Statistical Science
2024, Vol. 39, No. 1, 192–208
https://doi.org/10.1214/22-STS878

A Conversation with Stephen M. Stigler

Sam Behseta and Robert E. Kass

Abstract. Stephen M. Stigler received his Ph.D. in Statistics from the University of California, Berkeley, with a dissertation on the asymptotic distribution of linear functions of order statistics. Starting in 1967, he taught at the University of Wisconsin, Madison, then in 1979 moved to the University of Chicago where he taught from 1979 to 2021. Stigler has worked on a variety of topics in mathematical statistics, ranging from asymptotic theory to the theory of experimental design, and on applications of statistics including in anthropology, forensic science, paleontology, psychology, information transfer and sports. In recent years, he has concentrated on the history of statistics, with inquiries ranging from the development of statistical methods in astronomy and geodesy and their spread to biological and social sciences, to lotteries, to the modern development of statistical theory. He has published four books, *The History of Statistics* (1986), *Statistics on the Table* (1999), *The Seven Pillars of Statistical Wisdom* (2016) and *Casanova's Lottery* (2022). A recent research focus has been upon the way the work of Francis Galton on the statistics of inheritance led to the creation of modern multivariate analysis and made a true Bayesian inference possible, and on how R. A. Fisher's transformation of Karl Pearson's path breaking research led to a modern period of statistical enlightenment.

Stigler is an elected member of the American Academy of Arts and Sciences and of the American Philosophical Society; he has served as President of the Institute of Mathematical Statistics and of the International Statistical Institute. In 2005, he received the Humboldt Foundation Research Award; in 2010, he was elected Membre Associé of the Académie royale de Belgique, Classe des Sciences. Stigler served as Theory and Methods Editor for the Journal of the American Statistical Association 1979–1982. He was a Guggenheim Fellow in 1977, and received awards for undergraduate teaching at the University of Wisconsin (1971) and University of Chicago (1998).

This interview with Stigler was conducted remotely in July 2021.

Key words and phrases: Statistical training, history of statistics, University of Chicago, academic life, statistical narrative.

EARLY YEARS: CARLETON AND BERKELEY

RK: Steve, take us through the early years.

SS: I was born in Minneapolis. At the age of five or thereabouts, we moved to Providence, Rhode Island, for one year and then my father moved to Columbia University. I don't remember anything about Minneapolis or Providence, which was a brief stay, but at Columbia, we lived out in Scarsdale, New York, and I went to a good school there. I still have some friends from there, but don't remember it very well. I left after 10th grade. My father accepted a fellowship at the Center for Advanced Study of the Behavioral Sciences at Stanford, so I spent a year out there. During that year, my father accepted an offer from Chicago, and we moved there for my senior year in high school.

SB: Your father, George Stigler, was a major figure in economics and a Nobel laureate! He was a fellow of ASA, and he wrote on the history of economic thought. Did any of that affect your upbringing and your thought processes?

SS: Personally, I had no idea, and the fact that he became a famous economist doesn't mean that he was such a famous economist for me, as I was growing up. He was

Sam Behseta is a Professor of Mathematics, California State University, Fullerton, 800 North College Blvd., Fullerton, California, 92831, USA (e-mail: sbehseta@fullerton.edu). Robert E. Kass is the Maurice Falk Professor of Statistics and Computational Neuroscience, Carnegie Mellon University, Pittsburgh, Pennsylvania, 15213, USA (e-mail: kass@cmu.edu).

FIG. 1. *Stigler Family in Paris in* 1955. *From left*: *George*, *Margaret*, *Steve* (*who set the timer for the picture*) *and brothers*, *David and Joseph*.

just an economist and we thought it was normal that Milton Friedman came by our house often and we called him "Uncle Milty." All we could tell is they were having a good time and he didn't work hard to explain economics to us, so we weren't getting any lessons, but we knew that he was with a bunch of economists, and they were having fun. The people in the statistics departments where I studied and first taught wouldn't have had any particular reason to have heard of my father anyway, and I never had any real sense of worry of thinking about having to follow in his footsteps or anything like that. It's just that I saw interesting stuff. He was interested in the history of things and had an immense library of books, a lot of which I've given to the University of Chicago's library.

SB: So then you went to Carleton College. You had a minor in history?

SS: My wife, Virginia, and I will be celebrating our 57th wedding anniversary in a couple of weeks, and dedicated my latest book to her as "my winning ticket in the lottery of life." She and I met in college. She was a year behind me and was a history major, so I took some history classes to be with her. In my senior year, I was in a seminar on the diplomatic history in the United States.

SB: Who were your early influences?

SS: I will say that a possible influence at one point was Kenneth May, who was a math professor at Carleton College. He was very interested in history and he left Carleton a year or two after I graduated and went out to Berkeley to retrain as a historian of mathematics. He then went to the University of Toronto and started a program in the history of mathematics and a journal, *Historia Mathematica*. May produced a huge volume of bibliography of history of mathematics. He was a very interesting fellow and someone who had an unusual life history, in that he got his Ph.D. degree with Jerzy Neyman out at Berkeley. He subsequently was publicly disowned by his father, who was a professor of history at Berkeley, because he was showing communistic leanings at the time of McCarthyism!

