing and teaching an SOS course requires considerably more effort than just picking up a textbook, say "Introduction to Statistics for Economics," and then lecturing. We will not only need more cross-disciplinary knowledge, but also more insightful understanding of the pedagogical needs of other disciplines. Some general efforts in this direction are underway, as highlighted by making "Statistics in Other Disciplines" a required course in the curriculum of a planned statistical education program (see Garfield et al. 2009).

Locally at Harvard, under the leadership of our CoDirectors of Undergraduate Studies, David Harrington and Joseph Blitzstein, we started joint explorations with economics, psychology, engineering, life sciences, etc. The journey is clearly long and circuitous. However, regardless of how successful our SOS courses will eventually be, the very fact that we took the time to sit down with faculty from other departments has been exceedingly well received. Indeed, given our very limited faculty resources, we initially planned to experiment with such SOS courses only with economics and psychology, two of our long-term "clients." The word, however, is out. I was soon greeted during chairs' meetings by other department chairs saying, "Hey, don't forget us!" or, "You guys really should talk to us too!"

These requests, of course, are not unexpected. But the dialogues also revealed something less anticipated. For example, one department chair told us, "We often look at each other and don't know what to say when a student presents a thesis that uses quite a bit of statistical methods—we just don't know enough to judge whether they are right or not. If you can offer a course for us, I want to sit in myself!" Several other faculty members echoed the same sentiment. Evidently, the task we face is even greater than educating the students from other disciplines. Indeed, the more statistically-oriented they become, the more demand these students will impose on their discipline's professors!

5. DEVELOPING MORE APPETIZING HAPPY COURSES

Equally important, and time consuming, is to design general introductory courses that would truly inspire students to learn—and learn *happily*—statistics as a way of scientific thinking for whatever they do, not a collection of tools that they may or may not need some day. Such courses are particularly effective for students who have not decided on a major, and therefore, are not compelled by the need (and requirement) of any particular discipline to study statistics. Obviously it is among these students where we have the greatest chance of developing future statisticians. Many current introductory level textbooks and courses do make a great effort to attract such students, but as Brown and Kass (2009) noted, "introductory courses too often remain unappetizing."

To make statistics more appetizing, somewhat literally, we last year launched a module-based undergraduate course, *Stat 105: Real-Life Statistics: Your Chance for Happiness (or Misery)*, after two years of preparation by what is now known locally as my *Happy Team*, which has included, over the years, eight Ph.D. and masters students. The central feature of this course is that the materials are organized by real-life topics instead of statistical ones. In the first offering, the five modules were (1) Finance (e.g., stock market), (2) Romance (e.g., on-line dating model), (3) Medical Science (e.g., Viagra trial), (4) Law (e.g., the Sally Clark case), and (5) Wine and Chocolate Tasting (depending on a student's age). The statistical topics are covered whenever they are needed by a module, which means that they may be "out of sequence" or appear multiple times.

Judging from the students' feedback and local media coverage we received (see *AmStat News*, April 2008, or *http://www.news.harvard.edu/gazette/2008/02.14/11-stats.html*), the students responded well to such a "real-life module" approach because it makes statistics a much more "alive" and tangible subject than they previously perceived. To keep up the "aliveness" of the course, this past spring we offered a new module on voting and election (as an alternative to the law module), given the historic election we all just witnessed. The course has been approved to become a Gen Ed course as Harvard launches its Gen Ed Curriculum next year. Eventually we hope to prepare a textbook and web media, with the ultimate goal of encouraging others to develop more such *Happy Courses*, so labeled to emphasize their key goal—to make students happy to learn statistics. (A brief summary of *Stat 105* can be found in a CAUSE webinar *http://www.causeweb.org/webinar/2008-11/*.) Of course, a happy course can focus on one real-life subject, instead of multiple ones, such as the course on sports and statistics planned by my colleague Carl Morris.

Incidentally, the direct involvement of graduate students (i.e., the *Happy Team*) in designing an undergraduate course itself serves as a great training opportunity, a model now formally instituted at Harvard as a Graduate Seminar in General Education; see the list of seminars at *http://www.gsas.harvard.edu/news_and_events/graduate_seminars_in_general_education.php*. Interestingly, one seminar listed is on distinguishing between "probability" and "statistical frequency," but it is offered jointly by a professor of philosophy and a professor of molecular and cellular biology! While this is no cause for fear of any kind, it is an acute reminder of the need of developing more courses, on our own or jointly with others, in order to meet the substantially increased practical and intellectual demand of our beloved discipline.

6. DEEPEN OUR FOUNDATION WHILE EXPANDING OUR HORIZON

Evidently, by now, few would question the ubiquity of statistics, to a point that some of us actually worry about too much fragmentation or our identity becoming too diluted as our horizon continues to expand. Indeed, some may have reservations about Brown and Kass's (2009) call to loosen the definition of a statistician out of similar concerns. The broad context in which Brown and Kass casted their definition, particularly their call that "the primary goal of statistical training, at all levels, should be to help students develop *statistical thinking*," makes it clear that the real issue here is how to elevate our general pedagogical effort so that many more people can appreciate statistical thinking in real terms, and put it into use for their own benefit, regardless whether they would be labeled as statisticians or not.

This brings a key point*: To foster more statistical thinking and to effectively prevent fragmentation, we must continuously deepen our foundation as we expand our horizon*. By "deepen our foundation" I mean to engage ourselves, and encourage others to do the same, in deep statistical thinking whenever possible, and not to be contented only with the methods or results we produce. This includes efforts such as revealing how several seemingly unrelated methods or applications actually share the same core, or identifying what part of a new area of applications is within the realm of existing principles and theoretical insights, and what part needs extensions or even a whole new set of concepts and principles.

The emerging area of "large p small n" demonstrates well the latter need. Indeed, the quest for the appropriate theoretical and methodological frameworks for dealing with "large p small n" distinguishes professional statisticians from ad hoc "data miners," i.e., those who immerse themselves in finding "features/signals" in the dataset at hand without seriously worrying whether the finding is statistically and scientifically meaningful. A key sign distinguishing a professional from an amateur is the person's ability to assess what can be done, what cannot be done, and what should not be done even if s/he has all sorts of incentives to do so (e.g., Thou shalt never substitute a casual analysis for a causal study).

The critical importance of such foundational understanding at the individual level and foundational deepening at the disciplinary level is perhaps best illustrated, unfortunately, by the Madoff or "Made-off" fiasco. Evidently Mr. Madoff gambled his giant, hollow scheme on people's lack of understanding of the fundamentals of investment returns and risks, or perhaps rather on people's tendency not to dig deeper when results appear to be desirable—why should I dig more when I already have what I wanted? This tendency or attitude, I believe, is responsible for a substantial portion of false discoveries in science, misinformation in media, and misguided policies in our society.

We statisticians, as a police of science (a label some dislike but I am proud of; see the next section), have the fundamental duty of helping others to engage in *statistical thinking* as a necessary step of scientific inquiry and evidence-based policy formulation. In order to truly fulfill this task, we must constantly firm up and deepen our own foundation, and resist the temptation of competing for "methods and results" without pondering deeply whether we are helping others or actually harming them by effectively encouraging more false discoveries or misguided policies. Otherwise, we indeed can lose our identity, no matter how much we are desired or feared now. Again, "Made-off," or more generally the current financial disaster, is a great reminder of an ancient wisdom: without a real substantial foundation, the larger a building, the easier it tumbles.

7. THE NEED TO INCREASE SCIENCE POLICING TO COMBAT "INCENTIVE BIAS"

As I argue above, a key reason to call for continuously deepening our foundation is to encourage ourselves and others to think harder and deeper, especially when incentives for rushing are so great. But could this lead to more "inaction," as Brown and Kass (2009) worried? Brown and Kass caution us not to instill excessive cautiousness in teaching our own students. I, of course, agree—nothing excessive is good. My worry, however, is that we are far behind in instilling the appropriate level of caution in scientists and their students. Too many false discoveries, misleading information, and misguided policies are direct consequences of mistreating, misunderstanding, and misanalyzing quantitative evidence. I am not referring to those deliberate efforts to mislead, such as infomercial statistics or unethical behavior (e.g., a highly cited author from another field told me, to my face, that he avoids precise model descriptions so readers can never be sure what he did and hence be able to challenge him). I am referring to honest mistakes made by scientists and policy makers, mistakes that could easily be avoided or caught if they themselves had been "instilled" with an appropriate amount of statistical thinking and caution.

I came to this realization after having worked with astronomers, engineers, geophysicists, psychiatrists, and social scientists. "Wait a minute, are you bragging?" some readers might question. "We don't see you publish in these fields much or at all!" Exactly—this is why I bring up my experiences. Over the years, especially after I joined Harvard, I have spent numerous hours (and taken many trips) to conduct collaborative research, attend project meetings, nonstatistical conferences, etc. I, however, have published very little in those areas, mainly for the following two reasons.

First, most of the time my role in these collaborative or consulting work has been "quality control" or even "damage control." I tell my collaborators what parts of their conclusions are primarily based on their belief or desire and not on the data analysis their research assistants did. I explain to them why the data they have could not possibly lead to the conclusion they hoped for, no matter how fancy the software their assistants adopt; or why their significant results are actually nonsignificant when more appropriate variances are used. All these, of course, do not lead to any publication (other than the current paragraph), but this is exactly what my professional duty calls for—one less erroneous "scientific discovery"!

Second, obviously, my police work is not always effective. When a backyard is dirty, but the host insists on having an open house (with the backyard closed for inspection) because the host is in desperate need of selling the house, all I can do as a backyard cleaner is to prevent my name from being used to vouch for the cleanliness of the backyard. Ironically, getting one's name off an article often requires more diplomatic skill than getting it on one. For a subject-matter article involving some degree of data analysis (not necessarily statistical!), a statistician's name in the authorship list is the most effective way of fending off (nonstatistical) reviewers' questions of the validity of the analysis. We statisticians would be doing science and society a tremendous service by refusing, as frequently as possible, to have our names used as evidence for sound statistical analysis unless it is indeed so in our uncompromised judgment.

Some readers may consider this far too noble or impractical. Many of us cannot afford investing time and energy without tangible reward—I cannot put on my CV that "I prevented three false discoveries" even if that actually is the most substantial contribution I have ever made to humankind. And indeed, how could anyone verify my claim? But this is exactly the source of the problem—our general reward and evaluation systems inherently incentivize false discovery. An article containing an erroneous statistical analysis is still an addition to one's CV. What is the punishment if I publish an article claiming strong evidence of discovering a disease gene, but later found not to be so? Not much. I would be just one of the many who have made similar claims, and I always have that "5%" statistical error to fall back on. But what if my guesstimation is actually correct? Someone has to win the lottery, right?

This is not a cynical view, but a serious reminder of the great temptation for all of us to succumb to a "leap of faith." We all have the tendency, precisely "for practical reasons," to produce and interpret results in ways that are more guided by incentives, however subconsciously, than by statistical or other scientific evidence. I, for one, despite my "noble talk" above, have items on my CV that you can certainly throw in my face with disgust: "Xiao-Li, so much for your police work—here is clear evidence that you have been on the wrong side of the law!"

If this reminder is still insufficient, let me further confess that I am committing this "incentive bias" crime repeatedly right now because I am using stories and anecdotes to support my arguments, an approach that is hardly scientific or statistical. But you are now warned, exercise caution when being intoxicated by my stories and anecdotes so you can stay on the right side of the law!

8. THE INCREASED DEMAND AND NEED TO HELP OTHERS SELF–POLICE

"OK, we can be noble to our heart's content. But why then would anyone want to work with statisticians, if we keep giving them trouble instead of what they want?" My response is that *our professional call, and ability to prevent others from using quantitative evidence erroneously or inappropriately, is precisely what makes statistics, as a discipline, unique, wanted, and increasingly so*. This is our profession's life line, something that I am not aware of any other discipline trying, or even having the desire, to compete for (at least so far), yet more and more scientists are requiring their students to develop "self-policing" ability.

Successful scientists comprehend thoroughly the importance of identifying and understanding limitations and impossibilities, and learning from failures and mistakes. Indeed, the most impressive part of the proposal, as listed in Section 1, is its call to teach students "to recognize when an issue cannot be settled on the basis of the available evidence," and to "familiarize students with some of the mistakes human beings typically make in reasoning and problem-solving." This is a clear call for increasing students' ability to self-police and to understand when conclusions cannot be drawn or should not be drawn. Multiple scientists at Harvard tell me that what they want us to teach are actually not the technical methods themselves. As one anthropologist put it, "I can teach my students how to use chi-square, but I need you to teach them when it is appropriate to use it, and more importantly, when it should not be used at all."

A basic reason for this increased emphasis on self-policing is the realization of the surge of false discoveries; the exponentially growing amount of quantitative information available online or elsewhere has made it much easier for data snooping, deliberately or inevitably, for anyone who is equipped with suitable software or a quantitative assistant. For example, one geneticist at Harvard told me that he now pays attention to any "gene discovery" study only if it uses a Bonferroni correction. He considers the rest "garbage" because it is his observation that only those with Bonferroni corrections ultimately have a chance to be confirmed. I found this observation intriguing, not because it goes against the theoretical extreme conservatism of the Bonferroni correction, but rather because I wonder whether the use of Bonferroni corrections is a telling sign of the study investigator's quality and integrity as a scientist, or instead a reflection that the evidence is so overwhelming that the investigator was not incentivized to report anything else. Either way, the moral of this anecdote is that the surge of false discoveries is, perhaps ironically, providing convincing empirical evidence of their grave negative impact, which has encouraged scientists to do more self-policing and call for more training in that regard for their students.

As another example, during a recent seminar presentation by an MIT computational biologist, I asked her what type of error is considered more serious in her field, false positive or false negative. Her immediate response was, "by far the false positive." Intrigued by her assertiveness, I asked her why. "Well, the reason is very simple," she responded, again without any hesitation, "Even if the false positive rate were zero, we still don't have nearly enough resources to experimentally verify all the claims."

Brown and Kass (2009) criticized a potential "cavalier attitude" by statisticians. Again, I agree that if, as they point out, all we do is to "shudder" then we are not helping anyone, but only harming ourselves. My emphasis is that through our *action,* not "inaction," we will help to instill "inaction" in others whenever there is not sufficient statistically sound evidence to support their actions. That is, our action is to help others to not *overreact* to the quantitative evidence they have. Serious scientists appreciate this role of statisticians and want more and more of their own research assistants to have such "self-policing" ability. Here is another personal story, with details blurred for confidentiality reasons. The story also illustrates the importance of maintaining, at least for some of us, a certain degree of "detachment" to the subject matter we are asked to help with, much like the importance of maintaining independence between the three branches of the U.S. government.

Pat, a well-respected social scientist, was going to present a major finding that would provide empirical evidence against a previously theorized difference. Realizing that this finding could cause considerable controversy, Pat called me in several days before the delivery, as I was known to Pat as a "Statistical Policeman." I didn't know much of the subject matter, nor did I have time to dig into the details, so all I could do was use my statistical instincts. The difference estimates provided by Pat's assistant were strikingly and consistently small across several groups, which was what made Pat excited.

However, to me, a statistician detached from the subject matter, the same "strikingly and consistently small differences" pattern was a smoking gun, especially when viewed against the group sizes. I literally did not care whether the theorized difference existed or not; what I cared about was whether Pat's empirical findings were statistically valid, in my "unincentivized" judgment. (Of course, one can argue that I also have my own "incentive bias," that is, to maintain my "Statistical Policeman" reputation. But that works exactly in the opposite direction to Pat's "incentive bias," and indeed is what Pat called on me for.)

So I asked Pat's assistant what he did. He explained to me his understanding of what Pat wanted, as well as the difficulties of producing stable results because of the very small sample sizes of various groups. As he has juggled with such problems many times before, he pooled the data in various ways until he could find a stable fitting to the model. He then used the fitted model to predict the difference as Pat wanted.

Pat later thanked me repeatedly because I prevented a professional disaster—all these wonderfully small differences were, you guessed it, artifacts from the pooling. This was truly an honest mistake, or I should say, miscommunication between Pat and the assistant. And the incident, I believe, is not uncommon. Many big-name scientists are too busy to check the details of the analyses done by their assistants. Some of them, frankly speaking, do not even know what to check, or have too much faith in "computer results." (I had a collaborator who was quite surprised to learn that results from a statistical software may not be trustworthy.) They rely on their substantive knowledge to judge whether the results "make sense." This is, of course, what most of us do, just as I relied on my statistical common sense to spot the problem in Pat's results. But this very common practice is also a core source of the "incentive bias" when our sensory bag does not contain enough senses; as we all know, many "findings" can be rationalized in ways we hope for with "common sense," especially in areas where not much is understood or variability is high. Or, as the British science writer Hanlon (2007) put it, "The history of science is littered with spectacular claims . . . , usually made by charismatic and highly-qualified people, that fade into nothing." Having an independent check by an unincentivized party is an essential way to reduce such claims.

All these remind us time and again of the importance of teaching statistical thinking, especially to students from other fields, as many mistakes can then be easily spotted or even avoided in the first place. Teaching statistical thinking is, therefore, particularly important for the courses designed for other disciplines, such as SOS courses discussed earlier. Although the materials and emphases are different, much of the concepts and principles remain the same, and this can be conveyed to students with real-life examples that they can all relate to regardless of their subject interests.

Take again the bias-variance trade-off, one of the very few absolutely fundamental principles in statistics, one that should be taught in every introductory statistical course regardless of the subject orientation. Students should be told that it comes in many forms and shapes, such as efficiency-robustness trade-off, parametric-nonparametric trade-off, etc., but that they are all fundamentally the same. In my own teaching, the following "parking dilemma" has worked well for illustrating its ubiquity.

The parking garage I use has seven floors. In many wee hours, my memory is in sleep, leaving me walking up and down the stairs in search for my car. So I told my students, "Well, here is an example of efficiency-robustness trade-off. There is always space left on the seventh floor, so it would be very robust if I always park my car there, as I'd always know where it is. But of course this would not be most efficient in terms of the stair walk, because often there are also spaces available on a lower level."

"However, parking on a lower level is efficient only when my memory can be trusted, just like your model assumptions give you more efficiency only when the assumptions can be trusted. Otherwise, you would be better off by using a more robust approach, just as I would save time if I always parked my car on the seventh floor when my memory is not working!"

Several months ago a former student told me that he still remembers this "parking trade-off" even though I don't remember when he took my course! All these wee-hour disturbances disappear, when I think about how many future costly mistakes or frustrations are avoided because my students remember their professor's parking dilemma.

