Interview with George Cobb

Allan Rossman
California Polytechnic State University

George Cobb
Mount Holyoke College

Journal of Statistics Education Volume 23, Number 1 (2015),
www.amstat.org/publications/jse/v23n1/rossmanint.pdf

Interview with George Cobb

George Cobb is Professor Emeritus of Mathematics and Statistics at Mount Holyoke College. He is a Fellow of the American Statistical Association and a recipient of ASA's Founders Award. He received the USCOTS Lifetime Achievement Award in 2005. The following interview took place via email on December 30, 2014 – February 17, 2015.

Beginnings

AR: Thanks very much, George, for agreeing to be interviewed for the Journal of Statistics Education. *You wrote an article for the very first issue of* JSE *in 1993, and I'm very happy to be interviewing you for this issue. Let me ask you to go back further than 1993. Where were you when you were 18 years old, and what were your career plans at that point?*

GC: When I turned 18 in the spring of 1965 I was nearing the end of my first year at Dartmouth. The previous quarter I had completed the honors section of vector calculus, reputed to be the hardest course in the undergraduate curriculum. At one point in my life I might have considered that a big deal, but Dartmouth had put me in my place. There were at least two dozen other members of the entering class who completed that same course along with me.

I'd had two extraordinary high school math teachers. Imagine a *ninth grade* algebra course in 1959 rural North Carolina: a course that proved the algebraic properties of the reals starting from the field axioms, much like a college course in abstract algebra. Imagine also a *high school* calculus course in 1963 that taught such topological properties as Bolzano-Wierstrass and Heine-Borel. Those inspiring courses taught by Kenneth Walker (Guilford, NC) and Grant Fraser (George School, PA) led me to give up earlier career ambitions as herpetologist, as center for Green Bay, as surgeon, and as chemist, in favor of math.

I chose Dartmouth because John Kemeny and Laurie Snell had transformed the math department there and earned it a feature article in *Time* magazine. As that first year at Dartmouth ended, I still expected to become an academic mathematician. Little did I know that I was about to take a big detour.

AR: Maybe you should have become a novelist, because you've already created suspense with your first paragraphs. I'm sure that other JSE *readers will join me at this point in thinking that the detour involves statistics, but we could be wrong. What was this detour, and where did it lead you?*

GC: The detour was into Russian literature. Of course a change in career plans never feels like a detour at the time. Although I was eventually to find my way to statistics, that was still several years in the future. By the time I got my degree from Dartmouth, I had been accepted into graduate programs in Russian, not math or statistics, and I was expecting to spend my professional life teaching Russian literature. I had continued to love math and take math courses every quarter, but I also took Russian every quarter and spent a summer after my junior year in what was then Leningrad studying the language. My senior year I had my first encounter with data analysis, looking for rhythmic patterns in Pushkin's 5500 line novel-in-verse Eugene Onegin, but the data analysis was secondary. I was mainly interested in the poetry. It took the Vietnam War to convert me to statistics.

AR: Before I ask about your conversion, let me first ask what appealed to you about Russian literature, poetry in particular?

GC: Thanks, Allan. Your question got me to stop and think about issues that I should have explored at the time. I'm guessing that anyone who has read Tolstoy or Dostoevsky can understand why I liked the novels. An additional appeal for me was that at the time Dartmouth had perhaps 40 math majors in each graduating class, but there were only two or three Russian majors. I liked it that I could walk into the department and be recognized right away. As to the poetry, I now think it was the appeal of learning how to play with words. For my data analysis project, I had to key-punch the entire poem onto those old 80-column Hollerith cards, mentally transliterating from the Cyrillic alphabet to equivalent sounds using the English alphabet. I got good at the transliteration, but the key-punching was a very slow process, because I was the worst kind of hunt-and-peck typist, so there was plenty of time while my fingers were searching to locate the right keys while my otherwise-bored brain could be thinking about the sounds and the meaning. And of course there's no way to erase a hole in a punch card, so I ended up rereading and retyping many, many lines. Rereading is the best way to appreciate poetry.

AR: While we're on this detour from your life's detour, I'll ask: Have you continued to read and enjoy Russian literature over the years?

GC: Sadly, my language skills have deteriorated, but happily, my tastes have broadened.

AR: Now back to your student days. You had applied and been accepted to graduate programs in Russian literature. Did you ever enroll in one of these programs? Or did you veer into studying statistics at this point?

GC: Neither one, actually. I graduated in 1968, at the height of the Vietnam War. I came from a Quaker family, with a tradition of pacifism, and my draft board granted my application to be a Conscientious Objector, which meant that I would do two years of alternate service before I would be allowed to go to graduate school. I approve of the idea that everyone should do two years of national service, but I began to get impatient as I sat around for several months waiting to find out what my assignment would be. Then a letter arrived, to the effect that if I wanted to end the uncertainty, I had better find my own assignment. It could be anything that served the national interest, provided it didn't pay as much as a regular job. Long story short, I ended up programming computers in the Department of Biometry at the Medical College of Virginia. The department provided research support for doctors doing medical research – that was the national interest part – and the take-home pay was barely in four figures, $1800 per year – that took care of the other criterion.

AR: Very interesting. So your computer programming experience was another detour, after your turn from theoretical mathematics to Russian literature. Did this work related to medical research point you toward statistics?

GC: It definitely started the process, by giving me time to think about what life might be like if I got a PhD and academic job in Russian. For me, the attraction had always been the poetry and the novels, but I probably wouldn't get to teach that stuff very often. Mostly I'd spend my life teaching first and second year language. I realized I just didn't have the right personality to make that fun for my students or for me. I wasn't yet attracted to statistics, but I was part way there, back to wanting to do and teach mathematics. It was the Biometry faculty, James Kilpatrick, Ray Myers, Roger Flora, and Hans Carter who convinced me to give stat a try. They were very crafty about it: If I got my degree in statistics, I could still do as much math as I wanted, but it would be much easier to get a job.

AR: That crafty advice might be as relevant today as then. How did you go about selecting a graduate program to pursue?

GC: That was easy. I was newly married and my wife Cheryl wanted to go back to graduate school at New England Conservatory in Boston. I'd already been accepted into the PhD program in Russian at Harvard, and I knew that Fred Mosteller was in the stat department there, so I wrote and asked if I could switch my program from Russian to statistics. Meanwhile, the Biometry faculty had decided I might be more useful if I actually knew a little bit of statistics, so they gave me time off from work to take courses in their new program. By the time I completed my alternative service I was a semester away from a masters in Biometry, so I stayed on to finish the degree.

AR: What did you find most appealing about statistics at Harvard? Did your interests move from mathematics to statistics at that point, or were you mostly interested in the mathematical aspects of statistics?

GC: There was a lot to like. The department was a small one where it was easy to get to know the faculty. My first year I was able to take a seminar with Fred Mosteller, and was lucky also to be one of his teaching assistants for the introductory course. I got to take experimental design from William Cochran. Art Dempster was my teacher for linear models and for multivariate analysis. All that in my first year! I really liked the way Art used elegant mathematics to make statistical thinking simpler, I also liked his work on foundations of statistics, and I ended up doing my dissertation with him.

Another benefit to the small size of the department was that we had just a handful of students in each entering class. We were all taking the same courses, and we worked together in the little department library in Palfrey house back in those days before Harvard built its new science center. Nan Laird started the same year I did and was part of our little group in the library; Persi Diaconis was a half-year ahead of me.

As to your question about my interests, I remember a conversation with Allan Donner, who was a year ahead of me. Allan asked me if I was more interested in mathematics or applications, and without hesitation I answered “mathematics.” At this point I was thinking ahead to a dissertation, and Allan passed along some advice he got from William Cochran, his adviser: You make a bigger contribution if you find an approximate solution to a problem that has not been solved than if you find a way to polish an existing solution found by someone else.

Career at Mount Holyoke College

AR: That sounds like a very exciting time and place to have studied statistics. But then one’s time in graduate school (almost) always comes to an end. What kind of positions were you looking for upon your graduation from Harvard? What were your career aspirations at that point?