SB: How did you decide to go to graduate school?

SS: By then I'd figured that out by watching my father do what he wanted to do, which was spend the summers up in Canada, and have a lot of fun at work. I thought that sounded good. So, I was an easy mark for going to graduate school. But going to graduate school was not so easy in those days. We're talking 1963. The number of openings was not large.

SB: So, what inspired you to go to Berkeley?

SS: By exploring a lot of different areas in my first two years at Carleton, I had a rather miserable grade point average, but I had learned a lot, and not all of it academic; I played a lot of bridge. I was thinking of going to graduate school in statistics or math, because I was reasonably good at it and thought it would be fun. Guess who came to visit at Carleton College for three days and give three lectures and talk to people? Jerzy Neyman! This is late 1962 and it was some lecture series that he worked on by going to different places. He had a lecture style that nobody should ever imitate. He would call the students up to the board, and he would say, "now write the formula for a generating function," and they'd say, "well, I don't know what a generating function is." He'd hold the chalk in their hand and forcibly write the formula on the board, and that would get them engaged. During his visit, I got a chance to talk to him and mentioned I was interested in graduate study. He said, "why don't you apply to Berkeley," and Lucien Le Cam was chair. Well, I applied to Berkeley. I applied to a whole bunch of places. I got into two, and one of them was Berkeley, but I wasn't admitted until July. The other place I got into was Michigan where Jimmie Savage had recently moved. He didn't stay there very long before he went to Yale.

SB: And how was the experience in Berkeley?

SS: Berkeley was a fine place. I'd been reading Lehmann's book on hypothesis testing my senior year at Carleton, as a special project, and I'd been reading Loeve's book on probability that same year. It was sort of an eye opener in a way, when I came in to my first term at Berkeley. I did not have a scholarship or fellowship. I got a job grading papers for George Kuznets. He was a brother of Simon Kuznets who later won a Nobel Prize in economics. It was a course in agricultural statistics, if I remember correctly. I was doing pretty well in courses, but I looked around and every student was older than I was. There were guys there with master's degrees, there were people from overseas that had been doing statistics for a half a decade and been reading things. But what I soon discovered was that I had better preparation than they did. We'd learned how to do math; you know in basic ways, at Carleton.

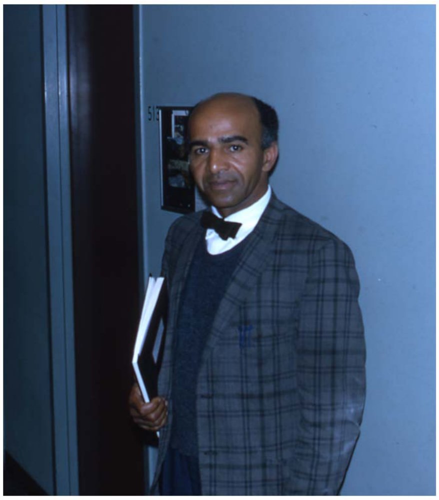

FIG. 2. *David Blackwell, in Campbell Hall at Berkeley in* 1967.

Lehmann, Blackwell, Le Cam and Hajek

RK: Lehmann seems to be someone who was really very attentive.

SS: Very attentive and very involved, He was a very good teacher and a very good lecturer, too.

SB: What year was this?

SS: I started in the fall of 1963, and I was at Berkeley for four years.

SB: Did you leave before the historic student protests at Berkeley?

SS: I would take my lunch over to the administration building, and sit out in the grass, and watch the police carrying the demonstrators out of the administration building. Things like that. I was not involved in that, but I was an observer. Mario Savio was around and lecturing at those times. It was very interesting. One of the things that stand out in my mind is I heard about a visit by Kerensky, who was involved with the Russian Revolution of 1917. He was in the student union building being interviewed by the students; he was based at that point at Stanford. And here was this guy who must have been very young at the time of the revolution, he wasn't very impressive in 1964, but it was sort of like a touch of ancient history, and you know it brings you a little closer to the historical times.

SB: That's remarkable, he was part of a transitional government that in effect was overthrown by the Bolsheviks!

SS: Yes, yes, this was like, wow! I had courses from Lehmann, I had courses from Blackwell who was an amazingly good teacher. I wrote a short piece for some math journal after he died. When he started talking about something in multiple dimensional spaces, he would use his hands to shape things out, and you saw those spaces. I don't know how he did it, but he had a way of getting deep ideas across. He spoke slowly and carefully, and he had a deep mind that got into all sorts of corners of mathematics. Anyway, Blackwell was immensely impressive—a genius, but also a great human being. The problem with working with Blackwell was that if you worked with him, you'd make an appointment to see him for five minutes once a week. There were so many people who wanted to work with Blackwell. And he didn't give you a problem, you had to find one. Some people worked with Blackwell for a very long time. Whereas with Le Cam, his door was always open. You could make an appointment, or you could just walk in. He was willing to talk, for as long as you wanted.

