Statistical Science
2016, Vol. 31, No. 4, 624–636
DOI: 10.1214/16-STS574

A Conversation with Jeff Wu

Hugh A. Chipman and V. Roshan Joseph

Abstract. Chien-Fu Jeff Wu was born January 15, 1949, in Taiwan. He earned a B.Sc. in Mathematics from National Taiwan University in 1971, and a Ph.D. in Statistics from the University of California, Berkeley in 1976. He has been a faculty member at the University of Wisconsin, Madison (1977–1988), the University of Waterloo (1988–1993), the University of Michigan (1995–2003; department chair 1995–8) and currently is the Coca-Cola Chair in Engineering Statistics and Professor in the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. He is known for his work on the convergence of the EM algorithm, resampling methods, nonlinear least squares, sensitivity testing and industrial statistics, including design of experiments, robust parameter design and computer experiments, and has been credited for coining the term "data science" as early as 1997.

Jeff has received several awards, including the COPSS Presidents' Award (1987), the Shewhart Medal (2008), the R. A. Fisher Lectureship (2011) and the Deming Lecturer Award (2012). He is an elected member of Academia Sinica (2000) and the National Academy of Engineering (2004), and has received many other awards and honors including an honorary doctorate from the University of Waterloo.

Jeff has supervised 45 Ph.D. students to date, many of whom are active researchers in the statistical sciences. He has published more than 170 peer-reviewed articles and two books. He was the second Editor of *Statistica Sinica* (1993–96). Jeff married Susan Chang in 1979, and they have two children, Emily and Justin.

This conversation took place in Atlanta, Georgia, on April 21, 2015.

Key words and phrases: Industrial statistics, data science, experimental design, computer experiment, EM algorithm, resampling methods.

1. EARLY YEARS

Roshan: Tell us about your early training and career.

Jeff: I was born and grew up in Taiwan. In general, my life there was very happy. Life was peaceful and education was nearly free. And there was no Cultural Revolution. We actually had a lot of political freedom, as long as we didn't get directly involved in politics.

I faced my first choice when I graduated from high school. Should I choose history or mathematics? I was very interested in history. I could remember historical facts very easily, and connect the dots with almost no effort. I also liked geography. These attributes are all connected with being an historian. However, I decided not to study history because history can be political. And Taiwan at that time was still an authoritarian regime.

So I chose mathematics. I was equally passionate about this subject. I went to National Taiwan University for mathematics in 1967. Most of the instructors were very inexperienced. I'll give you an example. In those days, for the sophomore year, we studied "the three highs" (as they were called in Taiwan and world-

Hugh A. Chipman is Professor, Department of Mathematics and Statistics, Acadia University, Wolfville, Nova Scotia, B4P 2R6, Canada (e-mail: hugh.chipman@acadiau.ca). V. Roshan Joseph is Professor, Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA (e-mail: roshan@gatech.edu).

FIG. 1. *Jeff Wu in high school*, 1966.

wide): higher algebra, higher analysis and higher geometry. The instructors of each of the three courses only had a master's degree from National Taiwan University. They probably knew only a little bit more than the students. However, our textbooks were all in English and written by the best authors in US or Europe. These books were my real teachers. In my time, my classmates were very motivated and worked very hard. Remember these were the go-go years—Taiwan was moving up fast. The students ran the seminars and often we chose textbooks that were above our level. So we probably didn't understand everything, but we had a lot of enthusiasm.

Hugh: Students chose the seminar?

Yes. It wasn't even a credit course. We just did it. Although the university could not provide much in terms of faculty or facilities, we still did well. Now, it's a much better learning environment. Taiwan has professors who are mostly Ph.D.s from top places in the west.

Hugh: What was it about statistics that attracted you to it?

Jeff: I always enjoyed doing math in my undergraduate. But in my senior year, I had the fortune of meeting Y. S. Chow, who visited Taiwan from Columbia University. He gave a graduate course on probability theory. I took the class and was the best student. In fact, I solved a problem he posed. So I wrote it up and submitted it to the *Annals*, and eventually it got accepted. But unknown to me, he also sent my paper to a new journal in Taiwan (*Bulletin of the Institute of Mathematics*) at Academia Sinica (Wu, 1973). When I found out later, I had to withdraw my *Annals* paper because that one had already appeared. For a young person that's very sad. Remember, I was only studying a bachelor's degree at the time, before two years of military service. So that's paper number 1 in my CV.

Roshan: And it would have been an *Annals* paper...

Jeff: But it's okay now. That paper actually gave me my entry ticket to Berkeley. I was the first one from Taiwan who went directly from undergraduate to Berkeley Statistics with support, and it was not easy to get that. I think it wasn't just high grades, it was also the paper.

The exposure to probability led me to explore the field of statistics on my own. I had two years of military service, 1971–73, and I used my spare time to read Ferguson's (1967) *Mathematical Statistics*. I solved all the problems in the exercises, except maybe one, so I knew that I could easily handle mathematical statistics on my own. I think I may still have the solution book somewhere—when I moved I didn't throw that away. The second book I read, *Experimental Design* by Cochran and Cox (1957), was more intriguing. It had a totally different flavor from Ferguson. I must admit that I did not understand the depth of the statistical ideas and methodology in the book, but I knew right away that this was the kind of thing I would be good at. I didn't know much about statistics, but I liked the intuitive thinking that went with the methodology.

Hugh: So you appreciated the mathematical side of it but could see that there was some depth to the philosophical ideas.

Jeff: Yes, those ideas would be what distinguished statistics from mathematics, actually. I sensed that, without the ability to spell out why. Looking back, I think my interest in history played a role and was related to my later success in statistics: history often gives you a different perspective on research—either existing methods or in a new area.