GC: Whether math or Russian or statistics, I’d known for a long time that I wanted to teach at the college level, so as I began looking for jobs, I was looking for positions in college math departments. Of course I asked Art Dempster to write for me. I also asked Fred Mosteller, even though it had been a couple of years since I had worked with him. To my amazement – and this is so typical of who Fred was – he said that he needed to get a better sense of the kind of job I was looking for, and so he took me to lunch in the law school cafeteria, and spent an hour talking with me about my interests and goals.

My wife Cheryl was job-hunting at the same time – teaching voice – and we hoped to find jobs that were reasonably close geographically.

AR: Did you go directly to Mount Holyoke from Harvard?

GC: I did. I was so naïve that I didn't even realize that the job wasn't tenure track. It was close enough to Boston, where Cheryl could continue her singing career, and I liked the faculty in the department. As it turned out, I had made a good choice.

AR: What did you teach at Mount Holyoke at the start of your career?

GC: Mostly calculus. Intro stat barely existed. At the time I joined the Mount Holyoke faculty, I think that nationwide there were only three other statisticians at liberal arts colleges. Very few colleges had anything like today's intro stat course, except perhaps in departments of psychology or economics. In the math department, Mount Holyoke had a sophomore-level course in probability and statistics taught every year, and an upper-division follow-up course taught every other year. Our teaching load was five courses per year, so in each two-year cycle I taught 10 courses, but only three were in probability and statistics. The rest was calculus I and II, or, if I was lucky, linear algebra. Unlike my department colleagues, who had taught calculus as grad students, I hadn't had to think about how to learn or teach calculus in the many years since I took it as a student, so what for them would have been no big deal was for me a major challenge.

AR: What was your teaching style at that time?

GC: Ouch! Your mention of teaching style triggers painful memories. My style was whatever it took just to get me through each hour still alive. I was more comfortable looking at the blackboard than looking at my students, and I came each day having written out an entire script. I never had to look at it when I was in front of the class because I'd obsessed over it for so long in advance. My notes were mainly a security blanket, a written version of sucking my thumb. In my head, I knew that I needed to work toward a better approach, but it took years of effort to change. Fortunately, I had wonderful students who were willing to work hard, who cared about learning, and who were a lot more comfortable than I was. I think what saved me was a fact I learned from David Moore decades later: When it comes to evaluating teaching effectiveness, one of the three things that correlates most strongly with student learning is the students' sense that their teacher cares. I did care, and they sensed it.

AR: You said "student learning." Did you mean "student rating of instructors?"

GC: An important distinction. Almost surely, you've got to be right, since student ratings are easy to measure, and student learning is harder to measure. All the same, I got this from David Moore's talk and article, "The Craft of Teaching" (1995), where he was careful to distinguish students' ratings based on mere popularity from ratings that correlated with actual learning.

AR: You also said "one of the three things," but you mentioned just one, caring about student learning. What were the other two?

GC: Clarity and responsibility. Clarity about what students were expected to do, and responsibility about showing up on time, ending on time, and getting written work returned promptly.

AR: Before I ask more questions about statistics, let me ask about Mount Holyoke more generally. You went from Dartmouth and Harvard to a college with only undergraduate students. Even more distinctive is that Mount Holyoke only has women students. How did you find the transition to such a campus environment?

GC: The transition from one institution to the next actually felt secondary. At Dartmouth, which was all male at the time, I lived in a dormitory. At Harvard, I was married and living in an apartment, so that was the big transition. At Mount Holyoke, I was so preoccupied with learning how to teach that yet again, the institutional change was secondary. However, after I began to get used to teaching, I did start to think about what it meant to be teaching in a math department at a college for women. At Dartmouth, not only were the students all male, so were the faculty. At Harvard, in statistics, the gender balance among the grad students was about 50/50 – in stark contrast to math, which was almost all male. At Mount Holyoke, the students were all female, but the faculty ratio was about 50/50. One of my deans used to say that the women's colleges were the only places with truly coed faculties. In retrospect, though, what stands out the most was something I learned slowly, over time. From time to time I would have a male student who had cross-registered from Amherst or Hampshire or UMass. The damping effect on classroom dynamics was striking, even when the male student kept quiet most of the time.

AR: Did you find that much changed over your time at MHC with regard to the culture of having all female students? If culture is not the best word for me to use here, did you encounter any noticeable changes over time in this respect?

GC: To the extent that there was a single watershed year, it was 1972, just before I joined the faculty. All-male colleges like Dartmouth and Bowdoin were becoming coed, and professional schools (medicine, law, business) were admitting women in much greater numbers. Until that point, the Seven Sister schools had been for women what the all-male Ivies had been for men. Then, for the first time, talented women could go to Princeton or Dartmouth, and from there to Harvard med or Yale law. It put tremendous pressure on the women's colleges to think seriously about admitting men.

AR: How did Mount Holyoke respond to this pressure?

GC: Some women's colleges didn't face as much pressure because they were paired historically with a men's college: Barnard with Columbia, Bryn Mawr with Haverford, Pembroke with Brown, Radcliffe with Harvard. Some of the others (Vassar, Wheaton, MA) chose to admit men. Still others (Mount Holyoke, Smith, and Wellesley) were forced to reexamine their commitment to women's education in the context of a major new source of competition for students: With new options available to women, was there still a need for all-women's colleges? It was in this context that my experience in the classroom convinced me that the answer was definitely Yes.

AR: Do you think that you taught math and statistics differently, based on having all, or almost all, women students, than you might have with students of both sexes?

GC: I didn't set out to; I was too naïve at the start. Also, since Mount Holyoke was my first and only teaching job until late in my career, I had no control group for comparison. But it's clear that over time, being at a women's college with a coed faculty had a shaping influence on my teaching. For example, in the '70s, the Moore method, developed by the topologist R.L. Moore at Texas, was very much in vogue, as an early antecedent of active learning. It was effective for the right kind of student, but its emphasis on being the first to find a proof made the classroom atmosphere extremely competitive. At women's colleges like Mount Holyoke, we wanted a more welcoming, cooperative approach to active learning, even if that meant offering a math major less focused on getting students into top PhD programs. I remember that in 1976, our graduating class had 36 math majors; down the road, Amherst had only six. I expect all six of them went on to earn PhDs.

AR: Can you comment on some ways in which you achieved such a cooperative approach to active learning in your courses?

GC: You make it sound more deliberate than it actually was. I more or less stumbled into it, much as I stumbled into statistics. One semester I decided to teach our probability course using Fred Mosteller's *Fifty Challenging Problems in Probability* (1965). There was no textbook. Instead of proving theorems, students would solve the problems, and instead of competing as in a Moore-method course, they worked in small groups. It worked really well, better than when I did more of the talking. I also started requiring semester projects in my design course, and students could choose whether to work alone or on a team with one or two others. From that point on, I relied on term projects in almost all my courses. Then I had the good fortune to be invited to serve on the advisory committee for Dick Scheaffer's *Activity-Based Statistics* (1996) project, and what I learned from Dick and his collaborators about using activities had a big influence on my teaching.

AR: How did you come to introduce more statistics courses at Mount Holyoke?

GC: It was a long, slow process, because I had to deal with two major challenges, course slots and course content. When I arrived, there was space for only one and a half courses per year in probability and statistics, and constant enrollment pressure in beginning calculus meant there was no room for another stat course. Moreover, four decades back, "data analysis" had zero name recognition, and "statistics" meant either theorems or recipes, and there was no place for recipes in a math curriculum. My first change was to replace the sophomore-level course in probability and statistics with a course based on the 1978 book by Freedman, Pisani, and Purves (4th ed. 2007). Meanwhile, I had been helping some biology faculty with their data, and they started recommending my course to their students who might be doing honors theses in biology. It became clear that what those students most needed was an introduction to experimental design, so the course gradually moved more and more in that direction, until I was teaching design and ANOVA as a first stat course.

AR: Am I correct in assuming that your experimental design book (Cobb 1999) emerged from that course? Can you describe how that course and book differed from more conventional ones, such as the classic text by Box, Hunter, and Hunter?

GC: Yes, that course did lead to my design book. It's not really comparable to Box, Hunter, and Hunter (2005). Their book is written at a higher mathematical level, and it covers a lot more material. For example, it does a lot with designs like fractional factorials and with fitting response surfaces. Their examples come mainly from chemistry, engineering, and process improvement, and the book is intended to be of use to experienced practitioners. I wrote my book with a different audience in mind, undergraduates or graduate students having no previous experience with statistics, and I used examples from biology, health sciences, and psychology.