RK: And Le Cam was your Ph.D. thesis advisor, correct?

SS: Yes, but it was not from Le Cam that I got my thesis problem. It was from Jaroslav Hájek who was visiting for a year, and I really admired him. He was one of the people who processed Le Cam's work into a form that could be understood by mortals. Le Cam was wonderful and amazing and always available, and you could go in and talk to him, and the first 10 minutes would be absolutely wonderful, but after half an hour you'd be off in the second adjoint of a Banach space and lost. And it didn't help that in his approximate sufficiency paper, some definitions change in the middle of the paper. I have a copy of his dissertation that he sent me where he had annotated everywhere, everything that was wrong, and what parts I should forget about. He was full of ideas. He was absolutely brilliant but that doesn't mean that everything came out right, and he didn't always go back and correct it. I couldn't carry his ideas to the point where they would have been useful, so I gave up that project.

RK: But Hájek gave you your thesis topic?

SS: Hájek had a book with Šidák that was in press and he was lecturing about that, involving analyzing rank statistics asymptotically, by projecting them on the space spanned by sums of linear independent random variables, upon which the ranks were based. It occurred to me, sitting in his class, that it should also work with order statistics. I went and worked it out and did the conditioning that you have to do for the projections, and went and talked to Hájek, and he said, "that's pretty good, go ahead." I worked that out, and then took that to Le Cam because Hájek was going back to the Czech Republic (Stigler, 1969). I kept in touch with Hájek, though he didn't live long enough.

Charles Stein

SB: Was Charles Stein there?

SS: By then, he was at Stanford. I knew Charles because Berkeley students would go down there for a lecture each quarter, and there was also a regular visit at Berkeley with a speaker from Stanford. For each of these, there'd be a party afterwards where the graduate students could go and get free booze. So, we went and met faculty that way, but Stein was not a boozer. Years later, we

FIG. 3. *Lucien Le Cam, in Campbell Hall at Berkeley in* 1967.

gave him an honorary degree at Chicago. He came and had a great time, despite the fact everybody had told me that he didn't like personal honors. The thing I remember about Charles, though, is a time that I went down to a Stanford talk given by Hájek. Stein was in the audience. These were in small rooms, you know, there were not that many graduate students or faculty, maybe there were 30 people sitting around in loose chairs, and I happened to be sitting right next to Charles Stein. So, Hájek lectures in a very slow, deliberate way, and he went carefully through a series of steps involving the asymptotic behavior of maximum likelihood estimators. Charles tended to speak with a quivering voice. At one point, Hájek made some comment, and stated something, and Charles held up his hand and said, "excuse me, isn't it true that such and such, which I don't remember, would cause a problem with your theorem?" There was some silence (we students had no idea what was up), while Hájek was thinking, and he eventually said, "oh no because such and such." Stein went "Oh," and then I heard him muttering to himself under his voice, "you dumbhead!" Charles was a very modest man in some ways, but he knew a lot and did a lot of deep things.

RK: It's interesting when you talk about Blackwell, partly because I saw him lecture myself when we gave him an honorary degree here, and it was one of the best lectures I've ever seen, and inspiring and very consistent with what you were saying. Did you have a sense of those guys being incredible figures?

SS: No. One of the reasons is that they were mostly talking about their own stuff, so it was all very insular. We barely heard the word Fisher, unless it was combined with the word Yates, referring to the book of tables. I'd read enough, I knew more than that. Lehmann was a very good teacher, but not all of his students' dissertations were very exciting. I mean, he had the Hodges–Lehmann estimate based on rank statistics, and so then, well, the natural question was how about the Hodges–Lehmann estimate for the two-sample problem or in analysis of variance?

RK: A franchise was born!

Neyman

SS: And that's the way a school of thought works. And that was the powerful thing about Neyman. He came up with a way of looking at statistics that wasn't just about hypothesis testing, it was the work for a school of further work. It generated almost automatically a lot of problems, many of them interesting, many of them leading to wonderful corners. Fisher's work wasn't like that. If you were working with Fisher, then you're sunk. Fisher was not willing to admit that there was anything that could go beyond what he'd done on it, and maybe he was right. But with Neyman, he had ideas that could be applied to all sorts of different places, and it was like an open book for dissertations, and the people who wrote those dissertations filled the upper echelons of North American statistics and beyond, for a decade or more. When you talk about a school of statistics, it has to propagate through the students.

WISCONSIN YEARS—GEORGE BOX AND GRACE WAHBA

RK: The next step after graduate school was Wisconsin right?

SS: I was there for 12 years, though two of them were spent on leave. When I finished with my degree at Berkeley, we did not do what people do now, namely apply everywhere. I remember Erich Lehmann coming down the hall one day and saying, "do you have any interest in going to Florida State?" And I said "no." I didn't know anything about Florida State. This was not a value judgment. Anyway, people would write to the faculty and ask them, who should we be talking to? I ended up getting interviews at two places: the University of Iowa and University of Wisconsin, Madison. Both of them were very congenial places. They both offered me jobs and I went to Wisconsin.