9. THE FIVE IDEAL QUALIFICATIONS FOR TEACHING INTRODUCTORY STATISTICAL COURSES

All the aforementioned teaching tasks reinforce a key point: in order to successfully meet the "sea-change" demand, we must make a tremendous collective effort to change the "Statistics is *easy to teach*, but *hard (and boring) to learn*" perception to one of a "Statistics is *hard to teach*, but *easy (and fun) to learn*" reality. Specifically designed, carefully prepared, and well-taught courses have the best chance of convincing students that statistics is actually fun, easy, and worthwhile to learn, especially for those students whose career goals are not to become statisticians themselves. Good statistical courses, especially at the introductory level for other disciplines, are not at all easy to teach. They are best taught by those who have

(I) extensive statistical knowledge;
(II) deep understanding of statistical foundations;
(III) substantial experience in statistical practice;
(IV) great communication skills; and
(V) profound pedagogical passion.

And yes, I mean all five, with no priority given to any single one, because lacking any of these could lead to a mediocre or even disastrous course. I can easily list a handful of statisticians whom we would all agree possess (I)–(IV), and yet they are not known (or perhaps don't want to be known) as effective teachers.

Am I too intoxicated by the blended Bordeaux? "Xiao-Li, you must be kidding me!" I can see my fellow department chairs shaking their heads, "Where am I going to find such people to teach Intro Stat???" Certainly the issue is too urgent and too important to kid around. Although many of us do fall a bit short in one or more of THE FIVE, the list provides a scorecard which shows what perfect marks should be. Many of us do not obey the speed limits, and we often get away, literally, with driving at 70 miles per hour (mph) when the speed limit is 60 mph. But awareness of the 60 mph limit has surely prevented the vast majority of us from driving 90 or even 80 mph, which would not be an uncommon driving speed if no standard were in place. It is with the same spirit we should emphasize the most desirable qualifications for teaching introductory courses, for there has been a tendency to lower, deliberately or subconsciously, the requirements for their lecturers' qualifications because such courses are often viewed as "baby courses" or "service courses." The undertone here is that they perhaps do not deserve our best teaching resources, a perception I believe is rather unfortunate and dangerous, as I shall discuss further in the next section.

The list also sets an expectation for future statistical educators, which surely should be higher than the current one. We certainly do not want our students to do only what we can do.

Incidentally, as pointed to me by a colleague, the first three desired qualifications were essentially the same as what Hotelling called for in 1940 (see the reprinting and discussions as Hotelling 1988), who worried over the same problem: that introductory statistical courses, especially those for other disciplines, were not taught by qualified people. The problem then was obviously much more severe than it is now. Nevertheless, it is precisely those mistaught courses, together with some taught by those of us who lack the last two (or more) of THE FIVE, that have given both statistics and statisticians a bad name in the general scientific community and beyond—many students would naturally assume that anyone who teaches a statistical course must be a *qualified professional statistician.* It is, of course, a logical assumption, and our job is to make it true!

10. THE QUINTESSENTIALITY OF GENERAL INTRODUCTORY STATISTICAL COURSES

Why, one may ask, should we put so much emphasis on having our best qualified teachers for those introductory courses where many students have no (serious) interest at all in statistics? Given the severe shortage we already face, shouldn't we reserve our most qualified teachers for our own graduate and undergraduate major courses? This is certainly an understandable practice, as surely we want to provide our own students with the best possible education.

But let us also consider the impact of those general introductory courses. Even if we assume half of the professional statisticians have been practicing bad statistics (a number I certainly hope is too high!), there would still not be enough of us whose individual research publications or collaborative work could be held responsible and account for much of the current level of misuse and abuse of statistics in general. The general introductory courses have a far-reaching impact, considering the sheer volume of students who have passed through (though not necessarily passed) all the statistical courses taught in the U.S. alone each year. I don't have any data on that (perhaps ASA does), but a publisher told me several years ago that the total annual market for introductory statistical textbooks in U.S. colleges is roughly about half a million books. Suppose only 10% of those students receive bad statistical training, never question what they have been taught, and only they would potentially misuse or abuse statistics. We would still have produced, annually, 50,000 too many potential statistical abusers and misusers. (And I may have easily abused statistics here myself, as one may need very strong assumptions to justify the half-million estimate here.) Now further imagine that 1%, and only 1%, of these 50,000 will be teaching "elementary" statistics someday, somewhere, because they have taken a statistical course.

If you don't trust any of these figures (I don't), let us instead think about how we acquired our essential mathematical skills for our teaching and research: from working with mathematicians, from reading mathematical papers or books on our own, or from taking introductory mathematical courses? Now imagine that many of us had been taught by "mathematicians" who told us that $AB = BA$ for any positive definite matrices A and B, and 10% of us never questioned it.

With their potential impact in mind, it is easy to see the necessity of having the most qualified teachers for these introductory courses, just as for more advanced ones. And if I had to make a choice (and sometimes I do as a department chair), I surely will give the general introductory courses the highest priority for a very simple and practical reason. If an advanced course is sabotaged by bad teaching, the chances are that it will only affect a relatively small number of students, most of whom would have, or already have had, another chance to study statistics and to be convinced of our beloved subject's beauty and importance.

In sharp contrast, if a general introductory course is badly taught, it often will affect hundreds, or even thousands, of students, and the vast majority of them will never take another statistical course, even if some of them initially had some curiosity or interest in statistics. This is very much like a badly taught AP statistics course that can do more harm than help, permanently turning away many of its students, as all they saw was "Oh, this is what statistics is about—boy, am I glad that there are many more interesting and relevant subjects in college than this!" Indeed, among the Harvard undergraduates I asked, the most frequent reason for not considering a statistical major was a "turn-off" experience from an AP statistics course.

"So what? It is *their* loss," some may argue. "I only have time for those who are interested in what I do/teach." Well, the following anecdote might cause those arguing to think twice, as it did for me.

11. WE ARE NO SHRIMP!

Back in 2006, a few statisticians joined an effort organized by the American Mathematical Society (AMS) to urge Congress to approve the Administration's proposed increased funding to NSF. We were divided into small delegations of four to five people each. Each delegation, representing several states, had 15-minute appointments with some congressmen/women and senators from these states. Or, more accurately, with their *staff members*, most of whom, with no exaggeration to any degree, look exactly like those students sitting in our introductory statistical courses. They are young, smart, full of energy, and clueless about what we do.

Our job was to educate them, literally, in less than 15 minutes. We had a few well-made "Mathematical Moments" (*http://www.ams.org/mathmoments/*), a sheet of past NSF funding records to the state, and a prepared statement that we wanted each staffer to pass on to his/her boss. If anyone complains that the 15 minutes contributed talks at JSMs are too difficult to deliver, well, try this one! None of the staffers appeared to have any knowledge of what statistics was about. (We were, of course, previously warned that many of them are fresh college graduates or interns with degrees in law, government, or similar fields.) The most encouraging feedback was from one sharply dressed young fellow: “Oh, I’ve heard of probability.”

“So what?” my arguers might ask again, “Who cares if these staffers do or do not understand what we do? Their job was simply to pass on our messages.” Well, we wish! These staff members are bombarded by lobbying groups, 15 minutes each, literally, and we often had to sit, or more frequently stand, outside the office watching many other groups and individuals come and go. We were advised by AMS beforehand that it would be critical to convince these staff members of the importance of this funding to NSF because they are not receptionists, but screeners, and indeed very overworked screeners.

The revelation of the critical importance of teaching more and better general introductory courses came to me as we were leaving a Congressman’s headquarters. The next group had already started their 15 minutes education program before we could walk out of the front door. “We represent the shrimp industry from Cape Cod, and we urge the congressman to support this critically important local business.” Well, at the end of the day, which presentation would leave a more savory taste when the young staffer chows down over his daily offering? Shrimp or shrinkage?

But what if he had a fond memory of losing sleep over Simpson’s paradox, and then indulged himself with a big glass of OJ (with Vodka and jumbo cocktail shrimp) when he finally nailed it down?

12. THE WORLD IS COMING DOWN ON US . . . BUT WE CAN!

If there is any silver lining in the recent financial crisis, it is that it offers a public lecture, or rather a horrendously expensive lesson, about the critical importance of understanding and assessing uncertainty and risk. The financial module of the *Stat 105 Happy Course* introduces the concepts of mean and variance, which register much more rapidly when we refer to them as “expected return” and “volatility.” It also uses the much-recommended “diversifying principle” to introduce the concept of correlation. I surely had an easy time this past spring in explaining the consequences of not appreciating variance or correlation!

The grossly improper assessment of variance and correlation, either out of ignorance or greed, has brought down the (financial) world. And now that the world is down around us, our professional duty compels us to do our absolute best to educate our future trainers and trainees, and through them the general scientific community and public, about what statistics can and cannot do and why it is as essential to modern civilization as an election is to a democratic society.

Speaking of the election, as I wrote a good part of this article in a Baltimore hotel during the historic inauguration (as I needed to chaperon my son for his attendance at the inauguration), the spirit of “Yes We Can” has certainly been with me. Like the economic challenges we all face, I am fully aware of the challenges we statisticians face, and fully understand that the many tasks and needs articulated in Brown and Kass’s (2009) article, and in this follow-up article, will take years and even decades to accomplish or to meet. I also fully realize the complexities of many issues involved in what Brown and Kass (2009) proposed and in my supplemental proposals. For example, as I have been well reminded by several colleagues and students, in order to have the most effective student training and faculty teaching, we also need to consider issues such as admission standards and polices, tenure and promotion criteria, resource availability in small liberal arts colleges, etc. These are all very complex issues and some have been the subject of much on-going effort (e.g., Harvard’s effort in increasing the teaching requirements in faculty promotion and recruitment; see *http://www.nytimes.com/2007/05/10/education/10harvard.htm*).

Nevertheless, I am a strong believer of “Yes We Can,” or to put it more practically, “No, we really have no choice.” We are now in the spotlight, whether we like it or not, and it is in our best interest, as well as in (almost) everyone else’s interest, that we double our effort. Nothing in Brown and Kass’s (2009) proposals, nor in my supplemental ones, will be a panacea. But we all can start with one student at a time, one course at a time, one department at a time, and one institution at a time. *Culture can be changed more swiftly than we realize when genuine, collective, and sustainable efforts are made*. We started our required teaching course *Stat 303* in 2005–2006. Last year a member of my *Happy Team* told me that a first-year student asked him “Is it true that we used to put up teaching fellows without any training?”

[Received March 2009. Revised June 2009.]

REFERENCES

Brown, E., and Kass, R. (2009), “What Is Statistics?” (with discussion), *The American Statistician*, 63, 105–123.

Cassidy, S. (2009), “Teaching Matters: Support for Graduate Student Teaching,” *Colloquy: The GSAS Alumni Quarterly*, Winter, 6–7. Available at *http://www.gsas.harvard.edu/images/stories/pdfs/colloquy_winter09.pdf*.

Craiu, R. (2009), “The Statistics Debutante’s Dilemma,” *IMS Bulletin*, April, 3–5.

Garfield, J., Pantula, S., Pearl, D., and Utts, J. (2009), “Statistical Education Graduate Programs,” report of a Workshop Funded by an ASA Membership Grant, available at *http://www.causeweb.org/research/programs/GradProgram.pdf*.

Hanlon, M. (2007), *10 Questions Science Can’t Answer (Yet)*, London: Macmillan.

Hotelling, H. (1940), “The Teaching of Statistics,” *The Annals of Mathematical Statistics*, 11, 457–470. Reprinted in *Statistical Science* (with discussion), (1988), 3, 63–108.

Meng, X.-L. (2008), “Harvard Statistics: Quintuple Celebrations and Celebrating Quintuples,” *Amstat News*, August, 8–9.

——— (2009), “Statistics: Your Chance for Happiness (or Misery),” *Op-ed for The Harvard Undergraduate Research Journal*, Spring Issue, 21–27.

This article has been cited by:

1. Elart von Collani. 2010. Response to 'Desired and Feared—What Do We Do Now and Over the Next 50 Years' by Xiao-Li MengResponse to 'Desired and Feared—What Do We Do Now and Over the Next 50 Years' by Xiao-Li Meng. *The American Statistician* **64**:1, 23-25. [Abstract] [PDF] [PDF Plus]
2. Robert G. Easterling. 2010. Passion-Driven StatisticsPassion-Driven Statistics. *The American Statistician* **64**:1, 1-5. [Abstract] [PDF] [PDF Plus]
3. Roger W. Hoerl, Ronald D. Snee. 2010. Moving the Statistics Profession Forward to the Next LevelMoving the Statistics Profession Forward to the Next Level. *The American Statistician* **64**:1, 10-14. [Abstract] [PDF] [PDF Plus]
4. David R. Fox. 2010. Desired and Feared—Quo vadis or Quid agis?Desired and Feared—Quo vadis or Quid agis?. *The American Statistician* **64**:1, 6-9. [Abstract] [PDF] [PDF Plus]
5. Brian C. Kotz. 2010. Thoughts on the Importance of the Undergraduate Statistics Experience to the Discipline's (and Society's) FutureThoughts on the Importance of the Undergraduate Statistics Experience to the Discipline's (and Society's) Future. *The American Statistician* **64**:1, 15-18. [Abstract] [PDF] [PDF Plus]

Passion-Driven Statistics

Robert G. EASTERLING

As Meng (2009) made clear, one of the statistics profession's responsibilities is to be "the first quantitative trainers of future generations of scientists, engineers, policy makers, etc." (not just statisticians). Evidence suggests we have not met this challenge. In fact, our traditional Stat101 courses and texts can poison the statistical well for the people who become our potential sponsors and collaborators. We need to do more than teach 'methods.' We need to show from the first day and throughout the Stat101 experience that our methods exist to help people learn interesting things about issues and topics *they* are passionate about. This message pertains to the rising generations of professionals and the citizenry at large and it applies to statisticians. Getting the message across may require radically redesigned 'service courses' and a new generation of uber-teachers as Meng (2009) advocated. In the meantime we should use existing materials in ways that show how subject-matter passion can motivate statistical analyses that reveal interesting and important subject-matter insights. As we develop new texts and other materials we need better quality control by authors, editors, and reviewers to assure that our teaching supports our "first quantitative trainer" responsibility.

KEY WORDS: Passion; Statistics; Stat101.

WE DON'T GET NO RESPECT

Archie Bunker (on the TV program, *All in the Family*) once told his son-in-law,

> "Don't give me no stastistics (sic), Meathead! I want facts!"

What I think he was saying (with typical accidental profundity) was: We statisticians get our kicks from *stastistics* (the technical aspects of statistical data analysis), while our sponsors and collaborators are turned on by the *facts* (the subject-matter insights provided by data). We create Archie's impression of a difference between statistics and facts early on by lifeless, sometimes clueless, textbook examples that seem aimed only at teaching formula plug-in; there are no apparent or interesting facts either driving the investigation or revealed by the analysis. No wonder we are so often disparaged. If we want people to be passionate (and intelligent) about the use of statistical methods in their work, we need to show that enthusiasm for their chosen fields can be connected to an appreciation of statistical methods that will help them learn more about their world. And we need to do it at our first and maybe only opportunity: Stat101.

Meng (2009) identified the challenges this need poses, especially in the context of Harvard University's General Education curriculum. Improving Stat101 has long been a cause pursued by ASA and much innovative work has been done by individuals and departments. But it has not been enough.

Statisticians know the story. Meng (2009) recapped it. You're at a social gathering. Someone asks you what you do. You say you're a statistician. He or she says: I took a statistics class and it was the (choose one or more) worst/hardest/most boring/most useless/most confusing class I ever took!

It's been this way for the forty-some years I've been a statistician. Why haven't we fixed the problem? (I regret not spending more time on it.) Will it ever be fixed?

Two stories with snappy endings:

1. A professor friend had negotiated a price on a new car. The time had come for the Sales Manager to come in and seal the deal: "Glad to meet you. What do you do?" "I teach statistics." "Oh, no. I took a stat class et cetera, et cetera." Instead of ducking his head and tugging his forelock, my friend threw out his chest and said, "You've just insulted me and my profession. No sale. Good-bye, sir." And he left. When a car salesman can put your profession down (and I mean no offense to car salesmen), you're in trouble.
2. Brad Efron (2004) was at a faculty reception prior to receiving an award. A "nice lady" at the reception asked, "And what do you do?" Brad: "I'm a statistician." Nice Lady (after a pause): "Uh… What did you win for?" Brad: "I invented the mean." Bingo! That's why he got the award and why he served as President of ASA. Every statistician should read Efron's series of *President's Corner* columns.

My generation didn't solve our image problem. The articles by Brown and Kass (2009) and Meng (2009) encourage me to think that this generation has a chance.

It's not just an image problem. There are serious consequences for souring generations of professionals on statistics. A colleague at Sandia National Labs told me that a consulting

Robert G. Easterling is Itinerant Visiting Professor, 51 Avenida del Sol, Cedar Crest, NM 87008 (E-mail: *rgeaste@comcast.net*). My thanks to Xiao-Li Meng for his and his department's work at Harvard University and for the challenging and on-target views expressed in his article. Thanks also to the editor and an associate editor for their supportive and helpful comments. I especially appreciate Editor John Stufken's decision to invite discussions of a published article.

 DOI: 10.1198/tast.2010.09180

client told him that “I really didn’t want to come see you because I was so turned off by my Stat101 course, but my boss told me to come anyway.” (Good thing he had an enlightened boss.) I believe that progress—personal, organizational, corporate, national, planetary—depends on how well we (individually and collectively) deal with data. Can statistics as a profession play an important role in achieving progress, or will we be content to play, as Meng (2009) via Tukey described it, in everybody’s back yard while the adults in the house make all the decisions?

One of my revered professors at Oklahoma State University, the late Carl Marshall, had a motto that stuck with me:

> “The nice thing about statistics is that the nouns may change, but the verbs stay the same.”

I like Marshall’s motto. It’s a major reason I have enjoyed my career as a statistician. I could work one day on a study pertaining to the reliability of a complex electro-mechanical device and the next day on a personnel issue—very different nouns. The perspective and methods—the verbs—I brought to those problems, though, were pretty much the same. As I’ve thought about it in the recent years that I have been teaching, though, I think the statement contains the nugget of our problem. Archie Bunker’s complaint crystallized it for me: He wanted facts (nouns), not sta*s*tistics (verbs)! We statisticians get turned on by the verbs. Our clients and collaborators make their livings on the nouns. Pedagogically, we write textbooks and teach courses that focus on verbs: here’s how to construct a histogram, here’s how to compute a chi-squared value (or ask a computer to), and don’t pay enough attention to whether we’re learning anything useful about the nouns. We can do research on x and y and ignore nouns completely. We’ve got to connect nouns and verbs in meaningful, informative, even exciting ways. We can’t just hope the students will make the connection by osmosis or by sudden revelation after they leave our classes.