I was inspired by having taught from the book by Freedman, Pisani, and Purves, and took it as a challenge to try to explain the ideas of design and ANOVA without using any formulas. After all, such ideas as crossing versus nesting, fixed versus random effects, treatment structure versus design structure don't really require notation. My goal was to prepare students to think about design issues when they planned their undergraduate honors theses or masters' theses. And to get back to the long, slow process, that book was 20 years in the making.

AR: Did you develop more statistics courses at Mount Holyoke, in addition to this design course?

GC: Yes. Over the course of my first fifteen years, my department supported me in building to the point where we were able to offer a major in statistics, consisting of calculus, I – III, linear algebra, three applied statistics courses (design/ANOVA, applied regression, and a seminar in data analysis) plus three mathematical courses (probability, mathematical statistics, and linear models). Throughout, there were twin challenges: "Where do the course slots come from?" and "Does this really qualify as a math course?" The design course, for example, replaced the sophomore-level course in probability and statistics, but that change meant the course no longer counted toward the math major.

Another way my department supported me and statistics was by deciding to designate a second position for statistics, which decision brought me a wonderful colleague, Janice Gifford.

AR: You don't mention enticing students to take the courses as one of the challenges. Was there considerable student interest in statistics courses from the start, or did you have to recruit students? From where did the students for these courses come– mostly math majors, or from other disciplines?

GC: Looking back, I wish I'd had tried more actively to recruit students, but I didn't. Partly, I was trying hard to avoid doing anything that might leave students feeling pressured; partly I didn't want to appear to be competing for students with my mathematician colleagues. I now think I was being too careful. On reflection I think I was over-careful because I was a man teaching at a college for women, and I didn't want be pushy. I'm guessing that if I'd been teaching at a coed college, I would have been more aggressive in encouraging students to go into statistics.

For the most part, students came to my courses either because they were already interested, or because they had been in one of my calculus classes and liked the way I had taught the course. The first time I offered a regression course – a department "extra" made possible by a grant from the Sloan Foundation – I wrote and circulated a course description in advance, and 42 students signed up. That was enough to justify a follow-up course, which became our "seminar in data analysis."

AR: Please tell us about that follow-up course: What topics did you teach? What kinds of data analysis experiences did students engage in?

GC: The regression course had required a semester-long data-analysis project, and the follow-up course was based on extending those projects. I assigned two student consultants to each student project, so each student had her own project, and served as consultant to two others. Class time was devoted to presentations, first by the student doing the project, and then by the two consultants. Initial presentations were based on the final paper from the semester before. Initial consultants' reports pointed out possible follow-up directions and questions, additional data to gather, and possible shortcomings to be addressed. We cycled through a second time with intermediate results and comments, and finished with an oral presentation by each student on her completed project.

AR: I believe that you also developed a course on Markov Chain Monte Carlo (MCMC) methods, with support from NSF. Can you describe what you taught in that course and how you made the material accessible to undergraduates with limited background in statistics?

GC: The MCMC course had two goals, to attract prospective math majors to statistics, while also offering an intro to MCMC for stat majors. To serve the first goal, I needed to assume no stat background and make the course mathematically interesting. To serve the second, I needed to avoid duplicating the standard intro stat material and offer something new. The only prerequisite was a course in linear algebra, so hypothesis testing, confidence intervals, and Bayesian posteriors had to be introduced from scratch. By relying on simulation, I could offer stat majors a new way to understand tests and intervals while also making the concepts accessible to students with no previous stat background. Random walks on graphs offered an entre to different kinds of convergence and the eigen-stuff needed to understand rates of convergence. Somehow, we managed to explore Metropolis-Hastings, the Gibbs sampler, and an extended example of hierarchical modeling. We pretty much avoided calculus.

AR: That sounds like a fascinating course. Switching gears a bit, do I remember correctly that you took on an administrative position for part of your career? What position was that, why did it appeal to you, and what did you achieve in the role?

GC: From 1989 to 1992 I was head of the department of academic deans for students. This was a job filled from the teaching faculty for a three-year term. Compared with classroom teaching, I had much less impact on a vastly larger group of students, but also a much greater impact on a much smaller number of students, such as those applying for Fellowships like the Rhodes, and those on academic probation. It was very demanding of time, often emotionally demanding, and very satisfying. I was very glad I did it, and I vowed never to do it again.

Writings on Statistics Education

AR: Let me steer our conversation away from Mount Holyoke and toward your broader impact on statistics education, and I'd like to start with a personal reflection. I became a faculty member in 1989, and one of the first things I read was your 1987 JASA *article (Cobb 1987) that provided a framework for reviewing introductory statistics textbooks. I think this is a remarkable 20-page article in which you reviewed 16 books simultaneously and also provided a framework with which instructors could review textbooks and make adoption decisions. How did this project come to be, and how did you approach this seemingly daunting task?*

GC: It was another happy stumble. Judy Tanur, who was the book review editor for JASA at the time, invited me to be one of her associate editors. I had just gotten tenure, for which the booby prize was to become department chair. Between being swamped and being shy, I let the books needing review pile up, one after another, without managing to line up any reviewers. Finally, out of desperation, I decided to review them all myself. Once I'd backed myself into that corner, the easiest way out was to do a single comparative review.

AR: I have to confess that I laughed out loud at your description of this stumble. I was imagining that you had long planned to write a comparative review of so many books as a major scholarly undertaking. The most memorable aspect of your review to me was your refrain to "judge a book by its exercises, and you cannot go far wrong." I was struck both by the wisdom of the advice, which never would have occurred to me on my own, and also by your literary device of repeating that advice at strategically chosen points throughout the article. Had you gone into the project with this advice in mind, or did it occur to you as you reviewed the books?

GC: I started the review without any conscious preconceptions (which is quite different from starting consciously with no preconceptions). By 1987 I'd had many years' experience choosing textbooks, and could call on my memory for which choices had worked well, and which had been disasters, but I had never tried to think systematically about how to make a good choice.

What a marvelous question, though! It makes clear to me, with benefit of hindsight, what an important learning experience that book review was for me, and how it led me to an approach to thinking about statistics education that I was to continue to rely on, even though I never stopped to think about the process.

For the book review, I'd go through the books one at a time, making notes and jotting down thoughts as they came to me; then I'd review my notes, looking for patterns and questions to check out, and repeat the process. As structure began to emerge, I'd try making an outline as part of each cycle. Once the outline began to settle down, I'd start including bits of draft prose as part of the cycle.

AR: You began that review with three observations that I'll paraphrase here: that statistics is about analyzing data, that data analysis is interesting and intellectually challenging, and that most introductory textbooks do a remarkable job of concealing the previous two points. Do you think textbooks do better now than when you wrote that?

GC: Vastly better. Back then, only a small handful of the best books used real data. These days almost all books use real data. Back then, few if any exercises offered the chance to learn something interesting based on non-obvious patterns in data. That, too, has changed for the better.

AR: Let me turn next to your 1992 article that was published in the MAA volume Heeding the Call for Change: Suggestions for Curricular Action *(Cobb 1992), in which you report on recommendations for teaching statistics from a focus group that you chaired. How did this group and report come about?*

GC: As best I understand it, Lynn Steen, who was PI for the NSF-supported project "Heeding the Call for Change," invited me, thanks to a suggestion from David Moore. I moderated an e-mail discussion that extended over many weeks, and edited the responses into a report. I had the good fortune to moderate discussion among a wonderful group of thoughtful statisticians, who supplied wonderfully thoughtful ideas, so my own contribution was almost exclusively editorial. I worked hard at that, however, because I wanted to turn the good ideas into something that would be brief enough, and pithy enough, to be memorable, in the spirit of "Judge a book by its exercises …" So I claim no credit for the ideas in that report, but I do think my efforts at writing helped give those ideas the visibility they deserved.

AR: I think your efforts were quite successful. I must say that one of my all-time favorite sentences comes from that report: "Shorn of all subtlety and led naked out of the protective fold of educational research literature, there comes a sheepish little fact: lectures don't work nearly as well as many of us would like to think." Do you have any wisdom to share about how you achieve such imaginative and memorable ways to express key ideas?