RK: Box, I presume is the person who hired you.

SS: No, he had been the Founding Chair, but he was not the Chair at that time. I think Norman Draper was. They'd hired a bunch of people in previous years, and they were very active. It was a very democratic department.

RK: Was Grace Wahba more or less a contemporary?

SS: Her history is wonderful, but she was a late beginner. She was a single mother by the time she came to Wisconsin, which is the same year I came. But she'd been Manny Parzen's student at Stanford, and the things that she'd overcome! If you want to ask me about impressive people, Grace was absolutely one of them. George Box was absolutely one of them.

RK: Tell me more about Grace. You know, I TA'd for her and she was fabulous in multiple ways, as a person, and as a very knowledgeable and smart statistician.

SS: She was very independent. Most of the best students gravitated to Grace. Wing Wong, for example. She would have real problems and interesting ones for them. She was a fountain of ideas. I was there at the time that she was working with Kimeldorf on splines and cross validation and things like that. I was not part of the work, and I have no claim to any credit in that, but I was not just a distant person. My ears were open. She was very open to ideas, open to criticism. She was very athletic. You know that, until very recently, she would go on a long bicycle trip every year in Europe or in the United States! I was tremendously impressed. We got an honorary degree for Grace a couple of years ago at the University of Chicago and had a great time with her coming down here. George Box, on the other hand, was a very independent character, not on the same wavelength as those in the establishment of academia. He had some unexpected friends. Jack Kiefer was a very close friend of his. They were totally different in terms of their statistical thinking. But they got along wonderfully well. Kiefer was himself a brilliant lecturer and writer. The theory of experimental design was not the sort of thing that Box would work on. He would call it "these alphabet optimality conditions." Kiefer was a close friend, John Tukey was a friend. They had a mutual respect.

RK: How did Box end up at Wisconsin?

SS: Box had spent a year or two in Princeton with Tukey and declined an offer to stay at Princeton. Instead, he took an offer from Wisconsin. Wisconsin had been trying to form a statistics department for some time, and they made an offer at one point to Neyman, but it turned out Neyman was just playing, trying to leverage at Berkeley, and so he got the leverage at Berkeley and stayed at Berkeley. And they may have made some other offers, but then they got Box, and he was an extremely good catch for them: He built the statistics program. He got some very good students. George Tiao was there as one of his students. He was working on his time series book with Gwilym Jenkins. Box had an evening seminar at his house called the beer seminar. You'd show up at his house at eight o'clock and for two hours somebody would present a problem and people would chew it around and drink beer and he would come with comments; more like a consulting session. And this was wonderful to watch. But the thing that caught my attention was that, while he was working on the times series book, every solution that he proposed was a time series solution.

THE COHORT AT THE UNIVERSITY OF CHICAGO

RK: So, let's talk about Chicago. There was an amazing cohort of people there, in David Wallace's generation. Maybe we should list them to start with.

SS: Well, the way I would put it, there were six people that come into that category, really they are Bill Kruskal, Leo Goodman, Raj Bahadur, David Wallace, Pat Billingsley and Paul Meier. They all arrived in the 1950s. I think Kruskal may have been the first, and he certainly was older than the others because his own background was a little odd. He was one of three brilliant brothers. The family had a fur business. He stayed back and helped run the business with the family, while one after the other, the other brothers went off to college. And finally, it was Bill's turn, and so he was a latecomer in a way. But he was a very interesting fellow. When he was cleaning out his office, he would throw things out in the hall, and I probably shouldn't have done this, but I picked up some and kept them. There was one referee's file that I picked up. Someone submitted a paper to JASA, and Bill had been asked to referee it, and he wrote a very picky and long referee's report and in the report he was saying he didn't really understand why the results were true. So, this went off to the author, and he was incensed. The author said if he doesn't understand that, what is he doing refereeing to begin with? Anyway, he finally calmed down and wrote a revision. The revision got the same treatment, but now Bill asked different questions about other things, saying this doesn't really seem right to me or something. The author was again incensed. A second revision got similar treatment, but by now Bill understood what was bothering him. The author's response now was gratitude, with copious thanks to the referee, without whose help he might have published a paper with a serious error.

RK: There is a downside to that nitpicking personality. David's deep knowledge was hardly shared at all. And, I couldn't help feeling that, and it may be Kruskal and Wallace were similar in this way, that they were so critical of everything that it stood in the way.

SS: I think there's a lot to that and I think you're right about David. He had a lot to share that he didn't, and it wasn't because he was holding secrets. It was the case that he couldn't reach the level of excellence that he thought was needed. Allen Wallis was the editor of JASA for 10 years and Allen was one of the great delegators in the history of statistics. He learned this during the Second World War, at the Statistical Research Group at Columbia. Harold Hotelling was officially the head, but he delegated most of his decision making to Allen. And when Allen was editing the journal, if you look on the list of editors, David is not on that list, but he really was running the journal. There are different ways of editing a journal. You can be all consumed by it, in which case you're not going to do any of your own work. I was editor for three years, and I tried to compartmentalize, but heaven knows with 500 submissions coming in a year and revisions and all sorts of other things, you really are unable to give as much attention to anything else.