2. INFLUENCES

Hugh: Who has been the most important influence on your work?

Jeff: There are many, but I will be brief and only mention three.

I start with George Box because he represented a big change. When I was a student at Berkeley, I was basically a mathematical statistician. I read a bit of other work, but the whole Berkeley training in those days was more mathematical. I moved to Wisconsin, and was immediately influenced by George Box. He was a great scholar and a great lecturer. His opinions and passion for work were contagious. I really learned the Fisherian view of statistics, and the Fisherian tradition from Box. He was a son-in-law of Fisher. Box and I were not very close friends. I always felt a bit intimidated by him, but I respected him a lot. So even if I disagreed with him, I would try not to say anything public, unless he attacked me very hard. I just respected him so much.

The next person is my advisor, Peter Bickel. I've never really done anything big in the areas in which he excels. But I want to recall from my student days, something that Bickel did, which I was very impressed by. In my second year of graduate study, Bickel was on leave, so I had some free time to look around and explore. I took a course on optimal design from Jack Kiefer, who visited Berkeley only for one quarter. Then I picked up some research problems. These weren't from Kiefer, but his course got me interested in optimal design. I picked up one problem on my own and one related problem from Henry Wynn. I had basically done all the work when Bickel came back from sabbatical. He gave me a different problem, a difficult one that I couldn't solve. So, with some trepidation, I showed him the optimal design work. I was a bit worried. You know, this is a big professor, and I did something totally different. I could see immediately that he wasn't upset. He was very happy that I did something new and not in his field, and he happily corrected my English. I was very thankful he was so supportive.

Hugh: And he recognized the importance of what you had done.

Jeff: I'm sure he talked to Jack Kiefer.

Over the years, Bickel and I became good friends—I'd say teacher and student, but also good friends. Bickel's professional standard and behavior made him a role model for me to follow. He was an old generation intellectual, one who held his academic standards very high. Conversations with him were always about research or some intellectual thing. When I had dinner with him, the whole evening was so exciting—talking about subjects ranging from current affairs to philosophy and history. Like me, he knew a lot about history. So that's Bickel.

Jack Kiefer and I only overlapped briefly. He was a charming person and a great scholar in his own right. He visited Berkeley for one quarter when I was a second year student. He changed my direction to optimal design. In my youth, I was fascinated by the name and because of the mathematics. But more than that, I have a belief that you have to do the design right—you have to collect data correctly before you can do inference. Whether I made a good judgment to study optimal design, I don't know. But it doesn't matter—you're young, you do something exciting.

Here's a story. Kiefer had actually already agreed to give a course on sequential analysis. When I saw that, I went to see him. I was young and perhaps bold or even reckless. I said "Different people can teach sequential analysis. But, with optimal design, you are the founder—the creator. No one else can teach optimal design as you would. Can you teach that?" He was a bit surprised. But again, he was very friendly and receptive. He said "Well, I can, but you have to line up eight people to take the course." So I went around to line up eight people to take it for credit. Class started, and I think that within a few weeks, about four of the eight dropped out. But still there were some visitors. I think three visiting professors (Alistair Scott, Henry Wynn and one more) stayed to the end. So there were four students and three visitors and of course, I learned a lot.

Jack Kiefer died very young, at 57. He was very supportive of me, always very willing to answer my questions or help. I remember once I asked him this question during my visit to Cornell: "I seem to be doing different things, different topics, is this good or not?" He said "If this is what your nature tells you to do, just do it!"

Hugh: While you were at University of Wisconsin—Madison, your research focus changed to include industrial statistics and quality improvement, especially the design and analysis of experiments. What was the research community in Madison like in the 1980s?

Jeff: Madison was a great environment with Box as the leading light. There were other key players in engineering statistics like Bill Hunter, Norman Draper and Brian Joiner. I learned a lot from them and I wish I had written a paper with them, but I didn't. These were my formative years, because I moved from a very mathematical environment in Berkeley to Wisconsin, which was more balanced towards methodology and applications. In that period, between 1983 and 1985, Taguchi's robust parameter design entered the US. It was a new challenge and inspiration for the Madison school. I was there from 1977–1988.

Madison was my lucky place, not only for my career. More importantly, I met a young lady called Susan Chang. We got married in 1979. Our two children Emily and Justin were born in Madison. Susan gave me a lot of free time to pursue my work because she saw the tenure pressure I felt, at least in the beginning. For the first baby I rarely helped her in changing diapers. I do not take it as a badge of honor, though.

Hugh: Your interest in design of experiments started when you read Cochran and Cox, and your thesis was on optimal design. How did your interest change at Madison?

Jeff: Although I focused on optimal design at Berkeley, I was also a teaching assistant for a graduate course in design. So I read classical design on my own. But it was in Madison I really started appreciating the Fisherian way of statistics, including design. It was there that I started doing my major work in design. That environment helped me. I was a bit more mathematical than others at Madison. So the flavor of my research was varied: I did some applied work, and I also did some very mathematical work, like minimum aberration designs and resampling inference like the jackknife and bootstrap.

3. PERSONAL HISTORY

Roshan: Would you like to share any interesting stories behind your important works?

Jeff: Let me start with the EM algorithm. Somehow this paper (Wu, 1983) gave me a lot of fame although I don't deserve it. When I visit some places, like a biostatistics department, people may not know my work, but they know me as the EM guy. I'm not the EM guy, I just happened to prove something.