GC: I can only speculate, and certainly I have no wisdom to offer, but in my case I think word play is a form of rebellion, in the spirit of the two-year-old boy who resisted toilet training, saying "*No*, Mommy -- *I'm* the boss of my doo-doos."

AR: The three recommendations from that report were: 1) Teach statistical thinking, 2) More data and concepts; less theory and fewer recipes, 3) Foster active learning. How hard was it arrive at these three recommendations for the focus group to embrace? I can imagine that some members might have pushed for more recommendations, or differently worded ones. Was there much dissension or debate or compromise?

GC: Thanks for asking. I've already claimed credit for some editorial work I think I did well, but your question gives me the chance to acknowledge substantive contributions from others that I've sometimes gotten undeserved credit for. I think here in terms of form (me), substance (others), and inertia (me exploiting them). (1) Form: I had convinced myself in advance that all too often a "task force" proves to be more "task" than "force." For a collegial group like ours, the tendency leans toward "I'll let yours in if you let mine in." This ecumenical generosity can lead to a very long and unfocused list of recommendations. To avoid this bog of diffusion, I suggested to our group early on that we plan to condense our main points into a summary short enough for others to want to quote. It was not hard to get agreement on that approach. (2) Substance: We had lots of really good ideas. In fact, I don't remember any bad ones. Some recommendations were concrete and detailed; others were abstract. When it comes to concrete detail, of course, all curriculum is local in that what works for me in my classroom may not work for you in yours. But: our group's goal of aiming for a brief summary made it all but essential to focus on broad issues, which made agreement easier. In the end, our report relied heavily on David Moore, who is responsible for the wording in the three recommendations. (3) Inertia: As the saying goes, "It's easier to get forgiveness than permission." It took me several weeks to put together a draft report. By then, everyone else was busy with other things, and no one had much time or energy for complaining. Out of inertia, I was forgiven.

AR: In 2005 the ASA endorsed the GAISE (Guidelines for Assessment and Instruction in Statistics Education) report (ASA 2005), which provided an update of your group's 1992 recommendations. Do you think the timing was right for that update, and were you pleased with the new report?

GC: Definitely, yes to both.

AR: Did you have any concerns at the time, or since, that the number of recommendations was doubled from three to six? I'll also invite you to express worry that even though the recommendations were still quite succinct, they were slightly less economical in their use of words than in your report. (I'll be silly and calculate that the average number of words per recommendation increased from 5.0 to 6.83.)

GC: I'll be silly back at you and ask why you didn't compute a trimmed mean, and what reference distribution we should use for assessing the significance of the difference. More seriously, I think you and I would agree that the GAISE report, coming 15 years after the Steen volume, and many years after the success of the Advanced Placement course, needed to address a somewhat different audience with somewhat different concerns. *Whether* to teach statistics was largely settled, except in the backwaters, and a more detailed set of recommendations about *how* to teach statistics effectively was what we needed. GAISE delivered.

AR: Now let's turn to one of your articles co-authored with David Moore in The American Mathematical Monthly*, "Mathematics, Statistics, and Teaching" (Cobb and Moore 1997). Who was your primary audience for this article, and what were your goals?*

GC: I've always been grateful to David Moore for inviting me to help write that article. Ever since the second half of the 17th century, the relationship between mathematics and statistics has been fruitful, evolving, and more recently, uneasy. David had already written articles about that relationship, especially its implications for teaching: "Should Mathematicians Teach Statistics?" (Moore 1988) and "Teaching Statistics as a Respectable Subject" (Moore 1992). In deciding to write for the *Monthly*, we were choosing to address mainly academic mathematicians who cared about teaching, and our goal was to engage our mathematician colleagues in thinking about statistics as overlapping with mathematics, as different from mathematics, and as important in its own right.

AR: That article includes another delightful sentence that I admire, about how the role of context is much different in statistics as compared to mathematics: "The ultimate focus in mathematical *thinking is on abstract patterns: the context is part of the irrelevant detail that must be boiled off over the flame of abstraction in order to reveal the previously hidden crystal of pure structure." You and David go on to say that the crucial role of context in data analysis means that an effective statistics teacher "must, like a teacher of literature, have a ready supply of real illustrations, and know how to use them to involve students in the development of their critical judgment." Finally, I arrive at my question, which I'm afraid might put you on the spot: Please give us an example or two of your favorite examples/contexts/illustrations that you've used for this purpose.*

GC: I could go on at length, but I'll limit myself to five. (1) One of my all-time favorites I learned from you (Rossman 1994): The cases are nations of the world, and the two variables are life expectancy at birth and number of people per TV set. The fallacious causal interpretation is that watching TV makes you live longer. (2) Another example is in the same spirit, except that the lurking variable is more obvious and more easily measured: The cases are the US states, and two variables are the number of college students living in dormitories and the number of people living in cities. Once you adjust for state population size, the correlation goes from very strongly positive to clearly negative. (3) A 2x2 table from the murder trial of an intensive care nurse, in *Statistics: A Guide to the Unknown* (Cobb and Gehlbach 2006) offers a way to talk about what p-values can tell you and what they cannot, in the context of one of my courtroom experiences. (4) For complexities of multiple regression modeling, I like survey data for which the cases are 28 academic subjects taught at universities, the response is mean academic salary for the subject, and one of the predictors is the percentage of faculty who are women. For the data, search for "Chilean Journal of Statistics, Cobb" (Cobb 2011). (5) Finally, Dick Scheaffer and his NSF project "Activity Based Statistics" introduced me to student-generated data sets, and several of those activities became favorites.

Expert Witness Experience

AR: I've been meaning to ask about your experiences as an expert witness in a courtroom, so let me take this opportunity to do that. The article that you mention, co-authored by you and Stephen Gehlbach (Cobb and Gehlbach 2006), described some of the statistical issues in the murder trial of Kristen Gilbert. How did you come to get involved with this kind of work, and with that trial in particular?

GC: The phone call asking me to work for Gilbert's defense team came out of the blue, but I'd been doing occasional legal work for many years, and apparently my name had gotten around. It all started during my first sabbatical, which I spent back at Harvard. I was struggling to make ends meet financially, not quite reduced to a ramen noodle diet, but close. Art Dempster invited me to assist him with statistical analysis in relation to a possible lawsuit alleging employment discrimination. That not only rescued the budget, but gave me a wonderful learning experience assisting Art. After my sabbatical ended, I mentioned to the daughter of a friend and colleague that I had done this work, she mentioned it to a friend of hers, who (the friend) mentioned it to her own father, an attorney, who mentioned it to a fellow attorney in an elevator conversation. The rest, as they say … I never tried to compute the chance of such a string of improbables.

AR: Can you give us a sense for the kinds of legal cases that you have consulted for, and also for some statistical issues that have arisen?

GC: The Gilbert case was unique in my experience. Almost all of my legal work is of three kinds: employment discrimination, civil commitment, and Medicare fraud.

Employment discrimination: Have employees who belong to a "protected class" based on age, sex, or ethnicity been adversely affected, e.g., promoted less frequently or laid off more frequently? Statistically, the issue is one of comparing two groups, but the comparison is never straightforward, because individuals vary: with respect to education, years of service, job classification, etc.

Civil commitment: In many states, a person convicted of a sexual offense who has completed his sentence can be committed to civil confinement if he is judged likely to reoffend. Over the last few decades a small cottage industry has created large databases and used them to develop simple scoring systems that purport to estimate the chance of recidivism. Setting aside the important legal issues involving civil liberties, the statistical issue is whether and to what extent the scoring system is valid and reliable when applied to the individual in question.

Medicare fraud: Fraud and inaccurate billing costs the U.S. an immense amount. Think motorized wheelchairs, which are heavily advertised and too often bought for people who don't really need them. Our government has farmed out the recovery of unjustified payments to profit-making corporations that are paid a percentage of whatever they can recover. The corporation uses billing records to decide which health care providers to audit, the audit is based on a random sample of payments to that provider, and the total overpayment in the audited sample is extrapolated to the population of all payments. Here the statistical issues are related to the sampling process and to the validity of confidence intervals based on the Central Limit Theorem.