FIG. 4. *Table at the Quadrangle Club at the University of Chicago in 1973. From left: Paul Meier, David Andrews (hidden), Steve, William Kruskal, David Wallace and Leo Goodman.*

Leo Goodman and Jimmie Savage

RK: Goodman is an interesting character. I didn't know him hardly at all. I went to some of his classes briefly, but that was it.

SS: Well, I did know him, and we'd visit with him in California, almost annually, when we were out there for something else, and he was a very interesting person. He was a very good research statistician. He told the story about how he got into statistics. He went to Princeton as a graduate student in mathematics and he was interested in statistics. But he wasn't doing anything, and then one day he was at the mailboxes and Tukey came up to him and said, "what are you working on?" Leo said, "well, I'm just taking some classes" and then Tukey said, "look, I just got this this postcard from somebody asking me a statistical question. Why don't you go away and answer it and write it up and bring it back to me." And he did that, and it came back, and Tukey said, "Okay, I think with a little bit work you can publish that." Leo had connections in his life that you would never have guessed. He and his wife were very close friends with Sylvia Plath, the poet, and in her crisis period they were on leave in England and would drive her around a lot. He and his wife, Anne, are mentioned in her diary entries and the like.

RK: And he was something of a prodigy. Is that right?

SS: I don't know how to define that. Not in the Tukey sense. Tukey was homeschooled and never socialized with people and never saw the inside of a classroom until he was in college, which he finished off in two years. Leo was smart, definitely. He taught in Sociology. He came to Statistics department meetings, and he was present at department events. We liked him very much, but he didn't have very many Ph.D. students. The problem was he was a lousy teacher.

RK: I witnessed his style and it was, shall we say, unique.

SS: Well, I went to a big formal university dinner once and sat with a university trustee and his wife at the large table. She was a graduate of the University of Chicago and when she learned that I was a statistician, she said, "I took a statistics class," and I said, "who did you take it from?" "Leo Goodman" she said, "and it was the worst class I've ever had in my life." The thing is that Allen Wallis thought Leo was good, but Allen knew that Jimmie Savage was really special, and I'll quote one line from a memo. In arguing for raises to the dean, Allen wrote that Leo was a very good and productive statistician, but "Savage is clearly one of those exceedingly rare individuals who makes both a department and a university great." Jimmie was a pure mathematician when he was John von Neumann's research assistant at Princeton, and when he came to Chicago, he was sort of at loose ends. I have a huge box of Savage materials, which I someday might do something with, but they have some very interesting things about when he left our department in 1959 because his wife wanted to leave Chicago. A couple of years later, he was divorced, and he wanted to come back and the department wouldn't let him. He had been very rough on people for his last year there, and he was in emotional turmoil. You could excuse it but it was not his finest period, and still it was probably a mistake, because he had flourished at Chicago, and in fact, the Business School at that time made an offer to him which Jimmie declined, due to resistance in the Statistics department. In one of his memoirs, Milton Friedman said that in his whole life, he'd only known two people he would classify as geniuses, and both were statisticians: R.A. Fisher and Jimmie Savage. He later added Harold Hotelling and John Tukey to this group.

Raj Bahadur

RK: One of the most amazing experiences for me was to take this one class from Bahadur. It was a special topic class, and there were only five of us in that class and we all loved it. Wing Wong wrote up the notes, right?

SS: Yes, and when I was chair, we found somebody who was willing to type them up in Tex, and then we edited them a little bit, and it was published as an IMS monograph with a little bit of introduction. It was all done with elegant mathematics and far reaching, and if Wing hadn't worked hard at it, it wouldn't have been published.

SB: One question about Bahadur. Was he a contemporary of Mahalanobis?

SS: He was much younger, and there are stories that Mahalanobis was a bit of a tyrant. Raj did not get along well with him. He didn't fight back, but he made himself scarce. And that may be one of the reasons he came over to Chapel Hill for graduate study. It was there or at Columbia, where he visited for a while, that he met his wonderful wife, Thelma, in a class. She was an American woman, and their marriage caused some strife in the family back in India, and there's an interesting source on

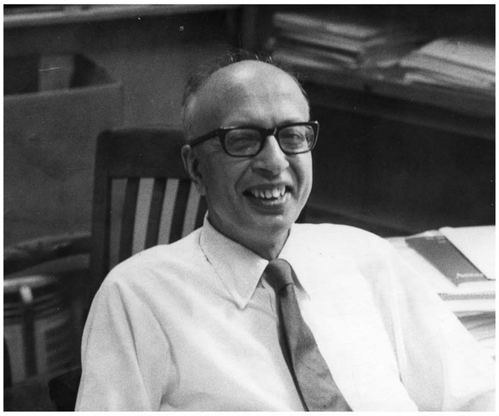

FIG. 5. *Raj Bahadur in his office, circa* 1980.

this, if you want to dig into it. There is another woman who has made a whole remarkable career on educating people on Indian cooking. Her name is Madhur Jaffrey. Maybe you've seen her cookbooks. There are lots of them. She was Raj's cousin, remarkable in her own right, and the reason this becomes relevant is that she, some years ago, wrote a memoir called "Climbing the Mango Trees," which you would not find easily in this country. I have a copy somewhere. And it's priceless; it's an unblemished story about her interactions with the whole family, including Raj and his father.