So what happened was this: Persi Diaconis visited Wisconsin from Stanford. This was the first time I heard from him that in the famous paper (Dempster, Laird and Rubin, 1977) the cumbersome proof was wrong. I read the paper and quickly realized why the proof was wrong. I tried to start working with Wing Wong on a correct proof. He was from Chicago, but he came to Madison often. We didn't get very far, and we dropped it.

Later, I picked up the problem again on my own. There was one night I remember clearly: I was taking care of our daughter Emily (Figure 2), who was less than a year old. I was always thinking about this EM algorithm. I suddenly realized that there was a theorem I learned in Berkeley called the global convergence theorem. I already tried proving convergence in a different way, so the math was already in my head, I just hit the right connection. I left Emily, ran upstairs, and started writing. I didn't know what she was doing downstairs.

Where did I get this theorem? It came from the third year of my Ph.D. at Berkeley. I had very little to do—I was writing my thesis. I also needed to learn more about optimization for writing optimal design algorithms, which was about half of the thesis. So I took a course offered by electrical engineering. The lecturer used a book written by her advisor who was a professor in Berkeley's EECS Department. Besides the standard stuff we learned, the early part of the book has a

FIG. 2. *Jeff with Emily in their Madison home*, 1982.

result by Zangwill, called the global convergence theorem (Zangwill, 1969). Somehow I remembered the theorem, but until then it had been hiding in the background. Once I realized the connection, I was able to write the paper in a week.

Getting the paper published was frustrating. It was submitted to the *Annals* and was rejected. I quickly recognized who the reviewers were. The EM algorithm has a lot of predecessors, and so people took it very personally if someone claimed "I proved it." A reviewer said "No, no, no, it's already been proven," even though the original proof was wrong. I felt it was hard to fight a negative decision, so I submitted to the *SIAM Journal of Applied Mathematics*. It was rejected, and I could tell at least one or two referees made mistakes and said the same things as in the *Annals* review.

At the time, I was on the editorial board of the *Annals*. For some reason, I mustered my strength and wrote to David Hinkley, the Editor. I said I felt that this paper was not treated properly. I gave a really long reply and also rewrote the paper. And I requested that perhaps a new AE should handle it. So he sent it to a new AE. It came back quickly with very positive reviews and was accepted with minor corrections. The new AE knew about the difficulty of proving the theorem. When you have the paper in the right hands, you get accepted. This was the roller-coaster history of the paper. I always tell people, that if there is something you really believe, you should fight for it.

Roshan: Now, this is one of the most cited papers.

Jeff: Yes, of my own papers. But other people have papers with many more citations. I wouldn't rate EM among my five most original papers. I just gave a proof, but somehow it gave me early fame.

Among my really original work, I want to mention another one, and this one has a more interesting story. This was the paper with Mike Hamada in the *Journal of Quality Technology*, on complex aliasing (Hamada and Wu, 1992). For a long time, I knew that factorial designs can be classified into two types: regular and nonregular. A regular design has a nice, clear-cut aliasing relationship because of its group-theoretic structure. For other designs, I used the term nonregular, for example, in the book by Wu and Hamada (2000, 2009).

I was challenged by practical experiments involving nonregular designs like the 18- and 36-run designs that Taguchi popularized. In the summer of 1986, I was part of a delegation visiting Japan along with George Box, Vijay Nair, people from Bell Labs and others (Figure 3). I remember very well, one afternoon in Nagoya, hosted by the Central Japan Quality Association. The presentations were all case studies, and most of the orthogonal arrays they used were 18 or 36 runs. There were almost no other designs because Japanese researchers followed Taguchi and these were the two designs he recommended. As you know, Taguchi never advocated the incorporation of interactions in data analysis, so these analyses did not consider interactions. Yet case by case, the result was successful and I'm sure the analyses were correct. So I wrote in my notes, "Why?" I thought it may be related to the design-theoretic properties of nonregular designs. When you do research, you need to have a theoretical reference. Then, when you see some phenomena, you put it into that framework.

I came back to Madison, where Mike Hamada was a Ph.D. student. I told him about the idea and we tried to work on it. We didn't get very far. Partly, he was discouraged: he went to see some professors in the department, very well-known ones, and they all pooh-poohed the idea. They said "These designs, including Plackett–Burman, have complex aliasing, so they are very hard to analyze."

Then we both moved to Canada, to Waterloo. But I remembered the problem—I never gave this up. One day, I returned to the topic. Mike reminded me that when he did the analysis, he ended up with many, many models. Looking at the computer output, we noticed that many of those models were incompatible, containing interactions without their parent main effects. Quickly we asked, what if we rule these out? So he redid the analysis that night. When I met him the next morning, I saw him with a big, big smile. The final chosen model was clearly the best in two real applications. In my class on experimental design, I mentioned this new method of ruling out models. I didn't have a good name for this idea and told the class I'd give $20 Canadian as a reward for a good name. In the next class, Randy Sitter came up with the name "effect heredity" so I gave him $20.

So this was the story. This paper has about 300 citations (Google Scholar), but I don't use citations to judge work. I think we should look at the intrinsic intellectual value of a paper's ideas, not the number of citations. In this case, the paper was smoothly accepted, but our path to discovery was torturous.

FIG. 3. *Visit to Japan*, 1986. *From left to right*, *Raghu Kacker*, *Genichi Taguchi*, *George Box*, *Madhav Phadke*, *Anne Shoemaker*, *Vijay Nair and Jeff Wu*.

The moral of the story is the importance of having a sense of *reference*. At the presentations in Japan, I was sleepy during the summer heat. Then I heard something that caused me to immediately wake up. When you do research, you need to have curiosity but you also need to have a reference. If you don't have a reference, like knowing that design has regular and nonregular two classes, you may not recognize the opportunity. Since I knew that classification, it was the reference point for my curiosity. I didn't write up this idea of two design classes until later and people didn't classify it that way until the Wu–Hamada book.