WHAT DO WE DO NOW?

The *Real-Life Statistics* course at Harvard developed by Meng and colleagues (2009) is organized around modules: Finance, Romance, Medicine, Law, ... “(S)tatistical topics are covered when needed by a module.” In other words, the nouns drive the verbs; the facts drive the statistics. That’s what happens in real life. Our courses should prepare students for their post-university life. Early results at Harvard suggest “happy students.” I hope Harvard will track these students as they move into their careers to see if statistical happiness continues and affects the way they pursue their passions.

Not all who are “the first quantitative trainers of future generations” can totally revamp a course and develop the materials and hire or prepare the instructors to make it work. Instructors can, however, use existing textbook problems in ways that better connect nouns and verbs. I tried to do that with respect to experimental design (Easterling 2004) by embedding selected textbook problems into realistic and maybe entertaining stories that show the interplay of subject-matter passion with statistical analyses. I’m sure many instructors already do this and I know I’m a Johnny come lately. This approach needs to be more widespread. We can also do a better job of quality assurance to assure that new textbooks, even if they follow the traditional Stat101 organization of statistical topics, which I have no major problem with, will provide a happy linkage of subject-matter passion with statistical analyses. More on these topics below.

WE COULD SCREW UP BIG TIME

Meng (2009) rightly worried about this outcome in the realization that the Harvard University Department of Statistics has been given the responsibility for “*being the first quantitative trainers of future generations of scientists, engineers, policy makers, etc.*” He asked whether we have a consensus on the qualifications of such trainers.

Wow! Does this mean we’ve been teaching all these service courses for years, often by graduate students (many of whom do a good job; I know I did, never mind that dampness behind the ears), and we didn’t realize that we were shaping/warping future generations of scientists, engineers, policy makers, and car salesmen?! I share Meng’s (2009) conclusion that badly taught Stat101 has contributed to statistics’ bad reputation. Moreover, we helped create the generation of executives who are now downsizing statistics groups in industry, government, and academia. We have “screwed up big time.”

Quite a few years ago I encountered the following problem (paraphrased) in a text that was new at the time and which I recounted in an earlier article (Easterling 2004):

> A shoe store owner records the shoe sizes for the last 20 purchases of women’s shoes in his store. Data are given. The student/victim is told: Test the hypothesis that the population median shoe size = 7.5.

I’m sorry, but I can’t think of any reason a statistically literate shoe store owner would do this analysis. (The regional manager calls and says a supplier hypothesizes that the median shoe size we should order is 7.5. “Will you test that for me?” Not likely. And what’s this ‘population’ business?) Will students trained on this sort of exercise be smart users of statistical methods in their careers or will they conclude that “This is the most useless class I ever took?” I’ve asked statistics graduate students, “Do you want to work for or with people who have been taught this sort of statistics? Chances are you will.”

At the time I just kind of shrugged this shoe store example off. Many statisticians in industry and government teach remedial statistics classes designed to help fellow employees recover from their Stat101 class experience. (“I’m Joe and I took Stat101 fifteen years ago.”) Hypothesis-testing overdose is one condition these classes treat.

In the last few years I have done some occasional university teaching—either Stat101 or Introductory Experimental Design. I’ve been dismayed at what I’ve seen in textbooks that I’ve considered or used. There are too many uninteresting (‘unhappy,’ in Meng’s 2009 terminology, actually sad) or technically incorrect examples. For example, they put data into histograms in their book’s histogram section even though the data have temporal, spatial, or other dimensions that should not be ignored and in fact are the most interesting aspects of the data. They do analyses that are driven by the section title, not the problem-context and the data.

Here's an exercise that I've seen in more than one text: Compare the annual home run data for Mark McGwire and Babe Ruth (we're talking baseball). Some students have baseball passion and most have at least some awareness of baseball so this example has curb appeal. Students are usually asked to create dot plots or box-and-whisker plots to compare the data. The result is pretty uninteresting. You would conclude that they're pretty similar home run hitters. However, these data are really time series. The time dimension could be important and if you implicitly tell students to ignore it, just for the practice of making a dot plot, you're telling them to ignore potentially useful information without even looking at it. When you plot the two time series, you see a rather dramatic difference in patterns. The first halves of their careers were pretty similar, but in the second half Ruth's HR productivity steadily declined while McGwire's "spiked," pardon the expression, sharply upward. The *possible* story (I have no proof: statistics means never having to say you're certain) behind this contrast in HR histories is: different dietary supplements then and now. The data, rightly plotted, give you something to talk and think about. Even the non-baseball fan can get the message: plotting data in clever (and correct) ways can be very revealing, and transfer that message into their own field. You can also use this example to talk about choice of measurement, say annual HRs or annual HRs per at-bat. You can also bring in Barry Bonds's data which may not have made it into the text you use. The textbooks miss all this good stuff.

In another example, the Olympic winning times in the 100 meter dash for men and women in the modern era are given and the student is instructed to make a histogram of all the data. Then notice that it is bimodal. There are situations where such a plot can reveal unknown "data streams," but not here. With only a minimum of subject-matter knowledge, people (including students) know that men run faster than women and that the event-times for both have decreased over the years. That knowledge says to display these data first as side-by-side or overlaid time series plots for men and women. Such displays could stimulate discussions of projections and predictions. We need to demonstrate the point that subject-matter knowledge should be used in shaping an analysis, not ignored just because the chapter is about histograms (I'm not picking on histograms, per se; that's where typical texts start and it is first impressions I'm worried about).

Many teachers can and probably do teach around these sorts of examples, and even use them (as I have) as horror stories of bad statistics (at the cost of destroying the credibility of a text they just asked their students to spend $125 for), but I worry about how many inadequately prepared teachers will simply present the text's material and how many students will passively follow. I worry even more about the sharp students who will be turned off of statistics by stuff they recognize as not making sense.

PASSION-DRIVEN STATISTICS

To explain this title I will repeat a story (Easterling 2004) that illustrates the importance of subject-matter passion in statistics.

The statistics group at Sandia has a small library and when we got a new book I would route it around to the group so they would be aware of it. One new book I routed had to do with graphical methods. Charlie Clark was thorough. He did more than skim the table of contents. One chart he came across was a scatterplot of automobile engine size (displacement) versus body weight. This plot showed a slightly curved positive association—heavier cars have bigger engines—and a couple of outlying points. The authors made the valuable (statistical) point that you couldn't 'see' the engine-size/body-weight relationship or the outliers in a table of the data and they commented that the outliers might be unusual cars or mistakes in the data. Then they went on to the next topic.

The outlying points were two cars with unusually large engines for their body weights. That means they would be high-performance autos, so Charlie (a car nut) got excited. He wanted one of those cars, so he looked up the source data (provided in the book's appendixes). Alas, they were the Opel and Chevette, which he knew were performance dogs. He then went to the original *Consumer's Reports* data source and found that transcription errors had been made. Sorry, Charlie.

The moral of the story is that Charlie found the true (unfortunately mundane) 'message' in the data, which is what statistical analysis is all about, not because he was a better statistician than the authors, but because he had a passionate interest in the subject-matter.

Examples abound of the role of subject-matter passion in inspiring and enhancing statistical analyses. Statistics is a team sport. We need to make and illustrate that point early and often, particularly in our service classes. I have looked at several handily available texts to see what kind of examples the authors have chosen to capture their students' attention in their first methods chapter. Is there an interesting story there illuminated by the data? Does it deal with a topic or issue that somebody could feel passionate about? Does the analysis depict good statistical thinking? Too often the answer is No.

WHAT DO WE DO NOW?

1. *Look more carefully at textbook or our own examples and modify and present them in ways that show the interplay of subject-matter insight and passion with statistical methods.*

Here's an example I use in teaching a Remedial Stat101 class at Sandia, modestly titled, *Essential Statistics.* The data come from an article by Gunter (1988) and were included in the text, *Chance Encounters* (2000) by Wild and Seber. Wild and Seber (2000) recounted the facts. I enhance the story a bit.

Ford Motor Company was experiencing problems in their cars' window mechanism: teeth were breaking on a small plastic gear and jamming the window, thereby incurring unexpected warranty costs and unhappy customers. Management formed a task force to investigate and solve the problem. Either the gears were encountering higher stresses than expected, or the gears did not meet their design requirement for strength.

The task force met. "We need data," they cried. "Let's get a bunch of gears and test them to measure their strength." Now here's where subject-matter knowledge first enters the story. To a layperson, these small 12-tooth plastic gears might look symmetric—no distinguishing features, front or back, or around the circumference. A process engineer, though, knows these gears are made by injection molding. A mold is filled with powder, then heated to liquefy the material, then cooled and out pops a gear. The injection port is on the end of one tooth and leaves a dimple on the finished product. You can't distinguish front and back, but you can identify the teeth by their radial position relative to the injection port. So, the process engineer on the team said, "We need to test teeth from every position (there are seven distinguishable positions or position-pairs) so that we can see if there are differences among positions that might tell us something about a problem in the process."

Some 163 gears were tested, with the number of tests per position ranging from 9 to 33. You can speculate why the replication was so unequal: Maybe no statistician was on the team. Maybe they didn't realize they could identify the test results by position until they had done the tests and fortunately the gears were still available for examination and all positions had been covered. (I envision a statistician being brought in to analyze the data. She asks lots of questions about the experimental units and the tests. That's when the process engineer makes the point that the positions can be distinguished and that could be important.) Other issues you can discuss are why 163 gears were selected for the study and where they came from. Can they be identified by production lot or date? Et cetera.

The breaking-strength data, when plotted by position, show a pattern: the strongest teeth are the two teeth adjacent to the injection port and tooth strength declines the farther you are from the injection port. The process engineer looks at the plot (Analysis 1. Plot the Data) and claps his palm against his forehead, "I know what's going on. Our supplier is shorting us! They're not injecting enough powder so we're not getting the material density we need throughout the mold." (If the team had not thought of separating the data by tooth position, they would have missed the whole story!) At this point, the statistician who is now on the team says, "I see the pattern you're talking about, but there's quite a bit of strength variability within positions. Let me do a bit of analysis to see whether the apparent differences could be just due to random variation among gears." She soon reports back that the pattern is real (if pressed, she can report a p-value, one of those "sta*s*tistics" that infuriated Archie). The team reports to the vice president in charge and recommends coming down hard on the supplier.

At this point it gets ugly. The VP says, "How come you and the supplier haven't been monitoring this process? How come you didn't catch this problem before we developed a major field problem? That's your job. You didn't do it. In the words of Donald Trump, You're Fired—all except the statistician, Mary. Mary, I want you to talk to the Executive Committee about how we could better monitor and improve the processes we're responsible for and how many statisticians we should hire to help us." (I am making this up.) End of story.

I realize that textbook authors cannot turn every example or exercise into a case study. Instructors, though, selectively can do this. Authors could provide supplementary teachers' materials that provide expanded stories.

2. *Self-Policing: Them and Us*

Meng (2009) discussed the statistician's role of "self-policing" studies—slowing down the tide of false discoveries. For example, Mary might have found that the eyeball analysis of the gear teeth data was not definitive in identifying real tooth-strength differences—it was all in the noise. Does this role mean that we should teach statisticians to police others and that we will teach future generations of professionals to ask statisticians to police their work? I doubt that that will sell (and Meng 2009 is clear that he wants us to teach 'statistical thinking,' not policing). I prefer a model of collaboration. If Mary had been involved from the start with the process and product engineers, she might have suggested they first run a pilot study to get an idea of variability, then she could have done some power calculations (which requires engineering input) to help design the main study (with more engineering input about costs and other resource considerations). Follow-up experiments with the supplier to find optimal and robust process settings (e.g., volume, velocity, time, and temperature) would have been an intelligent response to the finding that the process was off-target. (Another real-life lesson to teach: studies are seldom stand-alone; they lead to further investigations.) We want to foster the team concept, not the image of a statistical policeman arriving at the scene of a crime. Let's nip those false positives in the bud, not in the galleys.

Self-policing should start at home. We need to look at our existing course materials and ask: Do these materials meet the responsibility of being the first quantitative training of generations of scientists, engineers, and policy makers? Do they connect nouns and verbs or is it all verbs? Do they give some idea of why a person might be interested in the data and the message contained therein? The same sorts of questions apply to authors, editors, and reviewers of new texts and other Stat101 materials. When I scratch my head and wonder, "How did that get published?" I'm thinking of editors and particularly reviewers, too.

3. *Communication*

Lachenbruch (2009) and many others over the years have discussed the need for statisticians to be better communicators. Poor communication skills can contribute to our woes, but I worry more about bad ideas being communicated, especially via good communication skills! Some of the most student-friendly texts I've looked at, in terms of style and attention-grabbing examples, have been seriously lacking when it comes to illustrating statistical thinking.

THE DREAM

Fifteen years from now I meet a 30-something assistant in my Congressman's office. He asks me what I do. I tell him. He says, "Oh, I took Real-Life Statistics at Harvard. It was great! I see so many opportunities to apply the concepts and methods I learned there to issues that our office deals with daily. Right now we're looking at imported-shrimp quotas" (see Meng 2009, sec. 11).

[Received September 2009. Revised November 2009.]

REFERENCES

Brown, E., and Kass, R. (2009), "What Is Statistics?" (with discussion), *The American Statistician*, 63, 105–123. [1]

Easterling, R. (2004), "Teaching Experimental Design," *The American Statistician*, 58, 244–252. [2,3]

Efron, B. (2004), "President's Corner: But What Do Statisticians Do?" *Amstat News*, April 2004, 2–3. [1]

Gunter, B. (1988), "Subversive Data Analysis, Part II: More Graphics, Including My Favorite Example," *Quality Progress*, November 1988, 77–78. [3]

Lachenbruch, P. (2009), "Communicating Statistics and Developing Professionals: The 2008 ASA Presidential Address," *Journal of the American Statistical Association*, 104, 1–4. [4]

Meng, X.-L. (2009), "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" *The American Statistician*, 63, 202–210. [1,2,4]

Wild, C., and Sever, G. (2000), *Chance Encounters*, New York: Wiley. [3]

Desired and Feared—*Quo vadis* or *Quid agis*?

David R. FOX

The recent article by Meng (2009a) continues a long tradition of articles in this journal dealing with the future of statistics and statisticians. For over 25 years many of these accounts painted varying shades of the same grim picture—that our continued existence is under threat; the challenges are great; respect has been in short supply; and our future is bleak. In this article I suggest we spend less time scanning the cross-disciplinary borders for new intrusions and rather than shoring up the fortress, we open up the borders. I share Meng's upbeat enthusiasm for a bright future while recognizing much remains to be done to increase our relevance and effectiveness.

KEY WORDS: Accreditation; Communication; Future of statistics; Outreach; Statistical education.

1. INTRODUCTION

The article by Xiao-Li Meng (2009a) resurrected some old chestnuts about the relevancy of statistical science and statisticians. Following what seems to be a well-established tradition, I chose a title that poses a rhetorical question about the future of our profession. Loosely translated it means "*where to* or *what's the problem*?". While I believe it is healthy and appropriate for any organization to periodically engage in this sort of contemplative and reflective 'naval gazing,' there is a risk that introspection without intervention can lead to paralysis. Constant questioning of our worth, relevance, roles, function, and esteem with which others view us can undermine our self-confidence to the point where we lose our 'sense of place' and become unsure of our raison d'etre. Perhaps I am overstating it, but for as long as I can remember our profession seems to have been constantly searching for its 'place in the sun.' For me, the journey as a professional statistician has been, and continues to be, an exhilarating one. I could not have imagined that 37 years ago when I first enrolled in my undergraduate degree program that I would subsequently find myself working side-by-side distinguished scientists as a valued member of multidisciplinary teams. Nor did I envisage that my chosen profession and interest in its application to environmental protection would see me flying down remote and mountainous river gorges in Papua New Guinea or snorkeling over seagrass meadows in pristine waters off the coast of Western Australia.

There was a time when I believed that the role of an applied statistician was purely a supportive one—a bit like the ground crew that dutifully pulls up the rear in the Tour de France ready to provide assistance in case of a crash. During this period I remember reading a number of articles that sought to define what it means to be a statistician. One in particular left a permanent impression—not because it inspired me, but rather because it scared me. ASA past President Jon Kettenring's olfactory system was clearly on high alert when in 1996 he claimed "we smell trouble all around us." His statement (Kettenring 1996) that "other disciplines have been seizing opportunities that should have been ours" was, in my view, nothing more than naked paranoia. Thirteen years down the track and Meng (2009a) has reissued the apocalyptic warning that we still "have much to worry about or even to fear."

Rather than heed the call to arms and attempt to wrench back what apparently is rightfully ours, I decided a more effective strategy was to 'embed' myself in the research projects of others and to work from within to win over the nonbelievers. This was coupled with the provision of statistical training that was hands-on and delivered *in the context of the target discipline*. It was, as they say, 'hard yards' in the beginning and required a sustained commitment, but ultimately the strategy proved successful and I now find myself *directing* large multidisciplinary, multi-agency environmental projects as opposed to simply having cameo roles.

2. XENOPHOBIA?

Like many before him, Meng (2009a) challenged us to think about our future; to contemplate the possibilities; and to ask ourselves if we have the collective 'ticker' for "such a nerve-wracking task." His introductory comments and scene-setting quickly moved to the next section headed by yet another rhetorical question: "what should be our deepest fear?". My immediate thought on reading this was "*a lack of confidence in our own abilities*." To his credit, Meng (2009a) focused on the positives and provided many fine suggestions for 'lifting our game' and while I certainly wouldn't wish to detract from this enterprise, I am nevertheless struck with a profound sense of déjà vu.

In his Presidential address delivered on the occasion of this society's 141st Annual Meeting in Detroit in 1981, Ralph Bradley asked "what then is wrong with statistics and what should we do for its future?" (Bradley 1982). Noting that the shortage of doctorates in statistics "seems likely to reach a critical level in the very near future," Bradley's (1982) solution of attracting and retaining the best and brightest was, in the absence of a *strategy* for achieving this, yet another documentary on the apparent crisis in statistics rather than a survival guide for the future. In the same speech, Bradley emphasized "statistics as a science" and posed the question "have we failed to understand that experimentation and statistical analyses contribute only part of the information that goes into decision making?". Almost 20 years later, John Nelder warned us that "the public image of statistics is poor and may be becoming worse" and

David Fox is Professorial Fellow, Department of Civil and Environmental Engineering, University of Melbourne, Melbourne, Australia 3010 (E-mail: *david.fox@unimelb.edu.au*).

suggested that one of the biggest problems was the word 'statistics' itself (Nelder 1999). Like Bradley, Nelder (1999) was keen to establish a nexus between statistics and science, arguing that we should describe our discipline as *statistical science* rather than simply statistics. This was not a new suggestion and indeed Ruberg and Mason had promoted the same idea 11 years earlier (Ruberg and Mason 1988). Advance the clock to 2009 and we still find ourselves referring to our profession as 'statistics' and discussing ways to combat the perceived threats to our existence.

