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.