Hugh: You were interested in both history and statistics, but chose statistics. How does your interest in history inform your view of statistics?

Jeff: Keeping a historical perspective is good for research. When I study an existing procedure, usually I happen to know who has done what, so I can quickly connect people and ideas. For my contemporaries, I usually also know their ability and personality, and see this reflected in their work. It's important to have the ability to see the significance of a piece you are going to produce and if some of the past work has become very important, why that work is important. This perspective will give you some idea about choosing good directions.

Some areas have more historical context than others. When I worked on the bootstrap, it was already important. Likewise, the jackknife and general resampling inference were important. Other new areas were not obvious in the beginning. For me, three examples of nonobvious areas are minimum aberration designs, robust design and uncertainty quantification (UQ). For a new area you may not have a lot of direct references, but you have related work by others. You need to be able to connect the old to the potential new work, and do a bit of speculation using your historical perspective.

4. BUILDING

Roshan: Tell us some interesting stories from your long professional career.

Jeff: Okay, some interesting experiences. I'll mention two.

In 1988, I moved to the University of Waterloo from Wisconsin. I was still young—not even 40. I had, at that time, a successful career. I received the COPSS award in 1987. Many of my American friends were surprised because the US was supposed to be the number one place. Why go to Canada? When I had interviewed at Waterloo, I found out that they would give me a very unique opportunity. The department at Waterloo would let me build my own research group. It wasn't a question of money.

In the end, I was there for only five years, but I produced some excellent students including Hugh. I always felt very warm about those five years. It was very rewarding, I worked very hard. For example, they gave me a really big office, and I requested two neighboring offices for all my students. No other professors did that. Few professors in Canada worked in the office at night, but in Waterloo, after dinner, I went back to work. I didn't expect it, but a lot of my students would work in the office, too, and I would talk to them until 11 p.m. I did something a bit different from others, and the department gave me all their support. Those were very memorable years.

My second experience is of a different nature. In 2003, I moved from Michigan to Georgia Tech. It was attractive in several ways. It was warm. I had been in the cold weather for so many years; I was a bit tired of that. I found the challenge to build a statistics program within engineering very inviting and intriguing. So I decided to give it a try. I think that in some ways I have succeeded. Let's put this in a historical perspective: I think this is the first time that statistics has had a big program, and a successful one in an engineering college. When I was in Wisconsin, I saw that George Box always wanted to build statistics in engineering. But the timing was not right then. In the old days, academics were much more territorial. With his background in chemistry, Box tried to build statistics in chemical engineering. I think that's not the right place. In my opinion, only in industrial engineering can statistics be built. Unlike the rest of engineering, industrial engineers do not do experimentation. Industrial engineers, operation researchers and statisticians have mathematics in common. The mathematics unites these groups, making a place for statistics to flourish. For example, at Georgia Tech, statisticians have our identity within Industrial Engineering. One may argue "How about computer science?" I don't think a statistics department can be built within computer science. The entrepreneurial spirit of computer science may be incompatible with the more measured approach of statistics.

I found the positioning of Statistics within Industrial Engineering very rewarding because almost every member of the statistics group in Industrial Engineering at Georgia Tech has collaboration in physical science, engineering or information technology. I see this as one of my most important contributions to the

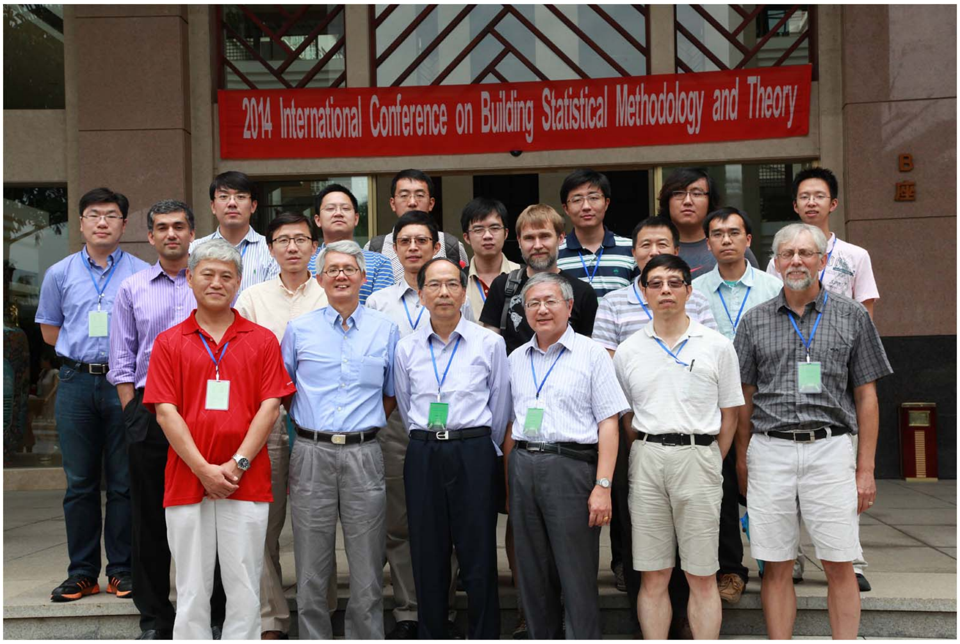

FIG. 4. *Jeff with some of his students and grand students in Yunnan, China. The picture was taken at a conference celebrating Jeff's* 65*th birthday, July* 2014.

profession. Statistics has found a home in other faculties. There are many biostatistics programs in schools of medicine or public health. There are a few statistics programs in social sciences. And now in machine learning, there are some collaboration between statistics and computer science, like CMU, Berkeley, etc. At Georgia Tech, our model is very different. I have found it very satisfying to be part of Industrial Engineering, instead of any other college.