I found the Nelder (1999) article compelling reading at the time and even now as I refer to it for this correspondence, I am struck both by its prescience and apparent lack of uptake of key recommendations. Just two years prior to the publication of Nelder's article, the American Psychological Society was contemplating banning the use of hypothesis testing in its journals (Shrout 1997). As noted by Shrout (1997), this was not the first time such calls had been made; he cited the infamous case of the *American Journal of Public Health* which advised authors that "all references to statistical hypothesis testing and statistical significance should be removed from the paper" and that you should "delete p-values as well as comments about statistical significance." The journal *Epidemiology* adopted the same stance under the editorship of Ken Rothman. Rothman's advice to would-be authors was blunt:

> "you can also enhance your prospects if you omit tests of statistical significance . . . we do not publish them at all. Not only do we eschew publishing claims of the presence or absence of statistical significance, we discourage the use of this type of thinking in the data analysis, such as in the use of stepwise regression" (Rothman 1998).

The philosophical debates about null hypothesis significance testing (NHST) have been with us for many years and the attempts of a single misguided journal editor to deny the existence of a well-established mode of statistical inference were inevitably doomed from the beginning. While Nelder (1999) was equally strident in his criticism of sloppy statistical practice, his calls were not to ban p-values per se, but to demolish the *culture* of uncritical thinking that had developed around the routine application of hypothesis testing and the attendant practice of "asterisk hunting."

So what went so terribly wrong whereby some of our scientific colleagues wanted to exorcise themselves of statistics? And indeed, where were the professional statisticians and our societies during these debates? I believe such instances of high levels of dissatisfaction with statistics and the issuing of 'statistical fatwas' are the result of guilt by association. It is not that the statistical methodology is flawed; it is the indiscriminate and inappropriate application in other disciplines that undermines the integrity of our science. This is not dissimilar to the renewed debate in ecotoxicology over the legitimacy of a phenomenon called *hormesis* (Kaiser 2003; Douglas 2008). Hormesis is the term used to describe the apparent beneficial response of an organism to very low concentrations of a toxicant. Examples relevant to humans include vitamins, alcohol, and fluoride. The decision to include or exclude test results showing hormesis in the fitting of a species sensitivity distribution (SSD) can profoundly alter the declared 'safe' concentration for that toxicant. While hormesis is real and observable, it fell from favor in the 1980s as comparisons were drawn with homeopathy—a largely discredited alternative therapy (Wahlberg 2007) that claims medical efficacy through the administration of extremely low doses of preparations. In statistical science, I refer to the *statistical hormetic effect*—the discrediting of our profession due to the dilution of sound practice! So this brings us to the question of who has the right to call themselves a statistician?

3. ACCREDITATION

Apparently we are in short supply and always have been. More than a quarter of a century ago Minton informed us that "the demand for statisticians will far exceed the supply for the years 1978–1990" (Minton 1983). While in the middle of writing this article, the latest copy of AMSTAT NEWS arrived—Xioa-Li Meng and his "happy team" are featured. Meng's article commences with a quote from Google's Chief Economist, Hal Varian who predicts that the "sexiest" job over the next ten years will be that of a statistician (Meng 2009b). I hope he is right and that we can erase the parenthetical linking of statistics and misery from the title of Meng's (2009b) article.

Meng (2009a, 2009b) noted the difficulty that most students have with our subject area and warned us that "we could screw up big time. . . if we do not offer enough good quality courses." No issue here—I think we are all agreed that targeted, relevant, and stimulating courses in *statistical science* will be critical to our ability to thrive into the future—indeed, even to be seen as "sexy"! However, this internal debate about what to teach and, more importantly, by whom seems to have locked us into a perennial second of February (2/2)—our statistical Groundhog Day.

Kettenring's claim that other disciplines had effectively been ripping us off (Kettenring 1996) echoed the more restrained assessments of Kish (1978), Minton (1983), Billard (1998), and a host of others. By 1990 the decline in University Statistics Departments had already commenced (Barabba 1990), further cementing the 'in-house' teaching of statistics in some institutions. On the related issues of appeal and difficulty, Duckworth and Stephenson (2002) gave us a 'heads-up' when they noted that "today's courses in statistical methods, for the most part, focus on the same methods that were taught 30 years ago" and suggested that part of the reason we find ourselves stuck at 2/2 is our professional inertia and "natural conservatism in our academic departments." It is thus refreshing to see Meng's statistical group doing so well at Harvard.

Consensus on the issue of *who* should be teaching statistical science has been much slower in coming. While it is relatively easy to describe who we *do not* want, it is clearly more difficult to agree on minimum standards and competencies for persons deemed qualified (dare I say 'certified') to teach our subject. The issue of certification for the ASA has been on the table for a good 15 years or more and was comprehensively debated in the May 1994 issue of this journal. Recently the ASA Board of Directors endorsed a recommendation of the Individual Accreditation Proposal Review Group to begin a program of voluntary individual accreditation of (ASA) statisticians (Bock et al. 2009).[1] In announcing this decision (available at *http://www.amstat.org/news/VoluntaryAccreditationofStatisticians.cfm*), it was noted that it brings the ASA into line with its sister societies: the Royal Statistical Society (RSS); the Statistical Society of Canada (SSC); and the Statistical Society of Australia, Inc. (SSAI) in offering such a service to its members. Participation in the ASA's PStat. certification program is to be entirely voluntary with admission based on considerations of such things as: experience; competence; ethical standards; and communication skills. Although there is far from universal support for the Pstat. scheme, it must surely represent substantial progress to the establishment of minimum standards (over and above an academic degree) required to be formally recognized as a statistical educator/practitioner/researcher.

There is, however, one area of the PStat. process I believe requires strengthening and that concerns reaccreditation. As it currently stands, the initial accreditation will be for a period of five years after which time it *may* be renewed. The announcement on the ASA's website acknowledges that "there are many details to work out" and presumably the process of reaccreditation is one of those. In making its recommendation, the Individual Accreditation Proposal Review Group noted a variety of approaches to certification exist, citing examples in accounting, aviation, and project management. As both a private pilot and a chartered statistician (CStat.) I can attest to this. The validity of both my 'licenses' is indefinite but the similarity ends there. To remain a CStat. I need only pay an annual fee, whereas to retain my Private Pilot's License (PPL) I must undergo a comprehensive medical examination *and* demonstrate competency in the cockpit every two years. Furthermore, if I wish to fly a different *type* of aircraft or even the same type but at a different location, I will be required to undertake training and/or a flight check before being allowed out on my own. There are clear and obvious reasons for this and, while I am not suggesting the reaccreditation process for PStat. be as rigorous, I certainly believe there is merit in adopting an evidentiary-based review process.

4. QUO VADIS?

I think it is true that as a profession statisticians 'punch above their weight'—we are relatively small in number although our reach and influence has been, and continues to be, great. While the pace of breakthrough advances in statistical science may have slowed somewhat since the heady days of Gossett, Pearson, Fisher, Yates, and Neyman, the challenges facing our profession in the 21st century and beyond are no less daunting—albeit of a different nature. The preeminence of organizations such as Google which are gathering and linking massive and disparate datasets on spatial and temporal scales spanning orders of magnitude will demand new modes of analysis capable of rapidly teasing out *information* from terabytes of *data*. Likewise, the omnipotent threat of climate change and climate variability will sharpen the focus on the statistics of extremes. At the other end of the data continuum, our risk-based approach to life will require credible and scientifically defensible assessments to be made of events yet-to-happen or which may never happen. Statistical modes of analysis for such data-poor environments will, I predict, become more common. Practitioners and researchers in the life sciences have discovered Bayes and are demanding more training courses in Bayesian statistics. The software engineers have been quick out of the starting blocks and have unleased new generation software like Nettica (Norsys Software Corp.) and AgenaRisk (Agena Limited) for constructing and analyzing Bayesian Belief Networks (BBNs). I suspect that in some quarters, the enthusiasm for training in Bayesian statistics is less than the enthusiasm with which the software is being adopted and used. Such a situation sets us up for another fall as the Bayesian paradigm is pushed beyond its capabilities and/or is inappropriately applied.

I provide these examples to make the following points: (i) the future is ahead of us—not in the rear-vision mirror; (ii) the directions for statistical science will increasingly be determined by the (unprecedented) challenges facing society and mankind; and (iii) statistical training at undergraduate and postgraduate levels in all disciplines will need to be built on a traditional core but with stronger emphasis on robust, fit-for-purpose model and tool development in environments characterized by massive amounts of data; virtually no data; and extremeness.

In closing, I believe our profession has an incredibly bright future ahead of it and as Gnanadesikan (1990) observed, the core of our discipline is (still) in excellent health. The opportunities to make a difference in all walks of life abound. That is not to say we have been sitting on our hands for the last 100 years or more! Indeed quite the opposite. Careful planning, consideration of alternatives, and evaluation of our decisions will always be hallmarks of our profession and approach to science. The trick is to know when to draw a line under the introspection—to avoid the situation of 'paralysis by analysis' and to simply 'get it going' rather than 'get it perfect.' As the Individual Accreditation Proposal Review Group suggested to the ASA Board (Bock et al. 2009): "the time has come to make a decision—either launch a program, such as the one we are suggesting, or say definitely that this is not for us. It's time to move on!".

[Received September 2009. Revised September 2009.]

REFERENCES

Barabba, V. P. (1990), "What Does the Future Hold for Statisticians?" *The American Statistician*, 44, 125–128. [7]

Billard, L. (1998), "The Role of Statistics and the Statistician," *The American Statistician*, 52, 319–324. [7]

Bock, M. E., Hoerl, R., Kettenring, J., Kirkendall, N., Mason, R., Morganstein, D., Nair, V., O'Neill, R., Oppenheimer, L., and Wasserstein, R. (2009), "Report to the ASA Board of Directors by the Individual Accreditation Proposal Review Group," American Statistical Association (July 24, 2009). [8]

Bradley, R. A. (1982), "The Future of Statistics as a Discipline," *Journal of the American Statistical Association*, 77, 1–9. [6]

Douglas, H. (2008), "Science, Hormesis and Regulation," *Human and Experimental Toxicology*, 27, 603–607. [7]

[1] Since first preparing this article, the ASA has announced the formation of an accreditation committee.

Duckworth, W. M., and Stephenson, W. R. (2002), "Beyond Traditional Statistical Methods," *The American Statistician*, 56, 230–233. [7]

Gnanadesikan, R. (1990), "Looking Ahead: Cross-Disciplinary Opportunities for Statistics," *The American Statistician*, 44, 121–124. [8]

Kaiser, J. (2003), "HORMESIS: Sipping From a Poisoned Chalice," *Science*, 17, 376. [7]

Kettenring, J. (1996), "Strategic Planning—The Future of Statistics," *Amstat News*, 231 (May), 3. [6,7]

Kish, L. (1978), "Chance, Statistics, and Statisticians," *Journal of the American Statistical Association*, 73, 1–6. [7]

Meng, X. (2009a), "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" *The American Statistician*, 63, 202–210. [6,7]

——— (2009b), "Statistics: Your Chance for Happiness (or Misery)," *Amstat News*, 387 (September), 42–45. [7]

Minton, P. D. (1983), "The Visibility of Statistics as a Discipline," *The American Statistician*, 37, 284–289. [7]

Nelder, J. A. (1999), "Statistics for the Millennium," *The Statistician*, 48, 257–269. [7]

Rothman, K. J. (1998), "Writing for Epidemiology," *Epidemiology*, 9, 333–337. [7]

Ruberg, S. J., and Mason, R. L. (1988), "Increasing Public Awareness of Statistics as a Science and a Profession—Starting in the High Schools," *The American Statistician*, 42, 167–170. [7]

Shrout, P. E. (1997), "Should Significance Tests Be Banned?" *Psychological Science*, 8, 1–2. [7]

Wahlberg, A. (2007), "A Quackery With a Difference—New Medical Pluralism and the Problem of 'Dangerous Practitioners' in the United Kingdom," *Social Science and Medicine*, 65, 2307–2316. [7]

Moving the Statistics Profession Forward to the Next Level

Roger W. HOERL and Ronald D. SNEE

In an important and timely article Meng (2009) has raised important questions regarding the future of the statistics profession. We elaborate on several of his points and offer some additional opportunities for the profession to consider. We argue that statistical methods and tools must be properly integrated into an overall approach to scientific inquiry in order to be properly understood and utilized. The discipline of statistical engineering, defined in this article, provides a mechanism to do this based on research and theory. Similarly, statistical thinking provides a clear framework to help students understand the "big picture" of statistics, and a relevant context for its application. Further, there is a natural, synergistic linkage between statistical thinking, statistical engineering, and statistical methods. We believe that teaching this linkage to students and utilizing it widely ourselves will enable the profession to move forward to a higher level of impact.

KEY WORDS: Statistical education; Statistical engineering; Statistical thinking.

INTRODUCTION

We applaud Meng's (2009) recent article and the *American Statistician*'s leadership in publishing articles addressing the future of the profession (see also Brown and Kass 2009). We believe that such discussion is particularly important at this point in our history. There is much to agree with in Meng's article and we will reinforce several of his main points. We will also provide a different perspective beyond that of academia. We have spent most of our careers in business and industry, and have also taught in major research institutions, worked with nonprofits, and one of us (Snee) has worked extensively in pharmaceuticals.

The important points we emphasize include Meng's suggestions that academic departments prioritize introductory courses, that these courses should actually be "happy" experiences for students of all backgrounds, and that they should emphasize statistical thinking over coverage of as many methods as possible. One point on which we may have a different view than Meng is the roles that statisticians should play when working with other professionals. Meng appears to be promoting a "bad cop" role—slowing down other researchers to avoid publication of bad statistics, while we would suggest a "good cop" role—facilitating other researchers to do good statistics in the first place, and thereby enabling them to move faster. Integrating this approach with Meng's other main points mentioned above should help our profession have greater impact.

Roger W. Hoerl is Professor, GE Global Research, Applied Statistics Laboratory, 1 Research Circle, Niskayuna, NY 12309 (E-mail: *hoerl@crd.ge.com*). Ronald D. Snee is President, Snee Associates, LLC, 10 Creek Crossing, Newark, DE 19711.

Meng notes that statisticians graduating today are getting jobs. That is good news. However, we are also concerned about the impact these statisticians will be able to have on society throughout their careers. Significant societal impact is a much higher bar than employment, or even academic funding. We suggest that much more change is required of the statistics profession.

STATISTICIANS AS LEADERS VERSUS FOLLOWERS (GOOD COP, BAD COP)

Meng rightly points out that statisticians can play a useful role in society by limiting the claims made by other scientists based on faulty statistical studies. He refers to this function as playing the "statistical policeman" role. We call this playing the "bad cop" role, in that bad cops fundamentally slow down the research of other disciplines. We agree with Meng on the need to avoid bad statistical studies. We further believe that our profession could play an even more valuable role by being proactive rather than reactive, by helping people do better science in the first place. This will ultimately enable them to move faster, not slower.

We would call this the statistical leadership or "good cop" role. Our discipline has the tools to help scientists reach valid conclusions faster than they can otherwise; this should be our primary message. Doing so will require a proactive approach, and true leadership. Being a leader will be a new venture for many statisticians, perhaps even an *adventure*. This is a critical step to helping our profession achieve its full potential, as displayed in Figure 1.

There are several questions that must be answered and skills to be developed to become leaders. The first question is "What is statistical leadership"? Leaders help people move from one paradigm to another paradigm, from one way of thinking, working, and behaving to another. For statisticians this means helping people and organizations effectively use statistical thinking and methods to improve their own work. In this "good cop" role statisticians in all arenas would be doing such things as:

- Acting as team leaders for projects, and accepting the requisite responsibilities
- Identifying significant, complex problems that are of major importance to society or the organization that employs them
- Designing statistical education and training systems that the organization needs to be successful
- Working with both managers and technical people
- Accepting, indeed seeking out, broad roles and accountability.

 DOI: 10.1198/tast.2010.09240

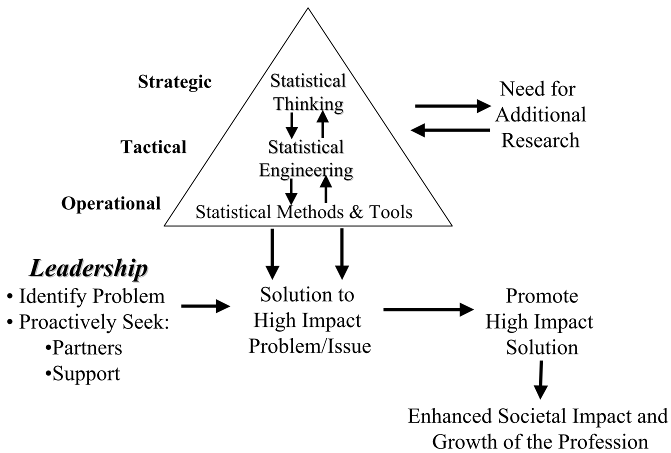

Figure 1. Statistical leadership.

For Figure 1 to become a reality, statisticians will have to assume more leadership roles; it is not likely that they will be handed to us by anyone! This means that we would be helping determine which problems are high impact and potentially solvable, not just waiting for others to make this determination. We then need to proactively seek support, financial and otherwise, seek out partners who can help us find solutions, and then work to ensure that potential impact is fully realized and sustainable.

The concepts of statistical thinking and statistical engineering are tightly coupled, and also link to statistical methods and tools. As illustrated in Figure 1, statistical thinking is the *strategic* aspect of our discipline that provides conceptual understanding and the proper context. It answers the question "Why should we use statistics?"

A well-developed discipline of statistical engineering would provide such approaches based on theory and rigorous research. This provides the *tactical* aspect of statistics, in that it answers the question "What overall approaches should be used?" Individual statistical methods and tools provide the operational aspect and answer the question "How do we implement these approaches?" The *strategic*, *tactical*, *operational* model of leadership has been around since antiquity, and has been applied to the statistics profession at least since 1990 (Snee 1990).