RK: The other thing about Raj was that he was a man who exuded dignity.

SS: One day, there was a distinguished Indian visitor, and he hung out, but for the evening he was going to go to Raj's for dinner. And I was in charge, not going to the dinner, but of delivering him to the door. Driving him to where Raj was living at the time, I wanted to make sure that he got in the house, and so, who comes to the door to meet him but Raj, and he was wearing a flowing white robe of the caste, which was serious stuff. I had never seen him in anything like that, but for this fellow he felt that was the right thing to do and he did it.

Kaplan and Meier

RK: So, there's also Paul Meier.

SS: He came to Chicago late. Paul was a brilliant biostatistician and also a hard person to argue with. Let's put it that way. He came with a contract that was half in the medical school and half in statistics. Also, for a while he was chairman of the department, and he did a lot of good things. Paul had very strong opinions, which he didn't keep to himself. He did not get along with Jimmie Savage. And I'm not saying that Savage was always right—sometimes he was wrong. I got along well with Paul. His most famous paper was coauthored with Edward Kaplan. Kaplan was an applied mathematician, at Bell Labs at the time. Both Kaplan and Meier wrote papers on survival; Paul's had a medical focus and Kaplan's considered telephone cables. Kaplan had learned about Paul's work from John Tukey at Bell Labs. They both submitted to JASA and Allen Wallis asked them to revise as a joint paper. Any way, they ended up being put together, in a marriage, and Paul thought that, well, his was the important part. But I never saw the two parts separately. One of the things that paper did was it went beyond mere biological applications, and that made it something that became fantastically successful and one of the most cited papers of all time until Cox regression came along.

Pat Billingsley

SB: Wasn't Billingsley involved with theater?

SS: Yes. While he was here, he would get active in the theater and then one day somebody came up to him and said, "would you like an agent?" And once he got an agent, he appeared in some major motion pictures, including "The Untouchables." There were some small motion pictures, where he had bigger parts, but he was a very good actor. He did this for fun. When they were making a movie in Chicago and were looking for somebody for a small part he was around. He was a very good teacher. He had been a student of Feller's. Pat had a good sense of humor, and he could spot an odd statement very easily.

SB: It's interesting because in his interview, Rob said he considered Feller's book the best statistics book ever written.

SS: Feller's is a great book—I absolutely agree, except that if you go to the second volume, it is an amazing mishmash of things that have been put together. The first one is like poetry or Mozart. The second one has got brilliance and wonderful things, but if you try and read it through, you've got to dive in and look carefully at different things. I've taught from parts of it and got a lot out of it, but, it is a mess.

RK: I could see that it doesn't easily lend itself to teaching.

SB: Billingsley's measure theory book is a masterpiece.

SS: Oh yes! I think he was a better textbook writer than he was a researcher. He was a wonderfully clear expositor in his book. It was written with clarity that makes you appreciate it more if you've tried to wade your way through some other books, as I have.

RK: His big thing, as I recall, was pushing as far as he could the use of indicator functions. He really felt that intuition should be built up from simpler building blocks. That seemed like a unique presentation.

STATISTICAL IDEAS AND HISTORY OF STATISTICS

RK: Steve, the driving force of your career, at least after getting tenure in 1971, has been to describe the development of statistical ideas through historical research.

SS: I didn't really switch to history directly in 1971. It was more general over the following decade, but it certainly shows up in some of my papers. I remember the first thing that caught my eye was something that Peter Huber wrote in a paper about M estimates and tracing them back to trimmed means, and to France in the 1820s. I said, "hey that's interesting" and so I looked up a few things, and some of this got into one of my papers in 1973 on history of robust estimation (Stigler, 1973a). I was working on robust estimation in odd ways through much of that decade. As I got a little deeper into the history by 1978–1979, when I was out at the Center for Advanced Study in the Behavioral Sciences, I was really starting to work on my book on the history of statistics.

RK: There's a part of people's attraction to history that I don't share, and I don't know if it's because I don't appreciate its importance, or maybe it's just not my taste. On the other hand, I do find history to be compelling when it helps us better understand the important ideas behind conceptual frameworks by seeing the way the concepts evolved.