Hugh: You've trained a large number of students. What role has that played in your career?

Jeff: First, I'd like to say how I educate my students. I follow the Chinese philosophy "Teach students according to their aptitude" (因材施教). I don't let my students compete. I don't let the next student do an extension of the previous student's work. If they did, they'd get into fights. I'm happy that quite a few of my students are working together.

What role have my students played in my career? Obviously, the most important is the intellectual satisfaction of seeing their success. I'm close to many of them. Some will call me up for advice on professional advancement: "Should I take this offer or not?" "How should I negotiate?" I am always ready to answer.

When you have a group of good people, working in a related area, they help advance the field. I'll name three such areas. First, my work in design of experiments, especially minimum aberration design, was like that. Quite a few students became major players. Another example is my work in engineering statistics, with a different group of students. In my latest work on computer experiments and uncertainty quantification (UQ), there is another group of students, this time having some overlap with the engineering statistics group.

Roshan: You played a role in the formation of the journal *Statistica Sinica*. Can you tell us about that?

Jeff: Yes, *Statistica Sinica*. Here is the background. A group of statisticians of Chinese origin, recognized strong growth in statistical research among their peers. George Tiao at the University of Chicago led this group. We saw a need to create a journal with Asian researchers as the main contributors, and that such a journal could also impact research in Asia. I followed that ideal when I was the second Editor. I told my editorial board that our dream was to be like an Asian *Biometrika*, although this would take a long time to achieve. That was my slogan.

I remember how we chose the name *Statistica Sinica*. We had a meeting at George's house for a whole afternoon, trying to choose a name. "The Chinese Journal of Statistics" didn't seem right. So Min-Te Chao used the Latin name: *Statistica Sinica*. "Sinica" is a lot subtler than "Chinese." Everyone jumped on it, and said "that's the name." Min-Te Chao was the Director of the Statistical Science Institute in Taiwan at that time.

We did not and do not want this journal to be an ethnic journal. I think we achieved that: we have many people who write and publish in the journal; there's no preference for any group. Obviously, there are a lot of Chinese authors, but nowadays, in the *Annals, JASA* and *Biometrika* there are a lot of Chinese authors, too. Well, there's an influx of many Chinese. After all, there's 1.3 billion.

Hugh: And Peter Hall was the editor also.

Jeff: Yes, exactly. That's a good example.

George Tiao was the founding editor and I was the second editor. I remember this was the most demanding time for my work. I was juggling three things simultaneously: editing the journal, writing the Wu–Hamada book on experimental design (the first edition was 650 pages) and chairing the Michigan Statistics department from 1995–98. I don't know how I survived. It was more rewarding for me to help launch a new journal and make it successful than being editor of an already established and top ranked journal. Generally speaking, it is very difficult for an editor to make big changes to an established journal.

5. INDUSTRIAL STATISTICS

Hugh: Quality improvement isn't as high-profile today as it was in the 1980s and 90s. Do core techniques such as design and analysis of experiments, reliability and process monitoring still have an important role in industrial areas? And as research areas?

Jeff: Let me start with design and analysis of experiments (DAE). I say DAE will never die. Why? Because this field always comes up with new ideas and tools to address a new challenge in practice. Let's witness the history: Design started with the work of Fisher and his collaborators in agriculture. Then in chemical engineering, there was George Box and the Wisconsin school. In the mid 1980s, there was quality engineering, inspired by Taguchi. I think there was a period from 1995 to 2005 when DAE was quieter, but it has recently emerged with strength.

I want to name two emerging areas in DAE. They give you some ideas about how design can rise to meet the challenge.

The first area is in Computer Experiments. Computer experiments employ space-filling designs which are different from factorial or optimal designs. Space-filling designs have been around for a long time. But recently, there are new classes of space-filling designs, such as nested designs. They were developed for multi-fidelity computer experiments. At a low fidelity, you have more points than at high fidelity. At high fidelity, the points are a subset and, therefore, nested. Another class of design is called sliced design, developed for computer experiments with qualitative and quantitative factors. For qualitative factors, you need to do slices. Both multi-fidelity designs and sliced designs were pioneered by Peter Qian (Qian, 2009, 2012).

As the field of uncertainty quantification is expanding rapidly, I think there will be other new methods, especially for very high dimensions. We know the way applied mathematicians do UQ; they use sparse grids. But sparse grids retain the tensor product structure, requiring a lot of points for even a 10-dimensional problem. In a real problem, such as designing a combustion system, can take 100–1000 input variables. There are no methods for choosing good designs in very high dimensions. You may choose space-filling, but this will not allow you to do fast computations. A space filling design abandons the tensor product structure, which allows you to do very fast polynomial approximation. Machine learning assumes that the data are already there and that data are cheap. But how about when data are *very* expensive? In some realistic situations, one finite element computation can take several weeks to run.

Hugh: This brings up a related question. In design for computer experiments, do you see computational methods or mathematical theory as the main way to generate designs? Or do they both have a role?

Jeff: I think both have a role. I believe for larger designs, computational methods will be more important. A good example is Roshan's recent work called MaxPro, which can be very useful for designs in high dimensions (Joseph, Gul and Ba, 2015). However, for the very high-dimensional case that I just posed as a challenge, I think deep mathematics will be required. That is, somehow you retain some tensor product structure to allow the orthogonal polynomial approximation, and also have some space-filling properties. Or, you use kriging to invert the correlation matrix. But you need to do inversion in a clever way. I have no clue how to tackle this problem. I think a good solution to this problem will be a major advance not only in computer experiments but also in applied mathematics.