A critical question that must be answered is "How do we train statisticians to be leaders in business, industry, government, and academia?" There is a large volume of leadership material in the literature. It seems reasonable that this material should be used as a basis for some continuing education courses on leadership as part of the Joint Statistical Meetings, or as a seminar series in academia. An underlying theory of leadership exists, and we should be doing more to understand and apply it.

The next step is to ensure that scientists and professionals understand how statistical thinking, methods, and tools should be properly leveraged, along with other disciplines, to reach the most valid solutions as quickly as possible. If scientists and engineers understand statistical thinking—how to think about statistics and its application, and are familiar with statistical methods and tools, they will police themselves, freeing statisticians to do higher value-added work that will bring greater impact, not to mention added respect, reputation, and influence for our profession.

With this "good cop" approach, statisticians will be preventing poor statistical practice from occurring rather than correcting problems after the fact. As noted by Meng, teaching other disciplines more and more statistical techniques has not worked well in the past, and is not likely to succeed in the future. Without a good conceptual understanding of why statistics is needed and how it fits in with other disciplines—the "big picture," statistics often appears to be just a miscellaneous collection of mathematical formulas.

CRITICAL ROLE OF STATISTICAL THINKING

We applaud Meng for his bold emphasis on statistical thinking, particularly as a focal point for introductory courses. Discussion of statistical thinking by our profession is critical, particularly at this time as we are discussing how to make the training and education we offer more effective and a "happy" experience for diverse students. A fundamental of good research is to review the literature and build on what has been done before, guided by available theory.

Dictionary definitions of theory are typically something to the effect of: Theory is a plausible or scientifically acceptable set of general principles offered to explain a phenomenon. Theory is conceptual, based on general principles, and is not limited to mathematics. Deming (1993) admonished us several years ago that "there is no knowledge without theory.... Experience teaches nothing without a theory.... Without theory you have nothing to revise, nothing to learn from.... You have no way to use the information that comes to you."

A rich literature and theory of statistical thinking has existed from the mid-1980s to the present (e.g., Moore 1990; Cobb 1992; Wild and Pfannkuch 1999; Chance 2002; Hoerl and Snee 2002; Snee 1990, 1999, 2008). A recent Google search produced more than 1.8 million hits for "statistical thinking." The American Society for Quality has published a definition (ASQ 1996) stating:

Statistical Thinking is a philosophy of learning and action based on the following fundamental principles:

- All work occurs in a system of interconnected processes
- Variation exists in all processes
- Understanding and reducing variation are keys to success

As does Meng, this definition emphasizes thought processes rather than a set of techniques. That is, the operative word in statistical thinking is *thinking*, not *statistical*. *Statistical* is only an adjective that modifies the noun *thinking*. We have found this definition to be relevant and applicable to people working in virtually all fields and disciplines, at least once properly explained.

It identifies the key elements of statistical thinking: process, variation, and data. By first understanding the process that produced the data, especially in light of the overall system involved, people develop a relevant context for any data. This proper contextual understanding often prevents people from blindly walking into the faulty data analyses mentioned by

Meng. We have also found the emphasis on understanding variation, and where relevant reducing it, to be quite useful in practice, rather than treating variation as a nuisance parameter.

This definition also illuminates the unique relationship between statistical thinking and statistical methods; a desire to understand and reduce variation naturally leads to use of statistical methods, once the context of the data (process view) is properly understood. The triangle in Figure 1 illustrates this point—through the connector of statistical engineering, and provides some underlying theory for the concept. For any idea to stand the test of time, a foundation and theory must exist and then continue to grow.

Agreeing with Meng that introductory courses should emphasize statistical thinking rather than statistical tools in isolation, with no unifying theory, leads us to another critical and difficult question: How should statistical thinking be taught? Fortunately, theory, literature, and relevant experience exist here also. Psychological and behavioral research has identified educational principles that can greatly increase the effectiveness of our teaching and training (Forester 1990; Hoerl and Snee 2002, pp. xiii–xvii). This research suggests that instruction is most effective when it proceeds from teaching the big picture first, such as overall approaches to scientific inquiry or improvement, and then demonstrating this big picture via case studies, prior to getting into the mechanics of individual tools.

Gale Bryce taught introductory statistics courses based on these principles at Brigham Young University (Bryce 2004); call it a non-randomized experiment! Bryce discussed the feedback he received from the students in the class. They reported that the approach used by Bryce enabled them to learn more and enjoy the experience more than they thought they would prior to taking the course, and that they perceived the students taking the traditional course did.

Meng suggested using a variety of real problems to help create this context in the "happy" course. We certainly agree with the need for real problems. However, Bryce also taught a unifying theory of how the tools fit together into an overall approach to scientific inquiry that transcended the individual problems. Such an approach helps students develop a theoretical understanding of how to attack unstructured problems, which ties directly to the concept of statistical engineering.

NEED FOR A GREATER FOCUS ON STATISTICAL ENGINEERING

One aspect of Meng's article that particularly intrigued us was his discussion of Stat 399 at Harvard, *Problem Solving in Statistics*. We are delighted to see this type of course on the syllabus at such a prestigious university. We applaud Harvard for offering a course that "…emphasizes deep, broad, and creative statistical thinking instead of technical problems that correspond to an identifiable textbook chapter" (p. 204). Such a course will undoubtedly better prepare graduates to address real-world problems, virtually none of which "correspond to an identifiable textbook chapter."

The existence of this course also raises some interesting theoretical questions that we fear our profession has not yet addressed. For example, what approaches are taught in this course for attacking deep, broad problems that require more than one technique to solve? Certainly students are not taught to make up a new approach for every problem faced; there must be some overarching themes that unify statistics as a discipline in this context. As noted previously, Bryce (2004) taught one such approach in his courses. Unfortunately, our profession does not have a rich literature on how to tie together various statistical methods into overall approaches to attack deep, broad problems. This is why those of us in the private sector have had to invent our own approaches, such as the Seven Step Method (Kume 1985) and Six Sigma (Snee and Hoerl 2003).

Second, what is the underlying theory given to students to justify these approaches over alternatives? To date our profession has developed virtually no theory, not even theory based on empirical experimentation, to justify utilizing one approach to attack deep, broad problems over another. We argue that statisticians simply have not felt that this was an important problem, certainly not one worthy of academic research. Hopefully this will change. We suspect that insightful students have already asked the professors teaching Stat 399 these same fundamental questions.

We believe that an underlying theory can be developed to answer these questions, if we recognize that statistics is both an engineering discipline as well as a pure science. So just what is an engineering discipline? Published definitions of engineering (e.g., *www.answers.com/topic/engineering*) are typically some variation of the following:

> Engineering is the study of how to best utilize scientific and mathematical principles for the benefit of humankind.

The statistical engineering discipline would then be the study of how to utilize the principles and techniques of statistical science for the benefit of humankind. From an operational perspective we define statistical engineering as:

> *The study of how to best utilize statistical concepts, methods, and tools, and integrate them with information technology and other relevant sciences, to generate improved results.*

In other words, engineers—statistical or otherwise—do not focus on advancement of the fundamental laws of science, but rather how these laws might best be utilized for societal benefit. This is not to say that engineers do not perform research, or do not develop theory. Rather, it suggests that engineers' theoretical developments tend to be oriented toward the question of how to best utilize known science to benefit society.

Interestingly, mathematics and statistics are perhaps the only disciplines that tend to equate "theory" with "mathematics." Biologists, geologists, and scientists in most other disciplines understand that theory may or may not be mathematical in nature. Madigan and Stuetzle, in their discussion of Lindsay, Kettenring, and Siegmund (2004, p. 409), made this point: "The issues we raise above have nothing to do with the old distinction between applied statistics and theoretical statistics. The traditional viewpoint equates statistical theory with mathematics and thence with intellectual depth and rigor, but this misrepresents the notion of theory. We agree with the viewpoint that David Cox expressed at the 2002 NSF Workshop on the Future of Statistics that 'theory is primarily conceptual,' rather than mathematical."

Pfeifer, Marquardt, and Snee (1988) provided one illustration of what we are referring to as statistical engineering when they suggested that statisticians "ingrain statistical methods in organizational culture," as a means to enhance results. They further proposed, "One strategy is to embody statistical expertise and methods in procedures, processes and or systems in organizational functions that ensure regular and routine use in meeting business needs." Hahn, Doganaksoy, and Hoerl (2000) discussed a case that illustrated this approach. The problem solution used standard statistical methods in a unique application—credit card collections. Statistical thinking and methods (e.g., design of experiments, run charts, Pareto charts, measurement systems analysis, and analysis of variance), along with knowledge-based tools (e.g., flow charting, failure modes and effects analysis, and control plans), were integrated with relevant subject matter knowledge of the collections process. The results were improved collections success rate, institutionalization of the improved approach, and generation of millions of dollars in savings.

In our opinion, the discipline of statistical engineering should be concerned with how to best utilize the laws and principles of statistical science to generate improved results. It should attempt to develop and utilize theory to answer such questions as:

- How should statistical methods be linked and integrated into an overall problem solving methodology, in order to attack deep, broad problems that do not correspond to an identifiable text chapter?
- What are the best approaches to deploying statistical methods broadly across an organization, while avoiding the obvious pitfalls?
- How can modern educational theory be utilized to better structure introductory statistics courses, so that they both are interesting and prove ultimately useful?
- What developments in statistical software would make it more robust to user misuse, that is, to alert people that they are doing something that does not make sense?

Too often, the jugular questions listed above are answered based on gut feel, intuition, or by the opinions of the statistical "gurus" who are currently in vogue. We believe questions such as these should be answered based on rigorous, academic research, and with a sound theoretical basis. As Deming admonished, we need theory to guide us, rather than being completely experiential. To date, we see no such theory that our profession has developed. This is a huge oversight in our opinion, one that contributes to our "bad name in the general scientific community and beyond," as mentioned by Meng.

SOME NEXT STEPS

Meng has provided important leadership for the profession with his bold emphasis on the importance of introductory courses and the need to focus these courses on statistical thinking, and for introducing an academic course on solving "broad, deep problems." Clearly, change is occurring in all arenas of statistics. We strongly suggest that rather than being relegated solely to the role of passive statistical policemen (bad cops), we should demonstrate leadership, showing others how statistics can be used to perform scientific work better and faster (be good cops).

A variety of approaches to do this should be researched and evaluated. But we must be guided by the relevant theory that already exists; we do not have time to guess or reinvent the wheel. Fortunately, a rich literature on statistical thinking and its theory exists today. We can revise and improve this theory, as we obtain more knowledge and feedback on what works. We also need more developed theory for statistical engineering.

Teaching introductory students the concepts of statistical thinking will develop demand for application of statistics to the deep, broad problems. Providing students overall approaches to attacking these problems based on the theory of statistical engineering will lead to more sound applications of statistical methods, and fewer blunders. A conscious focus on developing the theory of statistical engineering will help us to be even more "valued," and hopefully not "feared."

[Received November 2009. Revised December 2009.]

REFERENCES

American Society for Quality (1996), *Glossary and Tables for Statistical Quality Control* (3rd ed.), Milwaukee, WI: Quality Press. [11]

Brown, E., and Kass, R. (2009), "What Is Statistics" (with discussion), *The American Statistician*, 63, 105–123. [10]

Bryce, G. R. (2004), "A Statistical Thinking Approach to Introductory Statistics: Theory and Practice," presented at the *Beyond the Formula Conference, Rochester, NY* (August 2004). [12]

Chance, B. L. (2002), "Components of Statistical Thinking and Implications for Instructional and Assessment," *Journal of Statistical Education*, 10 (3). [11]

Cobb, G. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. A. Steen. *MAA Notes*, Vol. 22, Washington, DC: Mathematical Association of America, pp. 3–34. [11]

Deming, W. E. (1993), *The New Economics—For Industry, Government and Education*, Cambridge, MA: MIT Center for Advanced Engineering Study. [11]

Forester, A. D. (1990), "An Examination of the Parallels Between Deming's Model for Transforming Industry and Current Trends in Education," *Small College Creativity*, 2 (2), 43–66. [12]

Hahn, G. J., Doganaksoy, N., and Hoerl, R. W. (2000), "The Evolution of Six Sigma," *Quality Engineering*, 12 (3), 317–326. [13]

Hoerl, R. W., and Snee, R. D. (2002), *Statistical Thinking—Improving Business Performance*, Pacific Grove, CA: Duxbury. [11,12]

Kume, H. (1985), *Statistical Methods for Quality Improvement*, Tokyo, Japan: The Association for Overseas Technical Scholarship. [12]

Lindsay, B. G., Kettenring, J., and Siegmund, D. O. (2004), "A Report on the Future of Statistics" (with discussion), *Statistical Science*, 19 (3), 387–413. [12]

Meng, X. (2009), "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" *The American Statistician*, 63, 202–210. [10]

Moore, D. S. (1990), "Uncertainty," in *On the Shoulders of Giants*, ed. L. A. Steen, Washington: National Academy Press, pp. 95–173. [11]

Pfeifer, C. G., Marquardt, D. W., and Snee, R. D. (1988), "Statistics—A Time for Change," *CHANCE*, 1 (1), 39–42. [13]

Snee, R. D. (1990), "Statistical Thinking and Its Contribution to Total Quality," *The American Statistician*, 44, 116–121. [11]

——— (1999), "Development and Use of Statistical Thinking: A New Era," *International Statistical Review*, 67, 255–258. [11]

——— (2008), "W. Edwards Deming's "Making a New World": A Holistic Approach to Performance Improvement and the Role of Statistics," *The American Statistician*, 62, 251–255. [11]

Snee, R. D., and Hoerl, R. W. (2003), *Leading Six Sigma: A Step-by-Step Guide Based on Experience With GE and Other Six Sigma Companies*, Upper Saddle River, NJ: Financial Times/Prentice Hall. [12]

Wild, C. J., and Pfannkuch, M. (1999), "Statistical Thinking in Empirical Inquiry" (with discussion), *International Statistical Review*, 67, 223–265. [11]

Thoughts on the Importance of the Undergraduate Statistics Experience to the Discipline's (and Society's) Future

Brian C. Kotz

Xiao-Li Meng's recent article "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" (2009) was of particular interest to me as a former undergraduate statistics major and as an Associate Professor who teaches 12 sections of introductory statistics annually at Montgomery College, a two-year college in Montgomery County, Maryland (an adjoining county to Washington, D.C.). I approach my comments from these perspectives as I believe that these groups very much need to be represented/addressed in the discussion of Meng's observations and proposals. My remarks are also influenced by a *Washington Post* article published during the 2009 Joint Statistical Meetings that referred to statisticians as "superheroes," described some of the challenges we face, and ultimately presented a favorable light (in my opinion) on our discipline.

KEY WORDS: Introductory statistics; Statistical education; Two-year college; Undergraduate education.

1. CONSIDERING THE UNDERGRADUATE MAJORING IN STATISTICS

> "Currently we have far too few good statistical educators and communicators relative to the task at hand and the coming demand" (Meng 2009, p. 204).

In a Harvard class of roughly 1600 undergraduates, I was one of only three students who graduated with a degree in Statistics during the 1989–1990 academic year. While I sometimes felt isolated in my own major, from an undergraduate perspective, I was very well challenged, I received attention and mentoring that students in other disciplines envied, and I certainly could not hide from the rigors of the discipline.

The introductory statistics course I took in my freshman year was great, but it was not a "happy course" in the sense that Meng (2009) described as it was not necessarily designed to make students *happy* to learn statistics. I cannot speak as to whether there has been an increase in the number of statistics concentrators (majors) that graduate from the college each year, but Meng's article does confirm that things have changed considerably in the past 20 years at the college in terms of both undergraduate interest in statistics and the recent "happy" course design that has been successfully implemented. Nonetheless, I suspect that impediments still exist for attracting undergraduates to a statistics major (at Harvard and elsewhere), and I feel that statistics faculty in higher education and others in the discipline must keep these considerations in mind while addressing the issues that Meng presents.

1.1 Some Challenges in Getting the Undergraduate "in the Door" and Beyond Introductory Statistics

For one thing, several colleges still do not offer an undergraduate degree in statistics and most likely do not have the financial resources to begin doing so in the near future. Also, many students (and the general public) still see statistics as a specialized extension of mathematics—a fact that is only reinforced when many introductory (and advanced) statistics courses are only offered through a school's mathematics department (and designated as such in the course catalog, e.g., "*MATH 116—Elements of Statistics*"). Another consideration is student perception of the discipline, particularly when compared to other disciplines that are more well-known, possibly considered to be more lucrative, and/or potentially perceived as more transparent. Bear in mind that the Brown and Kass (2009) article that inspired Meng to write his article is titled "What Is Statistics?" I know firsthand of classmates in the late 1980s (as well as some of my own recent students) who enjoyed their introductory statistics class but had no interest in pursuing further coursework in the discipline for three main reasons:

(1) additional statistics coursework was simply not required for their concentration,
(2) many of the next classes in statistics required calculus (still a "gatekeeper" course and a source of fear for many),
(3) students who met the minimum math requirements for a statistics concentration had several other options as they also met (or well exceeded) the math requirements for chemistry, physics, economics, etc.—disciplines that might appear more appealing and/or viable to a 19-year-old being asked to declare a major.

1.2 Why We Need to Reach Out to the Undergraduates (Statistics Majors and Otherwise) and Secondary School Students

Given that it may be a challenge to develop an undergraduate statistics student (or even get one "in the door"), what can be done once the student is "in the door" with regard to the concerns that Meng described of "far too few good statistical educators and communicators relative to the task at hand"? One tacit assumption is that the undergraduate upon earning his/her statistics degree is appropriately proficient in THE FIVE criteria that Meng put forth: "extensive statistical knowledge; deep understanding of statistical foundations; substantial experience

Brian C. Kotz is Associate Professor of Mathematics, Montgomery College, Germantown, MD 20876 (E-mail: *brian.kotz@montgomerycollege.edu*). He thanks Xiao-Li Meng for his discerning article, John Stufken for inviting contributions for a discussion, and the reviewers of this submission for their encouragement and helpful suggestions.

 DOI: 10.1198/tast.2010.09185

in statistical practice; great communication skills; and profound pedagogical passion" (Meng 2009, p. 208)—or at least the student has a relatively firm foundation and has received some effective nurturing/mentorship in these areas. If so, these students are prime candidates to assist in addressing Meng's concerns both as potential future educators and as potential future stewards of the discipline.