AR: Have you often appeared in court? The Perry Mason fan in me has to ask: What does it feel like to be cross-examined by someone who perhaps desperately wants to discredit what you've said?

GC: I'm guessing maybe a dozen times. I'd work on one or two cases a year. Perhaps two-thirds of those got to the point of an affidavit based on my analysis, and of those, roughly half settled before a trial, and for the remaining half I'd be called to testify. I, like you, was a fan of Perry Mason, but I'm glad he never cross-examined me, because that would mean I was inevitably destined to be discredited. The second part of your question gets at one of the reasons why I decided to do this kind of work – as a shy person I wanted to challenge myself, and to find out what it would feel like to be cross-examined by a professional being paid hundreds of dollars to make me look like a jerk. For me, the main feeling was being anxious and self-conscious as the focus of attention in an adversarial proceeding. (I would have made a terrible professional athlete.) I was also, at first, overawed by the courtroom setting. Over time, with experience, I became better able to keep things in perspective. But I never got over feeling uncomfortable feeling caught in the middle, knowing that the client, and often the attorney also, wanted me to be more certain in my testimony than could be supported by the data.

AR: Can you comment on how you have used the Kristen Gilbert case, and the Robert Martin age discrimination case that appears in Statistics in Action (Watkins, Scheaffer, and Cobb 2004), and perhaps others, in your teaching?

GC: As you know, a major issue in the Gilbert case is the meaning of a single table of counts of 8-hour shifts classified two ways: Was Gilbert present? Was there a death on the shift? I've used the data to illustrate the mechanics of comparing two proportions using a normal approximation, testing for association using the chi-square test, testing using Fisher's exact method, and estimating the Fisher p-value via simulation. Beyond the mechanics, there is a lot of room for discussion about the meaning of a p-value when there is no chance mechanism. The Martin data set is richer, and offers scope for exploratory analysis, but there is also a relevant 2x2 table that counts employees classified as old or young, laid off or not. The cell counts are small enough that you can carry out Fisher's exact test by listing all the possible permutations, and here, also, there's an issue of what the p-value does or doesn't tell you.

AR: Your work as an expert witness clearly enhanced your teaching. I'm wondering if you also found a benefit in the other direction: Do you think your teaching experience helped your work in the courtroom? Did you feel that you were trying to help the judge or jury to understand statistics as you were testifying?

GC: Trying to get the substance across without oversimplifying was a constant challenge, and my years of experience working at that same sort of thing in the classroom was definitely useful. Two instances come to mind, both from the same age discrimination case against Ground Round. (1) To us, it's pretty much intuitive that 0.05 is a small p-value, and that 0.01 is very small, but my experience teaching had taught me to try to imagine what it might be like for others. So to illustrate a p-value, I held one end of a tape measure and asked the attorney for the plaintiff to back off a distance of 100 inches. To show 0.05 visually I pointed to 5 inches on the tape, and after that, to the actual p-value, which was, as I remember, less than .01. (2) A harder challenge was to explain the requirement of independence for the two-sample t-test. To compute the p-value, I had used a permutation test to compare the ages of those fired with the ages of those not fired, in the plaintiff's job category. The defense used a two-sample t-test to compare the average age before the layoffs with the average age after the layoffs for a much larger group that included employees in several different job categories, both before and after. The points of contention were the requirement of independent samples and the relevance of the sample. I showed the jury a quarter in one hand and a nickel in the other, representing those laid off and those kept: a difference of 25 versus 5. Then I asked them to imagine adding nine dimes to each hand, to represent the employees in other job categories who were present in both of the samples in the defense analysis. Now the average difference has changed to 11.5 versus 9.5. I asked the jury, "Which comparison is more informative?" The award for the plaintiffs set a record in the history of Massachusetts. Of course I was getting paid by the hour, regardless of outcome, so my take, re-expressed as a percentage of the total, was even smaller than the p-value. And honesty compels me to add that my role was very, very secondary: The twenty-something MBA responsible for the layoffs was on record saying, "We've got to take everyone over 40 and make them go away."

Statistics and Liberal Arts

AR: We talked earlier about the experience of teaching at a women's college. Now I'd like to ask about being a faculty member at a liberal arts college. Can you talk about some of the challenges and opportunities of being a statistician in such an environment?

GC: Two caveats: First, my sample size is the only positive integer strictly less than 2, and second, times have changed, especially when it comes to recognizing the practical importance of statistics for any area of research that claims to be scientific. Caveats deployed, I am convinced that Humanities faculty who choose to teach in a Liberal Arts environment are more open to recognizing the intellectual legitimacy of statistics. I have often pestered my colleagues with the reminder that statistics is one of the few interpretive enterprises that seeks to place itself within a rigorous deductive structure. We who analyze data are constantly challenged to struggle with the tension between abstract mathematics and interpretation-in-context. Liberal Arts faculty are inclined to value that struggle.

AR: Am I correct that you were a founding member of the Statistics in the Liberal Arts Workshop (SLAW) group? How did this group come to exist, and what was its purpose at the start?

GC: Actually, although I attended the first meeting of SLAW, and benefited mightily ever since, I can't claim any credit for the founding of SLAW. As I understand it, the idea emerged at a conference on statistics education held at SUNY Oneonta, thanks to Robin Lock, Tom Moore, Rosemary Roberts, and Jeff Witmer. Tom used Grinnell College's grant from the Sloan Foundation's New Liberal Arts Program to fund our first meeting, which was attended by Bob Hogg and David Moore, representing well-known small liberal arts colleges from Iowa and Indiana. Informally, the purpose was to create a "department in exile" (aka support group) for statisticians teaching in mathematics departments at Liberal Arts colleges across the country. There were so few of us back in 1987 that we were a small group at the start. As undergraduate statistics began to grow, we agonized about membership but decided to stay small for the sake of group dynamics – we wanted to stay more like a department than turn into a conference.

AR: Do you subscribe to the view that statistics is *a liberal art? If so, in what ways?*

GC: I borrow from Jaques Barzun, former president of Columbia. In his book *The House of Intellect* (1959), he argues that what we commonly refer to as "disciplines" are actually just subject areas. A discipline, according to Barzun, is a way of thinking, and there are only two, defined by Pascal and named by him as "the spirit of subtlety" and "the spirit of geometry." As I understand it, Pascal's two disciplines, his two ways of thinking, correspond to interpretation in context, and rigorous abstract deduction. As I understand it, also, a Liberal Arts college is devoted primarily to teaching ways of thinking, so a subject like statistics that struggles uneasily to integrate Pascal's spirit of subtlety and spirit of geometry should be central to the Liberal Arts. For a similar view, we can go back to Plato's curriculum for the philosopher king. The decade from 20 to 30 should be devoted to the study of mathematics, Pascal's spirit of geometry, as training for the mind, in preparation for another decade, from 30 to 40, devoted to the harder subjects involving the spirit of subtlety, such as sociology and psychology and political science.

AR: Would you recommend a career at a liberal arts college for young statisticians interested in teaching? I'm going to go out on a limb and assume that the answer is yes, so I'll proceed directly to some follow-up questions. What are the best aspects of such a position, and what is less appealing? What might you encourage someone to ask her/himself to assess whether a career as a liberal arts statistician might be right for her/him?

GC: Two things. I'll start with a metaphor from statistician Paul Rosenbaum, who did his undergraduate work at Hampshire College. "At most places," Paul said, "there are halls and walls. You can go quickly down the halls, but unless there's a door, you can't go through a wall. At Hampshire, you're in a swimming pool filled with molasses: you can take any path you want, but you can't go anywhere very fast." It's my sense that at Liberal Arts colleges, especially if you are the only statistician, you get to choose the path, but forward progress will often be slow. (When it comes to statistics, some math departments are more like molasses than others.) At colleges and universities with established statistics programs, it's more like halls and walls. Second, at Liberal Arts colleges, there is more openness to regarding curricular innovation as a form of scholarship. As my former Dean of the Faculty Don O'Shea used to say, "Liberal Arts Colleges are the places where cutting edge research from the universities is brought into the undergraduate curriculum."