The second example of an emerging direction for DAE is quite new. It is designed experiments for the internet, or e-commerce. I've just started working in this area. One of my recent Ph.D. students looked at it briefly in his thesis. I know that companies like Google, Amazon and eBay are using DAE. I think academics can do more fundamental or high impact work. I won't go into technical details, but in this case, the purpose is usually to optimize the revenue or increase the conversion rate, when the customer enters the webpage. You can use a factorial structure to change the web page design. But it is a lot more complicated. People keep visiting the page, so it's also a sequential design optimization problem. Industry currently uses stochastic programming to solve that problem, for example. I speculate this will be an area where DAE will go hand-in-hand with optimization and machine learning.

Let me move to reliability, the second area of industrial statistics you mentioned. Many people will say that reliability hasn't been active, with no breakthroughs and probably no future. I disagree on the last point. Let me explain the limitation of current research, which is a reason why I haven't worked in this area. I have long observed that most research in reliability is concerned with modeling "time to failure" data. However, for a high precision product, there are very few failures you can observe experimentally, so you have very little information. You can never get around this limitation unless you think differently, outside the box. That's the reason we have not seen much breakthrough in the last few decades. There's been good work, but nothing in the field that would make you say "Wow!".

I think the future for reliability would be bright if we can take a different approach. What if we use finite element simulation to generate failure data? Here's an example. Suppose you want to predict aircraft wing failure. Even with wind tunnel experiments, how many times do you observe failures? You have to do finite elements. Even finite elements will take a long time. It's a lot of repetitive computations until something happens. So instead, you could simulate material fatigue. Even if this simulation is very time consuming, it is doable using a fast machine or a cluster. You can generate some failure data from a computer model. Then you build an emulator, which is a surrogate model, on top of such data. This emulator can be used for reliability studies. I think this may be the way for reliability to go. If statisticians don't want to do this, some smart engineer with good mathematical and statistical training will be the first one to reach the goal post.

I am less sanguine about Statistical Process Control (SPC).

Roshan: You worked on optimal design in your early career but later you changed into other fields. What are your thoughts on optimal design?

Jeff: As I said, I was fascinated with optimal design and the terminology. I did a few papers on optimal design (including two in the *Annals*) in my early days and that's it. The reason for my change of direction was my very quick recognition that optimal design was so model dependent. In the early work on optimal design, for example, there are very few support points, not allowing for any model validation. In the first edition of Box, Hunter and Hunter (1978), there was a page (471–472) that scathingly attacked optimal design. This was a bit unfair, but they had a valid point. I felt optimal design was becoming more mathematical, but there were also some results that shed new light.

Here is an interesting twist of history and an example of how computing can be magical. With commercial software like JMP, optimal design has come back very strongly. Why? Consider when the design region is nonrectangular. I heard from a friend at Proctor & Gamble that he has rarely seen a rectangular design region; every region is chopped off in corners because of the constraints imposed by physics or engineering. Therefore, they almost always use optimal design. So after many years, optimal design has come back in a big way.

I want to point out something else. In the early days of optimal design, the most beautiful theory was the Kiefer–Wolfowitz (1960) general equivalence theorem, but this is for continuous design when the weights are continuous. This is the only case in which you can have such a general result, because you can take the derivatives, and with the derivatives you can do some calculations. If you have a discrete design, you have to do it case by case. Now, look at the real success of designs in JMP. They have nothing to do with general equivalence theorems. Of course I respect Kiefer a lot, but history has no mercy. Later on, the algorithm became key, and the algorithm is not necessarily very deep or fast, but it makes optimal design into a practical tool. Reflecting on this shift from theory to computation, my view of optimal design has changed dramatically over time.

Roshan: In this era of big data, do you think that small sample size DOE is still relevant?

Jeff: Let me first say that there still are important "small data" problems, and design is crucial there. In "small data" problems, the data are very expensive to get, like finite element simulation of a realistic system. For example, the simulation of an injector to ignite combustion can take weeks, so you cannot have many runs.

Let's return to big data. The concept of DOE, if not the specific details, will still be useful. For example, suppose you want to take a subsample. You don't do a random subsampling. You want to take the sample in a very clever way. This problem is totally different from survey sampling, which deals with human population or wildlife. This is a big challenge. I know some computer scientists are working on methods for subsampling, but I believe there will be further advances.

Another big data aspect is that having a lot of data doesn't mean that you can predict well. Companies may have huge databases, such as plant breeders like Monsanto. They use data mining to select potentially promising seeds. But field tests of the seeds often fail. It's not as simple as prediction and selection. For big data, even if we can do computation or inference, causality may not follow. We may recall similar challenges with causality in observational studies, which involve passive collection of data. A lot of knowledge has been built up in economics, in social sciences, in computer science such as the work by Heckman (1999), Pearl (2009), and Rubin (e.g., Imbens and Rubin, 2015). But big data is a different kind of phenomenon.

6. DATA SCIENCE

Hugh: You were one of the first people to use the term "data science." Can you tell us about how you coined the term?

Jeff: It was 1997 when I used "data science" in a public lecture (H. C. Carver Lecture at the University of Michigan). Beginning at least 5 or 6 years earlier, I grew dissatisfied with the term "statistics." It didn't capture all that we do.