If in fact the supply of "good statistical educators and communicators" is neither meeting demand now nor is expected to meet demand in the future as Meng suggested, statistics departments will have an obligation to inform both their own undergraduates and undergraduates in related disciplines of this scarcity, the opportunities that thus may exist in the future marketplace and in academia, and the support that exists from the statistics community, particularly the ever growing statistics education community. (It may require some deft handling to tell an undergraduate that there is a scarcity of good educators while assuring the student that that is "not the case at our school.") Moreover, if a statistics department believes that any of these undergraduates (statistics majors or otherwise) may not be future statistics doctoral candidates yet still might teach statistics at some point in their careers (future high school Advanced Placement teachers, future statisticians/consultants who will teach in an adjunct capacity at a local college, someone seeking a Ph.D. in another field, etc.), then the department may also have an obligation to offer and to encourage undergraduate coursework/experiences in statistics education and teacher training. (An undergraduate version of the *Stat 303, The Art and Practice of Teaching Statistics* course for Harvard doctoral students that Meng mentioned is one possibility.) Such offerings may also have interdisciplinary potential, for while statistics majors may be a target audience, it is possible that students from related disciplines and/or programs such as Harvard's Undergraduate Teacher Education Program (UTEP) would also be interested in (and benefit from) such experiences. It is also reasonable to assume that other, more direct interdisciplinary efforts at the undergraduate level could similarly assist in the development of successful future statistics educators and communicators, for when statistics departments reach out to undergraduates in diverse disciplines and offer the opportunity of collaboration with statistics faculty, an important secondary opportunity is created to discuss and to demonstrate effective communication and analysis. The "joint explorations with economics, psychology, engineering, life sciences, etc." (Meng 2009, p. 205) that are now in place at Harvard would appear to be good catalysts for such dialogues, and as Brown and Kass pointed out, "the very best quantitatively oriented students often come from other quantitative disciplines" (Brown and Kass 2009, p. 107) and may not be majoring in statistics per se.

Sparking students' interests during their secondary school years may also be critical to developing undergraduate statistics majors and/or the type of educators and communicators that Meng called for. While I agree with Meng that a poorly taught AP Statistics class (or similar class at a four-year or two-year college) may "do more harm than help, permanently turning away many of its students" (Meng 2009, p. 209), I wish that Meng would have similarly and specifically mentioned how a well-taught AP or other introductory college course in statistics may serve as an effective tool for inspiring some students to pursue an undergraduate (or advanced) degree in the field. It warms my heart when I learn of a former AP or undergraduate student who has decided to join the "family business" and pursue a career in the discipline, and I know that many other current and former AP/college statistics teachers have had similar experiences. I am concerned that Meng's anecdote regarding the results of his informal questioning of Harvard undergraduates who consider a concentration other than Statistics may in fact misrepresent the amount of "happy" learning and teaching that is occurring in many high school statistics programs.

All in all, I believe that the statistics community has both a challenge and a responsibility to encourage and to develop the undergraduate student, particularly if some of Meng's goals are to be realized. Furthermore, I believe that the right assistance from other disciplines and the secondary school community may prove extremely helpful if not ultimately necessary. I truly admire the faculty and staff of those universities who have developed thriving undergraduate statistics programs, and I would encourage these institutions to continue to share their expertise both in how they maintain and improve their programs and in how they attract undergraduates to the discipline.

2. WHY I HOPE THAT YOU CARE GREATLY ABOUT THE TWO-YEAR COLLEGE EXPERIENCE

> ". . . it is no longer just about helping others clean up their backyards, but rather about preparing whole generations of future scientists and policy makers" (Meng 2009, p. 203).

When I first read this excerpt, I wondered "Why did Meng stop with scientists and policy makers?" My reaction gets to one of the main reasons I became a professor at a two-year college—I am a member of this community, and so are my students. In addition to the groups Meng listed above, with my work here at Montgomery College, I am most likely teaching someone's future nurse, teacher, social worker, software engineer, doctor, accountant, investment analyst, lawyer, military intelligence analyst, and so on.

I have the opportunity to teach introductory statistics to 300 to 350 students per year, and I feel quite fortunate to be at a two-year college where I get to make a direct impact in the quantitative literacy of so many students. While there are many support resources here for our students beyond the instructor such as our popular Learning Centers, the supplemental materials that accompany the textbooks we use, etc., my contact with our students is direct (no teaching assistants or guest lecturers), and our introductory statistics class may well be the ONLY quantitative analysis course these students EVER take given current degree requirements both here and at most of the colleges to which these students may transfer. To put it more bluntly than Meng stated in his article, as an introductory statistics teacher, if I do not take the time and effort to make sure that the material is presented well, that the learning is facilitated well, and that the assessment is genuine and meaningful, there is a decent chance that I have truly shortchanged and compromised my students (and thus my community) in terms of the critical thinking, decision making, and general reasoning needs of their academic, professional, or even personal lives. And what of the subsequent impact to those students' clients, patients, customers, families, etc.?

Perhaps I am overstating the introductory statistics educator's influence on society (or *importance to* society), but consider the alternatives and consequences if we do not do our job well. Do we in fact have a "heroic" task (or "heroic" *responsibility*) before us? Meng seemed to think so (and I concur) as evidenced by his willingness to place his most qualified teachers in Harvard's introductory statistics courses and to develop the department's "happy courses." The recent General Education discussion at Harvard that Meng (2009) described is just another example of how colleges more and more are valuing (and defining, and debating) the quantitative analysis and decision making skills that are needed by their undergraduates.

Thus I also agree completely with Brown and Kass that "appreciation of statistical thinking should begin in introductory courses" (Brown and Kass 2009, p. 108). To me, that means that a teacher with a mathematics background cannot just skip the chapter about bias in sampling and polling just because he/she is not comfortable in covering the material (e.g., it was not in the statistics course he/she took years ago and he/she has not received any guidance, professional development, or training regarding the material). It also most likely means that a far more substantial writing component needs to be in place for a student's introductory *statistics* course as compared to that student's previous *math* courses. Again, to be blunt, if I do not take care of these things when I have the students in front of me, there is a chance that they will not be exposed to this material, this way of analysis, this way of writing, etc. in their undergraduate experience. Note: this is the best encouragement I can give for the American Statistical Association, American Mathematics Association of Two Year Colleges (AMATYC), Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), Mathematics Association of America (MAA), National Council of Teachers in Mathematics (NCTM), and other groups to continue their many initiatives for training teachers of introductory statistics, for developing guidelines and standards for instruction [see "Guidelines for Assessment and Instruction in Statistics Education" (GAISE) at *www.amstat.org/education/gaise/*], and for providing "actionable intelligence" based on the research of the statistics education community. I urge these groups to please maintain these important efforts, particularly for the faculty of two-year colleges as our enrollments keep climbing. (Last year at Montgomery College alone, over 3000 students enrolled in either *Elements of Statistics* or *Statistics for Business and Economics.*)

And allow me to add just one more reason why I would like you to care greatly about the job that two-year college introductory statistics faculty are doing—our students are coming to a university near you. In 2007, over 4500 Montgomery College students enrolled in 359 institutions in 44 states and the District of Columbia. The list is available at *www.montgomerycollege.edu/Departments/studev/destin2007.pdf*, and I am wagering that a fair number of the readers of this discussion are presently (or have at some point been) affiliated with at least one of the schools on that list.

3. CONCLUSION

As I alluded to earlier, while the 2009 Joint Statistical Meetings were held in Washington, D.C., the *Washington Post* ran an article titled "In D.C., Statisticians Flex Their Strength in Numbers" (*Washington Post*, August 5, 2009). To my surprise, the paper ran the story on the front page of the *Style* section, a section generally reserved for news/reviews of the entertainment industry or for coverage of the various celebrities from the political and entertainment world who drop by or live in the nation's capital. The piece by Monica Hesse contained the following:

> "Real superheroes... skip the capes and tights. Too bulky, too flashy, spandex doesn't breathe well, etc. Which is why they can be easy to miss when they're in town, even when there are 6,000 of them, super-number crunchers, data heroes, with powers of finding meaning in digits far beyond those of mortal men and women... Ladies and gentlemen: statisticians... beacon of hope for a nation of thoroughly confused individuals..."

Within the humor and despite some stereotypes presented, the remainder of Hesse's piece addressed the need for greater statistical literacy, made the point that proper statistical analyses really matter in people's lives, and included the prediction that statistics programs will become "the new hip destination" in the near future. In my opinion, the article also communicated our enthusiasm for our discipline, highlighting both the challenges we face and our willingness to address them. As I read this article during a break at the D.C. meetings, and as I sympathized particularly with the points about the need for greater statistical literacy, I then realized that I was at that moment in the midst of a few thousand similarly sympathetic colleagues—and I began to think that perhaps my feelings regarding the importance of introductory statistics education were well founded. (I also began to wonder if there were some ways to incorporate the superhero idea in my classes, e.g., if there were a statistics superhero, would the archenemy be named "The Outlier"?)

An important moment in my journey as an introductory statistics educator occurred just before my third year of teaching when a colleague encouraged me to read David Moore's (1997) "New Pedagogy and New Content: The Case of Statistics." This thought-provoking article (along with its terrific discussions) affirmed for me that I had joined the profession at an exciting time, but it also led me to wonder just how much of what the author and discussants had proposed, agreed upon, disagreed upon, etc. would really catch on in the next ten years. [Moore ended his article with the statement "Let a thousand flowers bloom" (Moore 1997, p. 136).] Thus, when I read Meng's article, I felt that in some respects I was reading a response to my questions from over a decade ago—a response that made me "happy" in my decision to become a career statistics educator yet also gave me concern about the areas of deficiency and challenge that Meng described.

I also must confess that in some portions of Meng's article, I found myself saying "Why is Meng telling the WORLD about OUR (the discipline's) current issues and shortfalls? Shouldn't we keep this within our own walls?" Not long afterward, in preparation of this piece, I realized that these are not just OUR issues, they are in fact the WORLD'S issues since we as statisticians and educators are in some ways responsible (some would say *greatly* responsible) for helping the WORLD to make its decisions effectively, to meet its challenges, and to effect change where needed. The importance that Meng places on the introductory statistics class—and thus the responsibility that he places on the effective statistics educator—was professionally and personally encouraging to me, but it also led me to wonder if many of the nation's four-year and research universities have the philosophical inclination or economic resources to attract, retain, or reward great introductory statistics educators who willingly postpone or abandon research/other teaching interests (or have little to no interest in research/publishing in the first place) in the name of educating the general undergraduate student.

Ultimately, upon reading Meng's article, I said to myself "there are stewardship issues here"—and as stewards of the discipline, we ALL (not just statistics educators) have a responsibility (perhaps even a large stake) in the resolution of the issues and concerns that he brings forth. For just as we want our students to be effective and engaged *learners* both in our midst and their whole lives through (as they will be our future scientists, policy makers, nurses, teachers, analysts, etc.), we too must be effective and engaged *participants* in the debate, development, and implementation of the experiences that we feel that our students—particularly our *undergraduates*—should have. Meng's article described and demonstrated some of the Harvard Statistics Department's commitment to these pursuits which I am glad to see. I know that many other colleges and universities have already been hard at work in addressing these concerns for some time, and it is my hope that others will take some of Meng's words to heart and help to develop as many great statistical educators and communicators (and "heroes") as their institutions' resources and responsibilities will allow.

[Received September 2009. Revised December 2009.]

REFERENCES

Brown, E., and Kass, R. (2009), "What Is Statistics?" (with discussion), *The American Statistician*, 63, 105–123. [15,17]

Meng, X.-L. (2009), "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" *The American Statistician*, 63, 202–210. [15-17]

Moore, D. (1997), "New Pedagogy and New Content: The Case of Statistics" (with discussion), *International Statistical Review*, 65, 123–165. [17]

Who Is Teaching Introductory Statistics?

Frank P. SOLER

The growing popularity of the statistical sciences has brought about an unprecedented student demand for undergraduate statistics courses, especially courses of an introductory nature. The question of "Who Is Teaching Introductory Statistics?" is at the core of whether over the next 50 years the discipline of statistics would be desired or feared. This commentary addresses compelling issues currently facing the status of statistics education in this nation.

KEY WORDS: Algorithmic; Bias; Critical mass; Desired; Developmental; Feared; Pedagogically; Randomization; Technically; Uncertainty; Variation.

Of all the challenges put forth by Professor Meng (2009) on his article "Desired and Feared—What Do We Do Now and Over the Next 50 Years?", *TAS*, Vol. 63, Number 3, pages 202–210, the one that can potentially cause the most damage is directly related to the availability of faculty who are capable of teaching introductory statistics courses well.

To be sure, this is not a new problem. In fact, over half a century ago, H. Hotelling (1940), in his article "The Teaching of Statistics," *The Annals of Mathematical Statistics*, Volume 11, pages 457–470, addressed a similar set of issues. More recently, others have addressed similar issues (e.g., David Moore and George Cobb). However, the situation has not really improved. Consider the fact that introductory statistics has become the course of choice for students seeking to fulfill the typical quantitative undergraduate requirement. This has caused the enrollment in introductory statistics courses to soar to unprecedented levels.

Even though I don't know enrollment numbers pertaining to the entire nation, I have access to local statistics at the institution where I teach (one of the three largest single-campus community colleges in the state of California with a fall quarter enrollment of about 25K students). Specifically, we offer about 30 sections of introductory statistics per quarter. Including our very popular summer quarter, the numbers are staggering: over 100 sections per academic year serving an enrollment of well over 4000 students. The corresponding teaching assignment is divided among twenty or so faculty (both full and part time) who have indicated an interest to teach such a course. Among these twenty faculty there are only six who have either an undergraduate or graduate degree in statistics. The typical statistical background of the remaining faculty consists of one or two courses, mostly on probability, with little or no significant exposure to statistical methodologies or statistical thinking. Needless to say, there is enormous variation in the type of introductory course a student will get. Most problematic is the fact that our mathematics colleagues treat the teaching of introductory statistics very similarly to the way they treat the teaching of developmental mathematics: heavy emphasis on formulas and algorithmic-driven procedures with little or no emphasis on many of the core components of statistical thinking such as uncertainty, variation, bias, and randomization.

Complicating matters further, public institutions of higher learning in the state of California (and probably in other states as well) operate under fairly restrictive tenure laws impacting the teaching load for both full-time and adjunct faculty. This translates into very carefully tailored teaching assignments where tenure and seniority considerations trump any other criteria. In the absence of a separate Statistics Department, statistics and mathematics faculty fall under the combined title of "mathematics faculty"; thus, tenure and seniority guidelines are applied to the combined faculty. This causes bizarre scheduling situations such as excluding faculty with a graduate degree in statistics from teaching statistics at all!

However, the problems we face are not only bureaucratic and administrative in nature. The question of how to attract talented teachers, competent in statistics, is most compelling. Such tall order starts by recruiting solid students in the mathematical sciences into the field of statistics. Even though the discipline of statistics is still very young, the subject has seen considerable growth and specialization over the last half century. This calls for statisticians to teach statistics at all levels. In fact, Professor Meng makes the argument that the most competent (both technically and pedagogically) statistics faculty should teach introductory statistics courses. (I couldn't be more in agreement.)

Unfortunately, this is not the current reality. That is, mathematicians will be teaching statistics courses, especially introductory statistics courses, for years to come. How can we make that situation better? At the local level, Faculty Associations (tantamount to Teacher Unions) must understand that current seniority and tenure practices are not in the best academic interest of students. At the same time, our mathematics colleagues must be brought to understand, for example, that when we talk about "robust" procedures we are not suddenly engaging in "wine-speak"! There is a lot to be done via in-service and locally organized workshops. A united effort is needed, not only locally but also regionally and nationally. Both of the major mathematical organizations (MAA and AMS) must engage in serious collaboration with ASA in order to address this troublesome situation.

At the high school level the staffing problem is exacerbated. Consider the current status of the Advanced Placement program in statistics. For the academic year 2009–2010 some 140K students are projected to write the AP Stats exam. And, just as many students take a statistics course in high school but choose not to write the AP Stats exam. Needless to say, the overwhelming majority of high school faculty teaching a statistics course have degrees in mathematics but not in statistics.

Frank P. Soler is Professor, Department of Mathematics/Statistics, De Anza College, Cupertino, CA 95014 (E-mail: *solerfrank@deanza.edu*).

However, from my own experience, there are extensive opportunities available for high school teachers to attend in-service activities and comprehensive workshops conducted by leading statistics educators. These are mostly sponsored by The College Board and Educational Testing Services. To my knowledge, this network of professional growth has not been duplicated at the post-secondary level.

In closing, as we approach the conclusion of the first decade of the new century, we also approach the moment of "critical mass" in the teaching of undergraduate statistics and, in particular, the teaching of introductory statistics. Our immediate and careful attention to this problem will go a long way in determining whether statistics will be ultimately desired or feared.

The Business of Desire and Fear

Rick CLEARY and Sam WOOLFORD

We respond to Xiao-Li Meng's provocative article, "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" from the point of view of business education and business practice. We suggest that if statistics departments were to adapt some of the standard features of a business education they might take an important step toward producing graduates who were more effective collaborators in both academia and business.

KEY WORDS: Business education; Communication skills; Graduate education; Statistics education.

We thank Xiao-Li Meng for his deep, entertaining, and thought-provoking article. We also thank editor John Stufken for inviting the statistical community to respond and, more generally, for encouraging this sort of deep discussion of the philosophy and the future of our discipline. As statisticians who teach undergraduate, master's, and Ph.D. students at a university with a business focus, we wish to share our ideas on how some of the excellent points raised in Meng's article, written in the context of academia, translate to modern business practice. Our frequent interactions with the business community, as consultants and mentors to our students, have led us to explore many of the issues Meng raises and to think about the role of statisticians more broadly.

The phrase 'desired and feared' resonates with us on many levels as we see both verbs in action. One might speculate that it is our faculty colleagues, the employers who hire our students, and our consulting clients that are the ones showing the desire for statistics. Likewise one might guess that it is our students who show the fear of rigorous quantitatively focused courses. This scenario does sometimes play out but our experience, like Meng's, shows that the dynamic is often considerably more subtle and complicated. Experienced faculty and successful businesses that have done well without a data-centric approach fear change, but they also have a concern that if they do not adapt to a more quantitative approach they may be left behind. Students who take good statistics courses lose their sense of fear and in fact begin to see the subject as a desirable commodity not just for their careers, but for changing the way they see the world.

There is much about the student development approach of the modern business school that should perhaps be emulated in statistics departments. Admission to competitive MBA programs depends on work experience and career goals as much or more than it does on previous academic achievement as an undergraduate. Putting students on interdisciplinary teams, having them write reports and make presentations, and balancing theory and application are hallmarks of the top business programs for both undergraduates and graduates. We believe that statistics students would benefit from similar broad exposure.