SS: What you're recognizing is that deep questions are not always apparent at the first sight, and one of the things that got me interested in this was that when I was at Berkeley, we very seldom heard about anything before 1950. Every curriculum in the country right now is mostly concerned with things since 2000. So, that's the way it goes, but then I started finding works from ancient times that were totally surprising to me. I could give you a couple of examples. One of my earliest discoveries was when I was on leave at Chicago. I found two odd things in the works of Laplace (Stigler, 1973b). One was something that looked like a partial anticipation of the concept of sufficiency, and one was an asymptotic derivation he made of the joint distribution of the mean and the median in a certain situation. And I hadn't even seen that in my graduate classes. And I thought, this guy lived in 1818, and he had thought about those! The method of least squares was invented in 1805 by Legendre. And then regression was invented by Galton, and now wait a minute, that was the 1880s! How could that be? Everybody knows that regression is least squares, right? But regression isn't least squares! So, trying to understand that change, and how that came about, and what kind of problems led people to ask the questions that prompted them to make that kind of advancement were most interesting to me at early stages.

RK: What you're saying I think is that it's really important to understand the context of everything, and that this tells us a lot. It would make sense to talk now about your book *The Seven Pillars of Statistical Wisdom* (Stigler, 2016). I first heard about the Seven Pillars in your 2014 ASA presidential lecturer at the joint statistical meetings in Boston, but I'm sorry to say, from my too-quick initial reading of the book a couple years later I didn't appreciate the depth of your arrangement of ideas. The first question is, how did you decide on this kind organizational device and how did you come up with a number seven?

SS: The explanation for the Seven Pillars, or at least the exact framework, the number, is going to probably disappoint you, but it's as following: I was teaching, and I got to the end of one class, and I thought, well, maybe I should try and isolate what are the major ideas in statistics. Now, this was at a time when David Letterman opened every one of his late nights with a list of 10 things and I thought about the 10 great ideas in statistics. But the attention span isn't up to 10. I actually started with five. And then I found that didn't cover enough, so I expanded it, but then I thought seven is a good number, as in the seven pillars of wisdom.

RK: Sam and I were talking about how it's a very unusual organization actually.

SS: Yes. I started realizing that there are some very simple ideas that carry through generations of statistical work, and then the first one was on the combination of observations. How about a simple average? Well, the least squares estimate is a complicated average. Why is this an idea? It's an idea because not only then, but still to this day, there are people who deny that averaging is a good thing to do. An average pushes out of sight the individuality of the different numbers. And the idea that by throwing away information, the information about individuals, you could actually advance your understanding is wonderful. That's still a challenging thing, and when it comes to a new field it's not old history, it's a real step that is not always a good one to take. So, I started working on other things and developed what made up that book. And then I had given a talk at a couple of places on five great ideas in statistics, and I was invited to give the President's Invited Address at the ASA meeting in Boston. I developed the talk and then turned the talk into a book.

SB: The seven pillars are: combination of observations, information, likelihood, intercomparison, regression, design and residuals. However, as for the residuals, you're not viewing it from the angle of testing the assumptions of the model.

SS: Yes and no. The concept is more general, but it's similar. This actually goes way back into the philosophy of science and John Stuart Mill and other people like John Herschel. You can explain as much as you can with a scientific model and then ask what you didn't explain. What goes beyond that? For example, look at Cox regression. You can see it as starting with a basic survival function and building it up in terms of a simple linear model that we can handle. That's a residual sort of thing: you take something, and look at the difference between what you can get one way and what you need, and then you add to the bit you already have to get something new. This concept appears in all sorts of scientific investigations; it's a basic tool, thought by some of those early philosophers to be the basic approach to building science. But that's still pretty vague. It's more constructive to see one step and ask why a model doesn't fit and what is left to be explained.

SB: And then you talk about nested models.

SS: Yes. Fisher discovered that, in a sense, we can only deal with nested models. We do not have a single, unified theory for dealing with nonnested models, especially in the normal world, the mathematics of looking for additional dimensions in a nested model is beautiful and one of the great glories of statistics, even though model assumptions aren't always met.

Models and Parameters

SB: Since we're talking about models, do we know, Steve, who came up with the term statistical modeling?

SS: You know, Herb David was at Iowa State and wrote a whole book on first appearances of terms. And then people added to that, and so there may be an online source for that sort of thing now, but in a way it's not as interesting to me as a number of other statistical terms that signal a real change. Like when Fisher started using the word "parameter." It was a major change in how you think about the model. Fisher was looking at parameterization in a way that he could get mathematical handles on it and do things that nobody could do before. And that was the beginning of modern estimation theory. Sometimes terms can signal something that is going on, but the term itself doesn't cause that.

RK: Fisher's new view of parameterization is a good example of how history can change your appreciation of a concept. It was you who kind of unearthed this fact. I experienced it by reading what you wrote about Savage's rereading Fisher, and I just found it amazingly interesting. But I guess the question is, how did that discovery make you change your notion of what you were thinking about parameters?