The history of the word "statistics" is interesting. Going back to the root in Europe, its meaning was "state of the nation." When nations started collecting taxes, they needed to have a good count of the people they ruled. When you look at the translation of statistics into Chinese, Japanese or Hindi, they all mean the same thing. Here's a good example: In Chinese, this is very graphic. I have "statistics" and "accounting." "Statistics" I will write in Chinese "tongji" (Figure 5), which means collecting and counting. "Accounting" is in Figure 6. You can see that they share the same second character.

I remember in the old days, when I told people I was a statistician, they would say "Oh, you're an accountant." I was never happy about that. In many departments in Asia, in social science and business schools, accounting and statistics are one single department called Accounting/Statistics (会统). I was not happy because statistics is much more than is suggested by the name for descriptive statistics. I recall what Confucius said in his famous philosophy essay called Analects

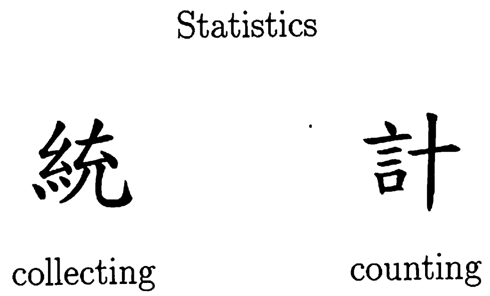

FIG. 5. *Chinese symbols for "Statistics."*

(論語): 名不正,语不顺. It can be translated as "If language is incorrect, then what is said does not concord with what was meant." I actually said this in my H. C. Carver lecture. This was my motivation; I wanted to change the name.

I had the idea of a better name, but the opportunity came when I had to give a public lecture to inaugurate the H. C. Carver Collegiate Chair. My lecture was entitled "***Statistics* = *Data Science?***." There I characterized statistics as a trilogy of data collection, data analysis and decision making. I was clearly talking about analytic statistics, rather than descriptive statistics. I suggested a change of our name from "statistics" to "data science" and "statistician" to "data scientist." I remember I even jokingly said in the lecture that by merely changing the name to data scientist, the salary will go higher. This is true nowadays. It's interesting.

Hugh: Should we all be calling ourselves data scientists and not statisticians?

Jeff: I think data science now has a much broader meaning, thanks to the influx of computer scientists who have the tendency and ability to take over a new field. My feelings are mixed: I think we need to keep our tradition, but we do not want to resist a new trend. We could call ourselves statistical scientists, but I don't think this term will go anywhere. "Statistical scientist" sounds a bit strange in the current environment. I prefer either statistician or data scientist. I notice that among my recent Ph.D. students who went into industry, all of them bear the title "data scientist." We cannot fight the trend.

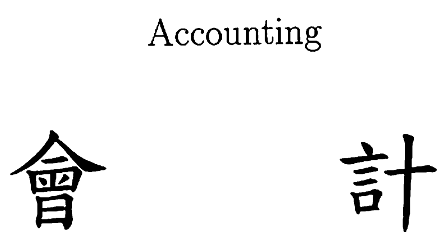

FIG. 6. *Chinese symbols for "Accounting."*

Hugh: Back when you gave the Carver lecture in 1997, you suggested some bold changes to statistics: making our education of statisticians more balanced and science-driven, putting greater focus on complex and large data and interfacing with other disciplines. Do you feel that these changes have happened? Is this still the direction statistics should be heading?

Jeff: Some changes have been made in statistics departments, but not on a massive scale. For example, I suggested changing the curriculum to be more applied or relevant to applications. I see one reason why changing to a more applied focus is difficult, and I do want to say it here. Our traditional curriculum requires three general areas: mathematical statistics, probability theory and computational/applied statistics. In many departments, there are three qualifying Ph.D. exams. But this can take up the first two graduate school years, simply to learn and pass exams. Once they are through the exams, students have very little time to accumulate research experience. Furthermore, most statistics students do not and cannot write big computer programs. I think this is an important reason why we lost to computer scientists in the big data challenge. Because we cannot handle big data, we fail before we can get started. I have, for a long time, asked the following question: Why should we insist that every Ph.D. student take and pass the qualifying exam in probability theory? The majority of them will go into industry. Even among those who become academics, only a minority work in probability research. We should have a more flexible system and course requirements in the qualifying exams. Some statistics departments in US and Canada have adopted these flexible systems, but the majority still have not.

Now the second part of the question, concerning the directions statistics should be heading. I think what I said then remains true, but of course there are new challenges in the last few years, like big data, and something I call physics and data-based statistical models in computation. Obviously, there are new challenges, way beyond what I could see at the time.

Hugh: In some of your work, there's an algorithm for solving a mathematical or statistical problem. Do you have any thoughts on the role that algorithms have played on your research and maybe the importance of thinking about algorithms in statistical research?

Jeff: We always need algorithms in statistics even if there is a theoretical solution. In most cases, the solutions are not in explicit form. Algorithms have become even more important nowadays. I'm not an algorithm person; my training was not computer science. I developed algorithms in many papers out of necessity. I believe when a problem or data are complex, there are few theoretical solutions. You need to do algorithms.

Roshan: And you prefer to solve using an algorithm than wait to solve it in a mathematically elegant way.

Jeff: Of course, if you do practical problems, you need to solve it right away; it cannot wait.

Hugh: So computation is part of a statistical toolbox. That ties in a bit with what you were saying about the training of our students in statistics.

Jeff: The current system does not allow enough time to hone computational skills. If you struggle to pass probability for one year, how can you learn to write code, or to master computation?

Hugh: Interestingly, even though statisticians have algorithms, we tend to be more tied to a theoretical framework. In contrast, a computer scientist will develop an algorithm very quickly, and if it's good enough, they're done.