Conference Presentations

AR: I've enjoyed hearing many conference presentations of yours over the years, and I've been especially entertained (and challenged) by some talks you've given at conference banquets. For example, you gave some remarks following the opening banquet for a roundtable conference on assessment in statistics education, organized by Beth Chance, in 2004. The title of your presentation was "Against Fairness" (2004). True to your title, you argued that instructors should not try to be fair in assessing student work. And then for good measure, and perhaps to demonstrate that you were in favor of something as well as opposed to fairness, you made a plea for grade inflation. Would you please summarize your arguments here for JSE *readers?*

GC: When I started teaching in 1974, I assumed, as I had been taught, that "fair" grading was based on a single scale that tried to measure how you the student compared with other students at the end of the course. Before I left the classroom a few years ago, I had convinced myself that a student's grade should reflect *progress*: the *quality* of a student's *learning* over the course of the semester, not an end-point inspection for defects. (Readers may recognize my use of Deming's language.) I'd also convinced myself that if teacher and student do a good job of directing that student's effort, then learning depends mainly on effort. In short, if I'm a decent teacher, a student who works hard will learn a lot, and deserves an A. The "one size fits all" scale allows some students to coast to an A while others struggle mightily to eke out a C. Finally, this approach helps students feel comfortable in a course like my MCMC course, where folks have very different backgrounds and goals.

Of course this learning-based approach to grading depends on two things I was fortunate to have at Mount Holyoke: comparatively small classes, whose size allowed me to pay attention to individual students, and an environment where students were serious about learning and willing to work at it.

Finally, I admit that a more accurate title for my talk might have been "Some reasons not to use a uniform yardstick for grading in small classes, and one possible justification for giving more A grades." If charged that in choosing "Against Fairness" I was sacrificing precision for the sake of grabbing attention, I plead guilty.

AR: I also have to ask about the third part of your "Against Fairness" presentation, which was all about a cartoon titled "Roger." Would you please tell JSE *readers about Roger and how this cartoon helped to shape your teaching philosophy?*

GC: The cartoon is by Gary Larsen. Roger plays cymbals in an orchestra, and as the piece nears its end, Roger needs to bring his cymbals together in a final crash. The bubble over Roger's head shows him chanting to himself, "I won't screw up … I won't screw up." But poor Roger is destined to screw up: One hand holds a cymbal, but the other hand is empty. He's so preoccupied with not doing wrong that he can't focus on what matters most. It's my hope that by grading students on how much they learn instead of how they compare with others in the class, we encourage them to think more consciously about how their effort relates to their own learning.

AR: Let me move on to ask about another of your after-dinner talks at a conference. You spoke at the first United States Conference on Teaching Statistics (USCOTS) in 2005, with the intriguing title of "The Introductory Statistics Course: A Ptolemaic Curriculum." Would you summarize your thesis in that presentation and explain the metaphor?

GC: Ptolemy started with a simple model: the earth is at the center of the universe; the sun and planets revolve around the earth in circular orbits. The traditional Stat 101 also starts with a simple model: the Central Limit Theorem puts the normal at the center of all inference. Over time, astronomers noticed anomalies, and invented epicycles to tweak Ptolemy's model to fit the facts. In much the same way, over time statisticians found anomalies, and added tweaks: t in place of z, and adjustments for unequal SDs. What had begun simply became increasingly complicated and removed from the simple model. In astronomy, Copernicus, Kepler, and Newton found a much simpler model for the solar system. In statistics, Fisher and Pitman found a much simpler model for inference, based on randomization, but, to paraphrase Fisher, we couldn't use the simpler model because we didn't have the computing power. In 1937, Fisher was right: we didn't. Now, we do.

AR: Six years later the theme of the 2011 USCOTS was "The Next BIG Thing." The consensus at the conference seemed to be that re-centering introductory courses around the core logic of inference, illustrated by randomization tests, was indeed the next big thing in statistics education. Several different people and groups have developed curricula and textbooks and faculty development programs around these ideas. A few examples include Zieffler (2013), Tabor and Franklin (2013), Lock, et al. (2013), and Tintle et al. (2015). Your 2005 presentation, and the subsequent article that you published in Technology Innovations in Statistics Education *(Cobb, 2007) are cited very frequently as providing the impetus for this movement.*

Now that we're approaching the ten-year anniversary of your USCOTS presentation, I'm wondering about your reaction to the statistics education community's response. Are you pleased, disappointed, surprised, or …? Has the reaction exceeded your expectations or left you feeling underwhelmed, or …?

GC: The implied premise of your question is overly generous. Many people, yourself among them, had been thinking about a simulation-based curriculum. In my USCOTS talk, I mainly tried to be persuasive about ideas that were very much in the wind at the time. That said, I've been somewhat surprised that so many people have been doing innovative simulation-based work in the last decade, and very pleased at the quality of that work.

AR: I happen to know that you've proposed a session for the 2015 USCOTS in which you contend that the introductory statistics course could be improved substantially by paying more attention to both cultures of data analysis described by Breiman (2001): one based on probability model and the other based on algorithms. Could you give us a preview of the argument that you'll make at the 2015 USCOTS?

GC: I think Breiman's article in *Statistical Science* was ahead of its time to distinguish two cultures in statistical practice. I'll first summarize Breiman's distinction, then say what it suggests for the future direction of statistics education.

Breiman's "stochastic culture" is based on a three-step model: Input -> Nature -> Output. To oversimplify, this approach tries to find a probability model for "Nature," the middle step that relates input to output. As I see it, our 60-year-old math stat courses are fossilized remains of the stochastic approach. Brieman's "algorithmic culture" treats "Nature" as a black box, and tries to find a pattern that relates input directly to output, much as behavioral psychologists try to relate stimulus directly to response without recourse to Freud to model the connection.

For me, the implications for what we teach are profound. I'll focus here on three: goals, criteria for success, and learning curve; and I'll rely on two examples, the two-sample t-test and a classification tree: t versus tree. The nominal goal of the t-test is to decide whether two population means are equal, but to carry out the test, you have to take a detour to create a probability model, and as a result, the learning curve is steep. You need to learn about sampling distributions, the central limit theorem, and the t-adjustment for using an estimated SD. After all that, how do we know that the t-test is a good approach? It's "beyond the scope of the course." Now consider a classification tree to tell the difference between spam and non-spam. The goal is obvious, the criterion for success is directly related to the goal (what's the percent error?), and there's no learning curve. You can present the problem on the first day of class, give students data, and let them take over the rest of the class period.

Bottom line: More and more of our practice is based on the algorithmic approach, and that approach is much more accessible to students. More and more of our teaching should reflect those realities.

AR: I like these examples, but suppose the two-sample t-test were replaced by a permutation test. Am I right that this would still belong in Breiman's first category, because it's still based on a *probability model? But some of your objections about the steep learning curve would go away. Might you propose introducing both randomization-based inference as well as algorithmic methods in our courses?*

GC: Yes, definitely. But I find myself wanting us to develop multiple entry points to statistics. I see randomization-based inference as the most direct way to embed statistics within the scientific logic of falsifiable hypotheses. That logic is not intuitive, and may need an entire semester. So I'm also thinking about a different intro level course based on, or at least introducing, the algorithmic approach. I've never taught such a course myself, but other people have, among them Amy Wagaman at Amherst and Dick de Veaux at Williams.

AR: I skipped from your 2005 USCOTS presentation to your upcoming session for the 2015 conference, and now I want to go back to the 2009 USCOTS, where you gave a presentation for which your abstract argued that "we should think hard about teaching Bayesian logic as part of the introductory course." Would you summarize your argument and speculate about where we might be heading with regard to this topic?

GC: As Eeyore once said, "Thanks for noticing." As I see it, leaning mainly on David Moore's "Bayes for Beginners? Some Reasons to Hesitate" (Moore 1997), there are three main arguments against teaching Bayesian logic in a first course: (1) Few published papers use Bayesian methods. (True when David wrote, but no longer true, thanks to Markov Chain Monte Carlo and hierarchical models.) (2) The dependence on prior probabilities makes Bayesian inference too subjective. (True then, but computing power now lets us assess sensitivity to choice of prior.) (3) It's too hard mathematically. (True then, but now a fallacy.)