For example, consider the very worthwhile proposal presented in section 3 of Meng's paper, "Supplementing Graduate Curricula With Professional Development Curriculum." This is an outstanding idea, and we applaud the work Meng has done in developing Harvard's innovative courses in professional development. Making sure that graduate students learn how to become good communicators of statistical advice and good representatives of the community is a great first step but we suggest that this does not go far enough. We believe that potential as a teacher, consultant, and collaborator should not just be a supplement to the curriculum, but should be part of the admissions process and then part of the assessment of students in a statistics program. If the ability to communicate effectively is truly important, why base admission of students solely on their ability to handle sophisticated mathematics? Moreover, is it presumptuous of a statistics department to believe that one or two courses that stress communication will make students competent in that area? After all, we as a profession have certainly been reluctant to believe that our colleagues in other fields can make a student statistically competent by teaching one or two research methods courses of their own!

Opportunities in business for those with ability and experience in statistics are likely to remain plentiful even through lean economic times. The recent explosive growth in ERP (enterprise resource planning) and CRM (customer relationship management) systems, among others, has fueled a need to take advantage of the growing quantity of data captured by corporations to create competitive advantages. Even at the top statistics programs, many graduate degree holders leave academia for business, and making sure that these professionals advance our field and help drive good decision making is a worthwhile goal for both business departments and statistics departments. The most successful of these will wind up moving toward management, at which point they are likely to spend most of their time as consumers of statistics, rather than producers of statistics. We have reported on efforts we have undertaken to support improved statistical practice, in both research and application, among business Ph.D. students (see Woolford and Cleary 2009).

Meng suggests that academic statisticians need to learn a client discipline to be effective collaborators, and business statisticians should likewise have an understanding of the area of statistical application. This understanding needs to be deep enough to not only comprehend the business context of the problem but to help statistically unsophisticated business leaders translate their 'business' problems into appropriate statistical frameworks. Then, in a critically important and undervalued step, the results of the statistical analysis must be translated

Richard Cleary (E-mail: rcleary@bentley.edu) and Samuel Woolford are Professors, Bentley University, Mathematical Sciences, 175 Forest Street, Waltham, MA 02452-4705.

 DOI: 10.1198/tast.2010.09263

back into the business context and presented in an appropriate and convincing fashion. This last step is hardest when the solution suggests that the 'conventional wisdom' is wrong and that the results suggest a counterintuitive business decision. This is the business equivalent of Meng's statistician as academic policeman; and it presents similar professional and personal difficulties for the statistician.

It is instructive that Meng's list of five ideal qualifications for teaching introductory statistics courses (extensive statistical knowledge, deep understanding of statistical foundations, substantial experience in statistical practice, great communication skills, and profound pedagogical passion) would make a great starting point for the ideal qualifications for the statistical analyst in business. (We might want to insert "a thorough understanding of the business" in place of the pedagogical passion.) Such analysts are scarce and highly desired by businesses just as statistics faculty with those skills are rare and in high demand. Taking a broader view of what it means to be a statistician, both business schools and statistics departments can help meet that demand.

[Received December 2009. Revised December 2009.]

REFERENCE

Woolford, S., and Cleary, R. J. (2009), "Statistics for Business Researchers," in *JSM Proceedings, Section on Statistical Education*, Alexandria, VA: American Statistical Association, pp. 154–159. [21]

Response to 'Desired and Feared—What Do We Do Now and Over the Next 50 Years' by Xiao-Li Meng

Elart VON COLLANI

Xiao-Li Meng has written a remarkable article that mirrors well the poor situation of statistics and of all of science. It contains the revolutionary statement that science needs help and should be controlled by a police in order to prevent wrong discoveries. My response aims at further clarifying this statement, which seems to be necessary in view of the question 'What Do We Do Now and Over the Next 50 Years?'. The global warming, the financial crisis, and the international terrorism and wars show that something is wrong with the decision making processes in human societies. These deficiencies must be identified and removed. Science plays a key role in decision making, and according to Xiao-Li Meng science is characterized by a fundamental weakness with respect to thinking—or in other words with respect to approaching reality. He states: 'We statisticians, as a police of science, have the fundamental duty of helping others to engage in statistical thinking as a necessary step of scientific enquiry and evidence-based policy formulation.' My response focuses on this statement and contains a proposal that differs from that of Xiao-Li Meng.

KEY WORDS: Logical thinking; Science; Statistical thinking; Stochastic thinking; Stochastics.

1. THE PROBLEM AND ITS SOLUTION

Xiao-Li Meng (2009) stressed the importance of statistical thinking as a 'necessary step of scientific inquiry.' Instead of relying on 'belief,' 'desire,' a 'leap of faith,' or an 'incentive bias,' scientists should be 'instilled with an appropriate amount of statistical thinking and caution.' According to Xiao-Li Meng, the current financial crisis offers a 'horrendously expensive lecture about the critical importance of understanding and assessing uncertainty and risk,' and he concludes that statistics 'is as essential to modern civilization as an election is to a democratic society.' I principally agree with Xiao-Li Meng, but two important points are missing that are necessary for a correct judgment of the present situation, namely:

- a thorough analysis why the current way of thinking of 'astronomers, engineers, geophysicists, psychiatrists, and social scientists' leads to wrong discoveries, and
- a justification that statistical thinking can in fact improve the situation.

The prevailing way of thinking in science is deterministic. Thus, if Xiao-Li Meng is correct, the many wrong discoveries in science are due to deterministic thinking, which prevents understanding and assessing uncertainty and risk. An uncertainty analysis must answer in particular the following two questions:

- What developments can occur in future?
- How likely are the different developments?

The first question implies that in any given situation many different future developments may occur. The second question implies that there is a structure defined on the set of possible developments that specifies how prone to occurrence each development is. Traditionally science models processes by causal relations and is therefore unable to identify the extent and the structure of future variability. This principal inability of traditional science to describe developments realistically has led, leads, and will lead to numerous 'spectacular claims' that 'fade into nothing' as Meng quotes Hanlon. However, before they fade the spectacular claims result in wrong theories and wrong decisions that have brought the system Earth to the brink of a catastrophe.

Presently, many scientists of all branches make attempts to take into account the inherent uncertainty of future developments. Not being qualified for this task by education, training, and thinking, they commit errors, and Xiao-Li Meng referred to the resulting wrong discoveries by 'astronomers, engineers, geophysicists, psychiatrists, and social scientists' as being 'honest mistakes that could easily be avoided or caught if they themselves had been 'instilled' with an appropriate amount of statistical thinking and caution.' To avoid wrong discoveries, an appropriate handling of uncertainty is necessary which must be based on an adequate perception of the relation between past and future. Before turning to 'statistical thinking,' let me introduce 'stochastic thinking' as the basis of an adequate understanding of uncertainty about future developments. Uncertainty is often divided into:

- aleatory uncertainty, and
- epistemic uncertainty.

Aleatory uncertainty refers to the future and is due to randomness that allows different future developments. Epistemic uncertainty refers to the past and is due to the lack of knowledge or ignorance about the initial conditions. Ignorance and randomness are the sources of uncertainty and a meaningful uncertainty assessment must necessarily consider both issues.

Elart von Collani is a retired professor of stochastics at the University of Würzburg, Sanderring 2, D-97070 Würzburg, Germany (E-mail: *collani@mathematik.uni-wuerzburg.de*). During 30 years of research and teaching, he found so many misconceptions and wrong approaches in statistics that he began to develop a new way of handling the uncertainty about future developments, resulting in a set of rules and a unified approach for controlling uncertainty. In order to utilize his findings, he founded the company Stochastikon GmbH (*http://www.stochastikon.com*). He thanks Xiao-Li Meng for his exciting article and his advice of the editor's invitation.

If science should not be based on belief, ignorance about the initial conditions must be represented by the set of all those initial values that cannot be excluded in a given situation. The larger the ignorance, the larger is the corresponding set. A singleton represents complete knowledge. Randomness was quantified about 300 years ago by Jakob Bernoulli, who introduced the concept of the 'probability of a future event' as the degree of certitude of its occurrence. Probability thus quantifies the force of attraction that is exerted on the event by the initial conditions. The totality of probabilities of all future events represents the above introduced structure. Accordingly, an objective science quantifies ignorance by a simple set and randomness by a probability distribution. Stochastic thinking means to think in sets and structures, that is, to link the set representing the past and the sets representing the future by a set of probability distributions. Starting from this, the development of a unified approach called *Bernoulli Stochastics* for dealing with uncertainty (see von Collani 2004a, 2004b, 2008a) is possible without having to rely on 'belief' or 'desire.'

2. STATISTICAL THINKING

Xiao-Li Meng demanded 'statistical thinking,' but did not explicitly specify what it means. Therefore, we turn to Brown and Kass's (2009) article 'What Is Statistics?', where the following explanation of statistical thinking is given:

> What exactly do we mean by this? Different statisticians would use somewhat different words to describe what defines the essential elements of our discipline's approach, but we believe there is general consensus about the substance, which can be stated quite concisely. Statistical thinking uses probabilistic descriptions of variability in (1) inductive reasoning and (2) analysis of procedures for data collection, prediction, and scientific inference.

At first glance, statistical thinking seems to be similar to stochastic thinking as it uses *probabilistic descriptions of variability*. However, looking at the explanation more closely reveals that it does not say what statistical thinking is, but what it uses and for what it is used. Its meaning remains vague and unclear and, as we will show, allows no unique interpretation. The decisive statement refers to 'probabilistic description,' and its meaning should therefore be the key for understanding statistical thinking. To understand 'probabilistic description,' the word 'probability' must be understood as the basis of statistical thinking and indeed of the whole of statistics. Searching for the meaning of probability in statistical textbooks is rather frustrating, as in most cases no meaning can be found. One of the best sources for the interpretation of probability that I know is the *Stanford Encyclopedia of Philosophy*, available on the internet. It contains five main interpretations and a number of secondary ones. (Jakob Bernoulli's interpretation, by the way, is not included. There is the 'propensity interpretation,' which comes close to it, but is not the same.) In statistics, the two most frequently used interpretations are the frequency interpretation and the Bayesian (or subjective) interpretation:

- The frequency interpretation assumes a sequence of repeatable experiments and explains the 'probability of an outcome' as its relative frequency 'in the long run' (infinite frequency interpretation).
- The Bayesian interpretation refers to statements and explains probability as the individual's 'degree of belief' in them.

Both interpretations are not consistent with one another. Moreover, according to the frequency interpretation probability needs an infinite sequence of experiments, which obviously does not exist. In other words, the frequency interpretation does not explain the concept of probability, but leads itself to irresolvable interpretation difficulties. The Bayesian interpretation, on the other hand, approves 'belief' as a means to develop science and thereby puts it close to religion. As a result, the substance of statistics and statistical thinking, namely 'probabilistic descriptions of variability,' has not a unique (and objective) meaning but several (subjective) meanings that are not consistent with one another. There are some immediate consequences:

- The results of any statistical analysis cannot be interpreted in a unique way and therefore often lead to disputes.
- The methods developed in statistics depend heavily on the used probability interpretation. For the same problem there are different methods, yielding different results (with different interpretations).
- In most cases, statistical models violate reality in an obvious way, implying that, in general, it is impossible to judge the relevance of any obtained result.

The ambiguity of statistics yields many different ways of 'statistical thinking,' and adopting statistical thinking instead of deterministic or logical thinking again leads to problems and wrong discoveries. Furthermore, as another consequence of above inconsistencies, many additional uncertainty theories (fuzzy set theory, possibility theory, evidence theory, grey set theory, etc.) have emerged and are increasing the confusion about the appropriate way of thinking.

3. LAW AND POLICE OF SCIENCE

Xiao-Li Meng correctly stated that 'too many false discoveries, misleading information, and misguided policies are direct consequence of mistreating, misunderstanding, and misanalyzing quantitative evidence,' and he advocated a police of science. Quantitative evidence refers to the uncertainty of future developments, and treating, understanding, and analyzing it adequately must be based on an appropriate way of approaching reality. Before police can become active, a consistent law must be established, and complying with the law should ensure that uncertainty is taken into account appropriately.

Models that take uncertainty into account are often called statistical models in contrast to deterministic models. Indeed, statistical models consider uncertainty—however, they often do so in an inappropriate way. Statistics does not include specific rules on how to model uncertainty. Instead, models developed in mathematical probability theory are used without considering that mathematics follows intrinsic mathematical rules that do not take into account the restrictions of reality. For example, uncertainty is quantified by sets and structures, and in view of reality the sets are finite. Nonetheless, a majority of statistical models are based on the set of real numbers and use the normal law as structure. Whether or not these unrealistic assumptions yield misleading results is generally not checked. Because of this and for other reasons, I distinguish between statistical models and stochastic models. (A complete example of a stochastic model for industry is described in von Collani et al. 2008c.)

Choosing statisticians as police is tantamount to selecting an ambiguous law, and applying it will not reduce the number of relevant violations, but only change their nature. I agree with Xiao-Li Meng's wish for the installation of a police of science, but I do not think that statistics is suitable to serve as the corresponding law, and this opinion is backed by the many deficiencies of statistics and statistical eduction listed in Xiao-Li Meng's article.

The above mentioned Bernoulli Stochastics has none of the shortcomings of statistics. All the derived concepts follow reality, no belief or opinion is involved, and it is based on specific rules of how to model uncertainty by considering its two sources, that is, ignorance and randomness. Bernoulli Stochastics meets the requirements for a 'Law of Science': it is a unified approach, its rules are easily checkable, and its methods are transparent and comprehensible.

4. SUMMARY AND CONCLUSIONS

I consider Xiao-Li Meng's article as a milestone in the development of science, as it not only clearly pinpoints the weaknesses of contemporary science, but also describes the shortcomings of statistics. Contrary to him, I do not think, however, that a deeper understanding of the existing foundations of statistics could change anything, as the foundations themselves are ambiguous and contain misconceptions. A remedy must necessarily begin with removing the ambiguity of the most fundamental concept, that is, the concept of probability. General science, on the other hand, should accept that the deterministic approach must be abandoned as it leads to wrong discoveries and decisions. Abandoning the deterministic approach would result in a completely different science that would enable description of reality as we experience it. This fundamental change of science should be the task for the next 50 years. If this cannot be achieved, mankind will not be able to master the problems that have emerged due to the wrong way of thinking in science. Let me go back briefly to the three ways of thinking dealt with in my response:

1. There is the prevailing way of thinking in science that can be described as 'logical thinking.' It is appropriate for the development of mathematics, that is, the language based on numbers and logic, where the numbers assure a unique meaning, and logic assures consistency. However, reality does not follow logical rules. Mathematics should therefore be used in science exclusively as a language, but not as a way of thinking.
2. When in some applications it became obvious that the results obtained by logical thinking are useless, statistical thinking emerged. It is essentially based on mathematical probability theory and does not explain uncertainty clearly. Statistics did also not try to question logical thinking and science, until Xiao-Li Meng claimed that statistical thinking is a necessary condition to treat, understand, and analyze quantitative evidence.
3. Jakob Bernoulli recognized that a new way of thinking is necessary to cope with the problems of mankind. He introduced stochastics as the 'science of prediction' and quantified randomness by probability, using mathematics only as a means for recording and communicating quantitative evidence.

'Modern' science was initiated by Galileo Galilei and replaced scholasticism that was based on the 'Truth of the Holy Scripture.' When this truth proved to be inconsistent with reality, Galilei postulated that God designed the universe using mathematics, and that the divine truth could therefore only be discovered by mathematical (logical) derivation and not by empirical evidence. Galilei's postulate is still decisive for science (see von Collani 2008b) and prevents real knowledge, and it is time to clean up the 'backyard' as Xiao-Li Meng would put it.

[Received September 2009. Revised November 2009.]

REFERENCES

Brown, E., and Kass, R. (2009), "What Is Statistics?" (with discussion), *The American Statistician*, 63, 105–123. [24]

Meng, X.-L. (2009), "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" *The American Statistician*, 63, 202–210. [23]

von Collani, E. (2004a), "Theoretical Stochastics," in *Defining the Science of Stochastics*, ed. E. von Collani, Lemgo, Germany: Heldermann Verlag, pp. 147–174. [24]

——— (2004b), "Empirical Stochastics," in *Defining the Science of Stochastics*, ed. E. von Collani, Lemgo, Germany: Heldermann Verlag, pp. 175–213. [24]

——— (2008a), "Defining and Modeling Uncertainty," *Journal of Uncertain Systems*, 2, 202–211. [24]

——— (2008b), "Science—the Great Illusion, Stochastics—the Promising Alternative," available at *http://knol.google.com/k/elart-von-collani/science-the-great-illusion-stochastics/yn3zct4004h4/3*. [25]

von Collani, E., Binder, A., Sans, W., Heitmann, W. A., and Al-Ghazali, K. (2008c), "Design Load Definition by LEXPOL," *Wind Energy*, 11, 637–653. [25]

Rejoinder: Better Training, Deeper Thinking, and More Policing

Xiao-Li MENG

REPRESENTING AN ENGAGED POPULATION?

Perhaps due to the somewhat unusual nature of my piece—a discussion of Brown and Kass (2009) that ended up longer than the article itself—the Associate Editor (AE) who handled it had an unusual idea: inviting the general public to react to it. The AE's motivation is clear from his/her editorial comments:

> "My thought here is that we too often turn to the "usual suspects" to get reaction to such manuscripts, yet this is an issue that touches all of statistics and all statisticians. It would be interesting to get the perspective of the broader readership on the issues raised by Meng (and potentially Brown and Kass as well, if we wanted to open up that for general discussion). If no comment is elicited by the article, that perhaps says something as well (I'm not sure what)! I think that proceeding in this fashion would potentially open up a forum for a more wide-ranging discussion."

I was intrigued, and particularly liked the idea of testing what reactions (if any) would be generated without any targeted invitation, from a truly self-selected sample! As statisticians we worry deeply—and rightly—about biases in any self-selected samples, but here one could argue that the seven sets of discussants are a reasonable sample of the population of the "engaged participants," as Kotz characterized them. [If this characterization offends you ("I didn't have time to write because I was busy teaching!"), then you are in this population by definition!]

The AE's prediction of "a more wide-ranging discussion" is also accurate. Government, business, industry, and academia are represented by the discussants; so are North America, Europe, and Australia. The representation also contains deeper stratifications: two-year colleges and universities, nonprofit and for profit, on duty and retired, West Coast and East Coast, etc. Even the writing styles cover a whole spectrum, from humorous storytelling to almost a DoW (Declaration of War)! It is indeed quite remarkable that merely seven discussions can have such a broad and deep representation (but of course no claim on *proportional representation*)!