Jeff: I have an explanation. In the statistical world, mathematical statistics is really dominated by asymptotic theory. You prove theorems. Why is that? Because, if you publish in the *Annals*, you have a better career. Some of the *Annals* work is very important. But other *Annals* work is not. Some work requires tedious and laborious derivations work, which I call "glorified epsilon-delta." But that's how the system rewards. When I listen to some talks, after the first 5–10 minutes, I know what the asymptotic results will look like. On the other hand, there are important theoretical advances. A beautiful theoretical characterization that sheds new light or resolves what we otherwise don't understand. So there are some good asymptotics, and some not-so-good asymptotics.

Hugh: So although theory is a mainstay of statistics and may distinguish us from other data scientists, not all theory is essential.

Jeff: I agree, we need to stick to our roots, because we need to understand some theoretical performance; otherwise, we are no different from computer scientists. What I'm trying to say is that there are some not so interesting asymptotic papers, and they are flooding the journals. They could create a lot of careers, but they don't help advance the field. When people do research, do they think in historical perspectives?

7. STATISTICS: PAST AND FUTURE

Hugh: How has the field of Statistics changed since you began at Wisconsin? For better or worse?

FIG. 7. *Emily, Justin, Jeff and Susan, Bali Island, Indonesia*, 2008.

Jeff: There obviously have been a lot of advances since 1977. Let me just name a few of many. On the theory and methodology side, we have resampling methods, especially the bootstrap, and we have causal inference. In applications, biostatistics has made tremendous advances since the early 80s. There have also been advances in computational statistics, like machine learning and widespread statistical software. And now, we have the big data challenge. So statistics have never been very quiet.

Hugh: What do you think are some emerging areas of statistical research or areas where statistics can make an important contribution?

Jeff: At my age, it is not appropriate to predict the future. Often the predictions would be wrong. So I'll only focus on one area that I know well. Earlier, I mentioned computer experiments and UQ. This research area is at the interface between applied math and statistics. A major part of UQ work is to build a computationally efficient emulator based on the simulation output from running, for example, a finite element analysis. Remember, the finite element method is used for solving a set of partial differential equations, which models the underlying physics. So I think one future direction, and a bright one, is for statisticians to better understand the physics and closely incorporate such knowledge in the work of modeling and computation. In traditional statistics, there has been some work on incorporating physics, but it is much more limited. Usually, you have some knowledge about the physics, and that puts some constraints on an empirical model, like nonlinear least squares. With UQ, it is different. To understand the behavior of the solution of the PDEs, you often have to know the physics. I see this as a really interesting new direction that goes beyond just UQ. Here, I would like to give a reminder. Most of the work in statistics is to build empirical models, including the current work in machine learning and big data. I am not saying empirical model-building is unimportant. It's important, but the physics-based data-driven approach will be important in the future.

Hugh: You say physics-based models. What about other areas of science? Do you mean specifically physics-based or science-based?

Jeff: I meant physical knowledge. This can be a biological mechanism, chemistry, molecular and so on. Also, in physical sciences, the underlying model can be quite deep, and we really haven't explored that.

Hugh: Are statisticians going to have to learn all this stuff or just get good collaborators?

Jeff: My own experience is that I'm not good at physics. It's important to get good collaborators, and learn from them. But someone like Roshan is different; he has some background in mechanical engineering. We also should bring more people into statistics with background in the physical world and the subject matter area, especially engineering and physics.

Hugh and Roshan: Thank you for giving us this opportunity to talk with you today.

Jeff: It was my pleasure.

REFERENCES

BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978). *Statistics for Experimenters*: *An Introduction to Design*, *Data Analysis*, *and Model Building*. Wiley, New York. MR0483116

COCHRAN, W. G. and COX, G. M. (1957). *Experimental Designs*, 2nd ed. Wiley, New York. MR0085682

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **39** 1–38. MR0501537

FERGUSON, T. S. (1967). *Mathematical Statistics*: *A Decision Theoretic Approach*. *Probability and Mathematical Statistics* **1**. Academic Press, New York. MR0215390

HAMADA, M. S. and WU, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *J*. *Qual*. *Technol*. **24** 130–137.

HECKMAN, J. L. (1999). Causal parameters and policy analysis in economics: A twentieth century retrospective. National Bureau of Economic Research, No. 7333.

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics*, *Social*, *and Biomedical Sciences*: *An Introduction*. Cambridge Univ. Press, New York. MR3309951

JOSEPH, V. R., GUL, E. and BA, S. (2015). Maximum projection designs for computer experiments. *Biometrika* **102** 371–380. MR3371010

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad*. *J*. *Math*. **12** 363–366. MR0117842

PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166

QIAN, P. Z. G. (2009). Nested Latin hypercube designs. *Biometrika* **96** 957–970. MR2767281

QIAN, P. Z. G. (2012). Sliced Latin hypercube designs. *J*. *Amer*. *Statist*. *Assoc*. **107** 393–399. MR2949368

WU, C. F. (1973). A note on the convergence rate of the strong law of large numbers. *Bull*. *Inst*. *Math*. *Acad*. *Sin*. (*N*.*S*.) **1** 121–124. MR0322935

WU, C.-F. J. (1983). On the convergence properties of the EM algorithm. *Ann*. *Statist*. **11** 95–103. MR0684867

WU, C. F. J. and HAMADA, M. S. (2009). *Experiments*: *Planning*, *Analysis*, *and Optimization*, 2nd ed. Wiley, Hoboken, NJ. (First Edition, 2000). MR2583259

ZANGWILL, W. I. (1969). *Nonlinear Programming*: *A Unified Approach*. Prentice-Hall, Englewood Cliffs, NJ. MR0359816