The fallacy? The standard route to Bayesian posteriors starts with a mere technicality, not with the essential idea. The technicality is that P(A|B) depends on the marginal probability P(B) in the denominator, which is either a sum or an integral. In most serious applications, P(B) is a high dimensional integral: How can we possibly teach Bayes without at least Calc. III?

This technical question ignores the essential idea, due to Laplace, namely, that $P(\theta|y)$ – what you want to know -- is proportional to $P(y|\theta)$ – which you typically know, or at least assume you know. I like to call this early version of the likelihood principle "Laplace's Data Duplication Principle": a parameter value is believable in proportion to how easily it reproduces the observed data. If you can use $P(y|\theta)$ to generate lots of y-values for a systematic range of θ-values, your estimated posterior probabilities are just fractions. No integrals required!

I've been sorry not to see more people exploring this approach to teaching Bayes in a first stat course. On the other hand, if I'm really a "cultural change pessimist" as David Moore has called me, I shouldn't be surprised (Moore, Cobb, Garfield, and Meeker 1995). I should have assigned a high prior probability to being disappointed.

AR: I think Moore also made the point that probability is a very difficult concept for beginning students to grasp, that conditional probability is even more challenging, and that the logic of Bayesian inference depends on conditional probability. Many introductory courses have aimed *to minimize the study of probability. Do you think an emphasis on Bayesian logic would require greater emphasis on helping students to understand concepts of probability?*

GC: As I see it, there are two separate issues in your question, probability and conditioning. I agree with David that probability is hard, and there is plenty of research to back that up. But to me, conditional probability is no harder, unless we choose to make it hard. All probabilities are conditional – conditional on the choice of sample space. The basic idea is always the same: $P(A) = \#A/\#\Omega$, where Ω is the set of outcomes that are in the sample space, i.e., that satisfy the condition at issue. The reason students have trouble with conditioning is that we make too big a deal of the mathematical definition, $P(A|B) = P(A \text{ and } B)/P(B)$, which is not intuitive. In my experience, it makes more sense to students if we take $P(A|B)$ as a primitive concept and derive $P(A \text{ and } B) = P(A|B)P(B)$. In my experience students tend to find tree diagrams intuitive, and also to find it intuitive to recognize $P(A \text{ and } B) = P(A|B)P(B)$ as a formal version of the logic of the tree diagram. I'd be interested to know if there's been research on whether this approach works better with students than the approach based on the definition.

As to teaching Bayes, I think the hard idea is not the conditioning, but the Laplace principle that reverses the conditioning: $P(\theta|y)$ is proportional to $P(y|\theta)$. I'm not aware of any research into whether and to what extent this principle is intuitive to students, what the associated misconceptions might be, and how easily the principle might be extended to embrace prior probabilities. My hunch is that the ideas are not as hard as the correct meaning of a confidence interval, but I'd really like to see research.

AR: One of my concerns about teaching Bayesian thinking to introductory students is that the concept of a distribution *is quite challenging. It's hard for many students to think of a collection of values as a single entity. And then once novices start to become comfortable with the idea of a distribution, it's a much bigger jump to conceive of a statistic, such as a sample mean or a sample proportion, having a* distribution *under repeated sampling. I confess that I've never had the nerve to introduce the idea of a* parameter *having a probability* distribution *in an introductory course. But you suggest that many students incorrectly misinterpret a confidence interval as a probability interval for a parameter, so perhaps I'm overestimating the difficulty.*

GC: I think your question is not only spot-on in pointing to the main obstacle - the challenging idea of a distribution -- but your question also helps me clarify for myself the strategy I've used to try to avoid that obstacle in order to teach Bayesian logic. I think we all agree that teaching the idea of a distribution, especially a sampling distribution, is both important and difficult. As you say, for many students the notion of a collection-of-values as an entity remains a fog of words even after an entire semester. So I'm led to ask, "Can we introduce Bayesian logic without first requiring students to master the concept of a distribution?" (Agreed, we should continue to spend time over our semester teaching about distributions, but …)

I suggest that the logic of Bayes is simpler and more direct than the idea of a distribution. Here again, I blame our normal-centric curriculum, which misleads us into teaching continuous distributions ahead of more fundamental concepts of inference. The key idea is Laplace's principle that $P(\theta|y)$ is proportional to $P(y|\theta)$. If both θ and y are discrete, and we simulate to estimate probabilities, then Bayes is just a matter of proportional reasoning based on

#Yes/#Trials. The only necessary mathematics is part of the middle school curriculum. The underlying heresy that our curriculum resists is that continuous distributions are merely convenient approximations to a discrete reality, and not the reverse.

If we teach a simulation-based curriculum, we teach the normal as a sometimes-useful shortcut for a broader, discrete, simulation-based approach. If continuous-as-approximation-to-discrete works for Fisher and Neyman, why can't it work also for Bayes and Laplace?

AR: The topic of "big data" has become very popular in the news and in nearly all areas of business in the past few years, and the related idea of "data science" has become widely discussed among statisticians. Indeed, the 2014 curriculum guidelines for undergraduate programs in statistical science lists "increased importance of data science" as the first key point in its executive summary (ASA 2014). Do you have any thoughts on how statistics educators should be reacting to the "big data" and "data science" phenomena?

GC: Do I have thoughts? (Yes.) Are they any good? (Too early to tell.) I'm observing the phenomena largely from the sidelines. Caveats deployed, here are an observation and a worry.

The observation is that until recently, the major shifts in statistical practice have come from *within* our field. I'm thinking in particular about the renewed emphases on Bayesian and randomization-based methods, both of which are regarded as part of statistics. More recently, the pressure to change has come from *outside* the field. I see algorithmic data analysis as transitional. Some methods, like classification and regression trees came from within statistics; others, like neural nets, came from computer science. Still more recently, "big data" has come mainly from outside our field. In this sense, data science represents a new kind of challenge for our profession, the first time we have been pressed by competition from outside.

My worry is that our teaching may respond to that challenge in a way we come to regret. There's certainly no controversy about the importance of big data, and no controversy either about its value to students in getting a good job. For me an interesting, unresolved question is about the proper role of data science in an undergraduate curriculum, especially one with a Liberal Arts emphasis. In teaching about big data, what balance can we offer between creative and critical thinking, as opposed to mere rote and practice? So far, I'm not yet convinced that we have an approach to big data with much intellectual content or coherence. (In what ways does big data involve either of Pascal's two ways of thinking?) But perhaps that's just my inner conservative talking.

AR: If nothing else, I suspect that the big data and data science phenomena are spurring more statisticians to encourage their students to study more computer science than in the past.

GC: I agree, and I think that's a good thing.

Pop Quiz

AR: Now I'd like to ask a series of short questions that you can think of as a "pop quiz," and I'll ask that you confine your answers to a few sentences. First, please tell us about your family.

GC: My wife Cheryl teaches voice in the music department at Mount Holyoke. Our daughter LeeTae recently finished her master's degree in education at Penn and now works for a start-up company in Philadelphia. The four-legged branch of the family consists of two dogs.

AR: What are some of your hobbies outside of statistics and education?

GC: I do the crossword and Sudoku in the daily paper. I have an ambition to become a decent bluegrass banjo picker, and I like woodworking, but I can't claim to be very good at either.

AR: What are some books that you've read recently?

GC: William Manchester's biography of MacArthur (*American Caesar*), Doris Stearns Goodwin's dual biography of Teddy Roosevelt and William Taft, and Shirer's *Rise and Fall of the Third Reich*. I like biography, history, classic novels, and true crime.

AR: What are some of your favorite destinations to which you have traveled? Let me ask for at least one place that you went related to work, and at least one place that was just for fun.

GC: My two favorite professional destinations have been Marrakech and Cape Town, both for ICOTS. For just lazying around and walking near the ocean, I like Ogunquit in Maine.

AR: Do you prefer window or aisle? Do you use a PC or Mac? Do you consider yourself an early bird or a night owl?

GC: Window seat but not Windows computer; early bird.

AR: Now I'll ask a fanciful (perhaps silly) question. Suppose that time travel were possible, and you could take one trip. You can only observe, not change anything, when you get there. Would you travel to a time in the past or in the future? Why?

GC: Definitely the future. I can read about the past, but I find it hard to imagine what it may be like very far into the future.

AR: Here's another fanciful question: Suppose that you are offered dinner for four to discuss statistics and/or education anywhere in the world. Who would you invite for your three dining companions, and where would you dine?