Any author should be grateful for such wide-ranging reactions, even if most of them are disagreements or criticisms (not the case here!). My heartfelt thanks also go to the AE and the Editor, John Stufken, for providing the forum. In addition, John needs to be thanked for gently reminding me not to repeat history by making my rejoinder longer than the discussions. This freed me from trying to have 95% coverage, but instead to focus on 10% tails in either direction. This rejoinder therefore contains mainly stories inspired by discussants' excellent questions, points that received too little attention in my original piece, and responses to discussions that I need to pour myself a glass over because I have been given too much or too little credit. The responses are organized along the three main themes of my piece as highlighted by the discussants.

BETTER TRAINING

Kotz's and Soler's discussions should make us appreciate more the AE's creative idea. We all have seen many discussions over the years about statistical education, from K-12 to Ph.D. programs. Whereas technically two-year colleges have been included in these discussions, some issues Kotz and Soler raised are completely new to me, and I suspect also to many of AE's "usual suspects." For example, the "bizarre scheduling" situation Soler reported is not something most (any?) of us in statistics departments have thought about, yet now I can see how frustrating, complicated, and serious the matter is. Kotz is right that we all should care more about what goes on in two-year colleges, because such issues directly affect our entire profession; I was quite taken aback by the sheer number of students taking statistics merely in Kotz's and Soler's colleges, a combined annual total over 7000! (Speaking of numbers, this year alone we have 15 undergraduate students declare statistics as their concentration (major), to answer a question of Kotz.)

I also cannot have said better than Kotz's two "blunt" statements about the responsibilities on our shoulders, which remind me of a story from a statistician who joined a large pharmaceutical company after years of being a professor. His first task was to analyze a set of pre-clinical data. He told me that the night before his presentation, which he was told would *determine* whether the company should launch an estimated 30-million-dollar clinical trial, he literally felt sick to his stomach: "I was really scared; I had never felt this much responsibility!" I echoed that I could easily imagine how I'd have felt if I had been in his shoes. Retrospectively, however, I have been asking myself: have any of us ever felt sick to our stomachs the night before teaching because of the thoughts about the responsibility of training future generations, which surely should be heavier than any 30-million-dollar study? Of course most of us have not (at least I have not) for a very simple reason: the impact/outcome of our teaching is not immediately tangible or even measurable—I am sure many of us would have if we were told that tomorrow's lecture would *determine* 30 students' career choices. But this very fact should also remind us of the immensity and longevity of our impact through teaching and hence increase our sense of great responsibility. Perhaps it is not inappropriate to further intensify our "pedagogical sensation" by borrowing a phrase about passionate love—especially given that effective teaching also requires passion and love: it can be felt years after the sunrise. . . .

Passion-driven statistics is indeed the central theme of Easterling's humorous piece—I almost wanted to negotiate for a

Xiao-Li Meng is the Whipple V. N. Jones Professor and Chair of Statistics, Department of Statistics, Harvard University 1 Oxford Street, Cambridge, MA 02138 (E-mail: *meng@stat.harvard.edu*). I thank my colleague Joe Blitzstein and my happy team members (Yves Chretien, Paul Edlefsen, Kari Lock, Cassandra Wolos Pattanayak) for many insightful comments and suggestions.

new car so I could show my dealer what I am made of (but I won't tell him my shoe size)! Easterling is entirely correct that the current generations have a chance (and responsibility) to bring the passion to a new high, and our collective effort can start as simply as better utilizing existing textbooks; as Fox puts it, let's "get it going" rather than "get it perfect." Cleary and Woolford brought in another starting point that was only alluded to in my piece: better training should start with better admissions/recruitment—training with passion from the outset, whenever possible, is obviously more effective than passion injected afterward. Their excellent point on not repeating what we hope others won't do (i.e., equating one to two courses with competence) suggested that the verb in "supplementing graduate curricula with Professional Development Curriculum" should eventually be replaced by "integrating." This will take time, but its ultimate reality is an important assurance for Fox's prediction: "our profession has an incredibly bright future."

Fox also asked an excellent question: what are the minimum standards and competencies for persons deemed suitable to teach statistics? In my original piece, I answered a much easier question: what are the ideal qualifications? Fox's question currently has no enforceable answer. And *that's the problem* (*illa difficultas est*?), to answer Fox's "*Quo Vadis or Quid Agis*?" Surely any minimum standards should include having taken X courses in statistics with $X > 0$, right? As Mark Twain (or Will Rogers) was alleged to have said: "You can't no more teach what you ain't learned than you can come from where you ain't been." Well, even $X = 1$ would disqualify a good number of Soler's colleagues. And that is only for a single two-year college. Thinking about all the two-year colleges, four-year colleges, and AP Statistics, my stomach is now indeed turning.... (I know I am generalizing from $n = 1$, but I have a strong prior!)

Easterling worried about students getting turned away by bad teaching, especially the sharp students. I share that concern, as I detailed in my response (Meng 2009, Part I; 2010, Part II) to Rossman et al. (2009) regarding my observations that some Harvard undergraduates had been turned off by poorly taught AP Statistics, a point on which Kotz also commented. I sincerely hope that Kotz is right that the unintended perception my "Harvard observations" might generate is indeed a misrepresentation, but as I argued in my two-part response, any scientific assessment of the real impact of an educational program needs to study both the "turned-on" population, as Rossman et al. (2009) and Kotz reported, and the "turned-off" population, as I encountered. My two-part responses therefore include a suggestion to ASA to conduct assessment studies, which can help to assess whether the "Harvard observations" are merely local anecdotes or an indication of something far more worrisome.

DEEPER THINKING

This is another point that generates no disagreement, although von Collani asked that "statistical thinking" be replaced by "stochastic thinking," a concept whose meaning I yet need to find out. All discussions below, therefore, still center on statistical thinking as I understand it.

Hoerl and Snee correctly emphasized that statistical thinking should be coupled tightly with statistical engineering, a notion that was not discussed in my article but was advocated by John Tukey (if anyone can locate a specific quote, please let me know). A key component of this coupling, as I see it, is *efficiency*, a critical element that I wish the ASQ's definition of *statistical thinking*, as Hoerl and Snee quoted, had recognized, in addition to *process*, *variation*, and *data*. The maturity of a scientific discipline is measured not only by its accumulated coverage but also—and arguably more critically—by its demonstrated ability to establish *limits*, that is, the optimality and impossibility given constraints. I therefore like Hoerl and Snee's repeated emphases on how statistical engineering—like any other engineering—is about how to "best utilize" concepts, principles, theory, methods, etc. It is this adverb *best* that separates professionals from amateurs, and it is the quest for doing the best given the practical constraints that requires deeper thinking. Most people do not need to take a course in experimental design in order to try out "one-factor-at-a-time" (unless, of course, you are Easterling's unfortunate Sandia colleague!). But to be able to design optimal or even just cost-saving experimental designs given a variety of real-life constraints requires far deeper understanding of the principles of statistical experiments and modeling than most people are naturally equipped with; this kind of ability can have high societal impact, but it can be acquired only via a good dosage of interweaving statistical thinking and statistical engineering, to echo Hoerl and Snee's key point.

A local example illustrates well the importance of understanding optimality/impossibility in defining one's professional identity and hence being desired. My CS (computer science) colleagues here have been teaching all sorts of wonderful algorithms and programming for computing least-squares solutions and alike. However, they found themselves unable to explain satisfactorily the statistical models and principles underlying these solutions, nor could they answer seemingly simple questions such as "Why take squares?" I was thus invited last year to provide a guest lecture to one of their introductory courses. The 90-minute lecture was fully packed, proceeding from Gauss and Galton to the meaning of statistical models to the concept and wonder of MLE. The punch-line that the least-squares estimator is the MLE under the normal model, something we statisticians all take for granted, was an eye opener to both the students and my CS colleagues. It is particularly intriguing to them that once the normal assumption is made (an assumption few of them ever questioned), "taking squares" is the best one can do—as one of them told me: "This is really cool—I've got to look into this MLE thing!" Perhaps the best indication that the lecture got CS students' attention was the course evaluation comment, "you guys teach CS really well, but you should really leave statistics to statisticians," as one of the course instructors relayed to me in the following semester.

If you are thinking that I am using this example to show off how statisticians think more deeply than computer scientists, then bear with me for the other half of the story. Because of this guest lecture, I sat through the one immediately preceding it. It was equally an eye opener to me! Just as statisticians are well-versed in the limits of inference and the like, it is my CS colleagues' cup of tea to tell what is possible and impossible with algorithms and programming, among others. The lecture taught me that it is impossible to have an algorithm/program that can debug every other program correctly. Whereas logically it might not be hard to suspect such an "almighty" algorithm cannot exist, what demonstrates well the deep thinking by computer scientists is their ability to identify problems that seemingly have no connection whatsoever but in fact are equivalent to the impossible debugging problem. And hence they can immediately tell any amateur, "don't even try!" just as we statisticians can tell CS students not to waste their time trying to beat MLE asymptotically.

Hoerl and Snee also asked about what approaches are taught in Harvard's Stat 399 about attacking deep, broad problems that require more than one technique to solve. As I mentioned in my piece, the course was a result of responding to students' request that we help them to better prepare for Ph.D. qualifying examinations. Over the years our qualifying examination format has changed considerably, but one theme has remained—the problems are *not* designed around a set of textbooks or courses; rather, they come out of faculty members' research project problems or problems that teach deep thinking in statistics, such as applying the principle of bias-variance trade-off to investigate what is possible and what is not possible. That is, the problems are often multipart "nano research projects," mimicking their real-life counterparts yet doable in an examination setting. Such examination formats provide a forum for an intensified dialog between students and faculty, before, during, and after the examination. See the report by Blitzstein and Meng (2010) for detailed examples and discussions of the usefulness of "nano research projects." It is also worth emphasizing that the ultimate goal of repeatedly using real-life problems, as in Stat 399 and Stat 105, is not just to showcase the ubiquity of statistics, but more importantly—as Hoerl and Snee also emphasized via the cited Bryce's course—to demonstrate how statistics operates as a scientific discipline with a set of core principles, theories, and methods that can be applied to address an exceedingly wide range of problems.

MORE POLICING

This point is more debated, as several discussants expressed concerns about whether the label "police" would carry a passive image that we only react when someone does something wrong. Retrospectively I wish I had chosen a term that would not conjure such an image, because the whole message of my piece is how we can be more active than reactive, moving from everyone's back yard to the front yard and even living room. Perhaps it is my Chinglish, not understanding well all the connotations of the term "police." When I wrote that I am proud to be labeled as a "statistical police," what I had in mind was "We serve and protect"—a slogan seen on every police car in Chicago (where I spent 10 years)—we provide service to others and we protect them from mistakes.

I of course agree with Hoerl and Snee's "good cop" role, which is similar to Fox's "embedding" approach to work from within. Again much of my piece is about how to provide better quantitative training for future generations for other disciplines, which aims at helping others to move faster on their endeavors in the first place. However, there is either an apparent contradiction or a troublesome implication in the following statement of Hoerl and Snee's: "Meng rightly points out that statisticians can play a useful role in society by limiting the claims made by other scientists based on faulty statistical studies. He refers to this function as playing the "statistical policeman" role. We call this playing the "bad cop" role, in that bad cops fundamentally slow down the research of other disciplines." I do not see how slowing down the research of other disciplines is "a useful role in society," nor how avoiding bad/faulty statistics would slow down research. Isn't the whole purpose of avoiding or stopping mistakes, statistical or otherwise, to speed up real research progress? If someone can move his/her research faster by using *bad/faulty* statistical studies, would that logically imply that the *good/sound* statistical studies are actually antiscientific?

Hoerl and Snee could not possibly have meant what their sentences appear to imply, just as my being proud as a "statistical police" could not possibly mean that I am proud of being a "bad cop," as Hoerl and Snee characterized it. I surmise that what Hoerl and Snee really had in mind was that we should avoid making others feel that we are only interested in criticizing them, not helping them. This message I certainly agree with. We should be strategic in delivering the "bad news" so that we *consult*, not *insult*; and this is where effective communication skill plays a critical role. But this does not mean that we should avoid our "policing role" (though I certainly want to avoid all the negative connotations of "policing" if a better term can be found!). As some readers may have noticed, I tend to put significantly more emphasis on things I believe to require encouragement than on things that already come with good incentives. As I discussed in my original piece, "policing" is typically a thankless and creditless job at the individual level. But at our professional level, I believe it is a part of our identity that will remain unique to us even if "other disciplines have been seizing opportunities" away from us, precisely because we carry out the role for our discipline's integrity, as a critical part of general scientific integrity. Putting it differently, if someone is able and willing to carry out this role on a routine basis, I will have no trouble in considering him/her my fellow statistician. And in that role we sometimes do need to stop someone, not in his/her research, but in the potential harm s/he can do to others and indeed to an entire field. (Incidentally, the article "Fatal Flaws in Cancer Research" in the most recent issue of *IMS Bulletin* (2010, January, page 5) demonstrates vividly how faulty statistics can do harm to our society and how good "policing/forensic" work can stop it.)

A good example is in the literature of climate change, where decades of efforts have been made to understand and interpret apparent oscillations in running correlations among different climate time series/indices, often with conclusions that they represent some fundamental underlying climate dynamics in Mother Nature. However, Gershunov, Schneider, and Barnett (2001) demonstrated via simple simulations that such oscillation phenomena exist even if the two time series are completely independent white noises! As a part of our statistical thinking (and here it is not even a very deep one), any reasonably trained statistician would be concerned with the potential artifacts introduced by the *overlapping moving windows* used for computing the running correlations in the first place. Decades of efforts have literally been misled, but to make the matter worse, when an entire field is on a wrong track for a long time, it often would take an even longer period to put it back on the right track. No one likes to be told that he/she has wasted his/her (professional) life, so the force to defend the established answer or to at least find ways to "save some face" is often very strong (see Robinson, de la Pena, and Kushnir 2008 for a brief summary of the history of this debate). Of course the truth always prevails (let's hope!) and strong arguments, even or especially the wrong ones, could help us to think more deeply. Nevertheless, the real scientific progress in such cases is clearly delayed, not because policing or self-policing duties were carried out too soon, but rather because they were carried out too late.

It is also worthwhile to emphasize that the impact of such "policing work" can go far beyond what is originally intended. For example, Gershunov, Schneider, and Barnett (2001) work has apparently also influenced researchers in solar physics, where running correlations and alike are used to measure certain solar activities. In particular, Elias and de Artigas (2008) provided a detailed account of how a "spurious quasi-biennial cycle" induced by running correlations may be similar to the reported QBO (quasi-biennial oscillations) of the stratospheric equatorial zonal winds, parallel to Gershunov, Schneider, and Barnett (2001) findings; also see Elias and de Artigas (2006). I am especially delighted to see that Elias and de Artigas (2008) was featured as the leading Expert Commentary in the book on *Solar Physics Research Trends* (ed. Wang, 2008), and its abstract ends with what I consider as a good example of self-policing: "The results shown here do not rule out a physical origin, but point out that a result obtained after a statistical analysis carries, in addition to the physics behind, the spurious byproducts of the method applied."

Compared to Hoerl and Snee, von Collani gave me too much credit. Von Collani labeled my article as a "milestone in the development of science," and credited me as a revolutionary in policing science. While flattered, I must confess that I am at least a mile away in seeing the pictures von Collani is painting, and I am not sure if I would make a half turn or full turn in my grave if someone puts "Chief of Science Police" on my tombstone. Von Collani apparently is questioning the entirety of modern science and statistics and wants to replace everything by "stochastic thinking," a discussion topic that is the furthest from my original piece, certainly beyond my reflection antenna. But my statistical thinking compels me to express skepticism of just about any claim of one size fitting all, especially when it comes to matters as complex and grand as science and statistics.

THE POWER OF COLLECTIVE WISDOM AND ACTION

The experience starting from reading Brown and Kass (2009) to preparing this rejoinder reminded me once more of the power of collective wisdom. No matter how thoughtful, articulate, and well-intentioned each of us is, our individual contributions are inevitably idiosyncratic and can even carry ironies that are obvious to everyone but ourselves. For example, I almost choked on the great wine I was enjoying with the discussions when I read Kotz's question "Why did Meng stop with scientists and policy makers?" Indeed! Although I probably would not go as far as to include Easterling's car dealers and shoe salesmen, how could I forget to include "whole generations of future teachers" in an article that is mostly about teaching future generations?

This brings me to my concluding point, the same as the one in my original piece, and on another historic occasion, the first days of the new decade. I share Fox's and others' enthusiasm and optimism that our future is very bright, but to ensure that such enthusiasm and optimism will be carried over to future generations, we need collective action. So please do anything you can to help build the "statistical leadership" that Hoerl and Snee articulated: one lecture at a time, one speech at a time, one consultation at a time, and one publication at a time.

And a toast to the new decade: may we all have that sick-to-our-stomach feeling at least once before lecturing/speaking/consulting/publishing!

[Received January 2010. Revised January 2010.]

REFERENCES

Blitzstein, J., and Meng, X.-L. (2010), "Nano-Project Qualifying Exams: An Intensified Dialogue Between Students and Faculty," technical report, Harvard Statistics. *The American Statistician*, to appear. [28]

Elias, A. G., and de Artigas, M. Z. (2006), "The Quasi-Decadal Modulation of Running Correlations Involving the QBO," *Journal of Atmospheric and Solar-Terrestrial Physics*, 68, 1980–1986. [29]

——— (2008), "The Quasi-Biennial Oscillation in Time Series of Solar Activity Parameters," in *Solar Physics Research Trends*, ed., P. Wang, Hauppauge, NY: Nova Science Publishers, pp. 3–13. [29]

Gershunov, A., Schneider, N., and Barnett, T. (2001), "Low Frequency Modulation of the ENSO-Indian Monsoon Rainfall Relationship: Signal or Noise," *Journal of Climate*, 14, 2486–2492. [28,29]

Meng, X.-L. (2009), "AP Statistics: Passion, Paradox, and Pressure (Part I)," *Amstat News*, December, 7–10. [27]

——— (2010), "AP Statistics: Passion, Paradox, and Pressure (Part II)," *Amstat News*, January, 5–9. [27]

Robinson, L. F., de la Pena, V. H., and Kushnir, Y. (2008), "Detecting Shifts in Correlation and Variability With Applications to ENSO-Monsoon Rainfall Relationships," *Theoretical and Applied Climatology*, 94, 215–224. [29]

Rossman, A., Peck, R., Franklin, C., Hartlaub, B., and Scheaffer, R. (2009), "Letter to Editor," *Amstat News*, November, 5. [27]