GC: It's a wonderful question to think about. Assuming time travel is possible, I think I'd invite Fisher, Neyman, and some brave third person to referee. I could just listen and eat. With those dinner companions, the restaurant wouldn't matter.

AR: Back to reality and the recent past, what was your favorite course to teach?

GC: I liked them all, except for a very few for which I had chosen an unsuitable textbook, but if I had to list my top three, they would be design of experiments, linear models, and linear algebra.

AR: How long ago did you retire from teaching? What have you been doing since, and what are your plans for the next few years?

GC: I took early retirement from Mount Holyoke five years ago, although I did teach one more course during a delightful semester as a visitor at Eastern Kentucky University a couple of years ago. In addition, I'm involved with a few writing projects, I've done some consulting, and I've been taking banjo lessons. I expect I'll continue with the writing and banjo for the foreseeable future.

AR: I think you've already revealed many things that will come as a surprise to JSE *readers, but a standard "pop quiz" question of mine is to ask you for something about yourself that is likely to come as a surprise to those reading this interview.*

GC: Back in my hippie days (I actually *did* go to Woodstock!) when I was long-haired, bearded, bell-bottomed, and living in Richmond (VA) in the late 60s, I was headed for the bus station late one Friday night when I was grabbed from behind by two tipsy rednecks, while a third put a gun to my head, saying he was going to blow my (participle) head off. Needless to say, my heart and brain froze. What came out of my mouth was "I hope this isn't going to take very long. I have to catch a bus."

Fortunately, we eventually parted company on friendly terms, after they had showed me that the gun really was loaded.

AR: Wow, mission accomplished: I am surprised! And of course I'm delighted with the outcome of this encounter. I hope that you succeeded in catching your bus.

GC: I did, and the bus ride was uneventful, but after I got off the bus and was walking to my parents' house, a cop apparently thought I looked suspicious, and pulled over to question me. Another happy ending, albeit on a lesser scale this time: He gave me a ride the rest of the way home.

Parting Thoughts

AR: Among all of your contributions to statistics education, and I don't doubt that more are to come even in your retirement, can you pick one or two of which you are most proud?

GC: Instead of particular accomplishments, I'd rather suggest two themes that to me, looking back, have motivated much of my work. The first is the importance of the so-called "soft" aspects of our subject: tolerance of ambiguity as well as uncertainty, interpretation in context, and the like. I've come to regard the tension between Pascal's two ways of thinking – his "sprit of subtlety" and "spirit of geometry" as an essential energizing force within statistics. A second theme is the way computing allows us to make basic concepts and practice of data analysis more accessible, less reliant on technical prerequisites. Examples include diagnostics and reanalysis, our simulation-based approach to classical inference, a simulation-based approach (I hope) to

Bayesian posteriors, and a continued broadening of data analysis beyond formal inference through algorithmic thinking.

AR: Before I ask my last question, is there anything else you wish I'd asked about?

GC: Yes. I've been waiting, but you've modestly avoided giving me a chance to talk about your own influence on the way I think about and teach statistics, and I don't want the interview to end without a chance to mention that. Your *Workshop Statistics* series (Rossman and Chance 2012) forever changed the way I write homework exercises, and also the way I came to use class time. In addition, you've set a leading example for me, and I know for many others as well, in making your own ideas and examples freely available on the web for all to use.

Now that I've managed to get that in, I'm ready for your last question.

AR: Thanks very much for all the time you've devoted to this interview and for your thoughtful and thought-provoking responses. My final question is: What advice do you have for JSE readers who are just beginning their careers as teachers of statistics and/or as researchers in statistics education?

GC: Two thoughts. First, local conditions within a department or institution vary so much from place to place, in terms of opportunities and constraints, that I'd be uncomfortable trying to offer advice. But more broadly, I do have some advice: Don't hesitate to join our community. There are now lots of on-line resources. Moreover, and more important, national meetings of ASA, MAA, AMATYC, and NCTM offer many opportunities to present papers, learn from others, and meet others with similar interests. Statistics teachers are a remarkably friendly and supportive group. The future of our field depends on you.

References Cited in the Interview

American Statistical Association (2005), *Guidelines for Assessment and Instruction in Statistics Education: College Report.* Available at: www.amstat.org/education/gaise/.

American Statistical Association (2014), *Curriculum Guidelines for Undergraduate Programs in Statistical Science*. Available at: www.amstat.org/education/pdfs/guidelines2014-11-15.pdf

Barzun, J. (1959), *The House of Intellect*, Harper and Brothers.

Box, G., Hunter, J.S., and Hunter, W. (2005), *Statistics for Experimenters* (2nd ed.), John Wiley and Sons.

Breiman, L. (2001), "Statistical Modeling: The Two Cultures" (with discussion), *Statistical Science*, 16, 199-231.

Cobb, G. (1987), "Introductory Textbooks: A Framework for Evaluation," *Journal of the American Statistical Association*, 82, 321-339.

Cobb, G. (1992), “Teaching Statistics” in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. Steen, Mathematical Association of America, 3-46.

Cobb, G. (1999), *Introduction to Design and Analysis of Experiments*, Springer-Verlag.

Cobb, G. (2004), “Against Fairness,” ARTIST Roundtable Conference on Assessment. Available at: http://www.rossmanchance.com/artist/proceedings/cobb.pdf.

Cobb, G. (2007), The Introductory Statistics Course: A Ptolemaic Curriculum? *Technology Innovations in Statistics Education*, 1(1). Available at: http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art1/

Cobb, G. (2011), “Teaching Statistics: Some Important Tensions,” *Chilean Journal of Statistics*, 2, 31-62.

Cobb, G. and Gehlbach, S. (2006), “Statistics in the Courtroom,” in *Statistics: A Guide to the Unknown* (4th ed.) ed. R. Peck, Cengage.

Cobb, G. and Moore, D. (1997), “Mathematics, Statistics, and Teaching,” *American Mathematical Monthly*, 104, 801-823.

Freedman, D., Pisani, R., and Purves, R. (2007), *Statistics* (4th ed.), W. W. Norton and Company.

Lock, R., Lock, P., Lock Morgan, K., Lock, E., and Lock, E. (2013), *Statistics: Unlocking the Power of Data*, John Wiley and Sons.

Moore, D. (1995), “The Craft of Teaching,” *MAA Focus*, 15, Number 2, 5-8.

Moore, D. (1997), “Bayes for Beginners? Some Reasons to Hesitate,” *The American Statistician*, 51, 254-261 and 272-274 (with discussion).

Moore, D. (1988), “Should Mathematicians Teach Statistics?,” College Mathematics Journal, 19, 3-7.

Moore, D. (1992), “Teaching Statistics as a Respectable Subject,” in *Statistics for the Twenty-First Century*, eds. F. Gordon and S. Gordon, Mathematical Association of America, 14-25.

Moore, David S., Cobb, George W., Garfield, Joan, and Meeker, William Q. (1995), "Statistics Education Fin de siècle," *The American Statistician*, 49, 250-260.

Mosteller, F. (1965), *Fifty Challenging Problems in Probability*, Addison-Wesley.

Rossman, A. (1994), “Televisions, Physicians, and Life Expectancy,” *Journal of Statistics Education*, 2, 2. Available at: http://www.amstat.org/publications/jse/v2n2/datasets.rossman.html.

Rossman, A. and Chance, B. (2012), *Workshop Statistics: Discovery with Data* (4th ed.), John Wiley and Sons.

Scheaffer, R., Gnanadesikan, M., Watkins, A., and Witmer, J. (1996), *Activity-Based Statistics*, Springer-Verlag.

Tabor, J. and Franklin, C. (2013), *Statistical Reasoning in Sports*, Bedford, Freeman, and Worth.

Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., and Vanderstoep, J. (2015), *Introduction to Statistical Investigations*, John Wiley and Sons.

Watkins, A, Scheaffer, R., and Cobb, G. (2004), *Statistics in Action: Understanding a World of Data*, Key Curriculum Press.

Zieffler, A., & Catalysts for Change (2013), *Statistical Thinking: A Simulation Approach to Modeling Uncertainty* (2nd ed.), Catalyst Press.