SS: I was asked to be a discussant (Stigler, 1976), and Jimmie was no longer living, I never met him personally. I heard him lecture once but I had never met him. When you're discussing somebody's paper you want at least to come up with something different to say, and maybe a criticism or something like that. I spent a huge amount of effort, looking at a pre-1921 statistics literature everywhere; textbooks, articles and essentially nowhere did I find the word parameter. I think I cited a couple of exceptions, but they were talking about a different use of it. Then I thought about the role it played in Fisher's work. By the time I got through reading a lot of Fisher, I half-thought I really understood how he was thinking. I looked at enough of his theoretical expositions, and I was beginning to get inside of the kinds of approaches he was taking and, actually, here it comes back to residuals again. I discuss this in one of the papers where I go through Fisher's approach to maximum likelihood (Stigler, 2007). One of the ways he analyzed the efficiency of maximum likelihood invoked the idea of residual by considering how much information is left after you've got the score function. And I was beginning to see how everything fits together. This was a signal of a way he had of grasping things and carrying them to new levels.

RK: The contrast with Karl Pearson, Fisher's predecessor, is particularly striking to me because Pearson had this whole family of distributions, and when I was a student we still studied that, though I think the topic got dropped from curricula pretty soon after.

SS: Well, in Pearson's family, they were frequency constants, and it was likely the standard deviation and the mean and those could be parameters, if you have the family. But they aren't parameters. They are a quality, a measure of the distribution and Pearson was just thinking about it entirely differently.

RK: You know, the geometrical framework I worked with you in my Ph.D. thesis, as in Kass (1989), emphasizes two fundamental properties of parameters: they identify distributions in the family and they allow calculus to proceed, especially for likelihood expansions. An immediate consequence is that procedures can take place using arbitrary reparameterizations. Fisher recognized this when he considered the parameter-invariance of maximum likelihood to be a good thing. In the geometrical framework, parameter vectors become coordinate systems, and invariant methods get called "coordinate-free." In statistics, coordinate-free approaches appeared in some of Bahadur's work, and especially in Kruskal's abstract reformulation of ANOVA (Wichura, 2006), both of which were influenced by the coordinate-free treatment of sufficiency by Halmos and Savage (1949). Halmos clearly influenced Kruskal through his coordinate-free treatment of linear algebra (Eaton, 2007). Didn't Paul Halmos overlap with Kruskal at Chicago?

SS: Yes, but he left in 1960, just after Jimmie Savage left. By then, he felt Chicago mathematicians didn't value what he was doing.

RK: I would add that, in the preface to his book, Halmos said that he was motivated by the desire to make connections between finite-dimensional inner product spaces and Hilbert space. One of the early great successes of the Hilbert space framework was von Neumann's axiomatic treatment of quantum mechanics (von Neumann, 1955) and we come full circle when Halmos says his biggest personal influence for the book was von Neumann, whom he had known, and who impressed on him the value of coordinate-free approaches. The interesting thing is that Fisher seems to have, at some level, intuited the importance of parameters, even though he couldn't have seen all of these connections.

SS: Oh Fisher certainly saw the value—by limiting attention to a smoothly indexed family of distributions you could say much more than with no limits. You could do math. But it was key that the limitation is not too severe: the parametric family has to be flexible enough to cover a broad class of distributions.

RK: Another aspect of the geometrical framework is that parameters give a clear notion of dimensionality, and then when sufficiency holds, for exponential families, the natural sufficient statistic has the same dimension as the family itself. Geometrical thinking comes up again with degrees of freedom.

SS: In one of my papers (Stigler, 2008), I was able to point to exactly where Pearson went wrong on Chi-squared degrees of freedom, and how this idea was absolutely the key thing that he missed, and then Fisher used exactly that part of Pearson's work to try to show where he' had gone wrong and exactly what was missing. That term shows up beautifully in his expansion and accounts for the missing degree of freedom. Pearson made important errors and they got Fisher to do really great things. Fisher recognized how he'd gone past Pearson in a way that Pearson never understood, and Pearson was nasty to Fisher, which made Fisher resentful. However, Fisher never understood the intellectual debt he owed to Pearson, and how much he built on what Pearson had done. You can say Pearson made mistakes, but they were pioneering mistakes, and when you go out for the first time into a field, you're going to make mistakes, as Pearson did. Then Fisher comes along and figures it out and doesn't credit the guy before him.

Fisher and Le Cam on Sufficiency

RK: You just mentioned sufficiency. You wrote about the way Laplace followed a path similar to Fisher's but failed to arrive at sufficiency, as Fisher did (Stigler, 1973b). I have a question about sufficiency. Because of something I was working on recently, I realized I wanted to tie the concept, the fundamental idea behind it, more strongly to exponential families. So, I'm curious whether your historical perspective is consistent with that, or whether it deviates from it.

SS: Well, it certainly is consistent with that. I got interested in sufficiency at Berkeley and because it's such a beautiful notion. When you have sufficiency, it is a perfect answer. One time I started to write a book, which I never finished, on the design of experiments, where my starting point was not the usual one. It was Blackwell on information and games. There are cases where, given the model, the entire information is summarized in a statistic, and there's no way you can learn more than that from the data. It is the absolutely complete answer and Jack Kiefer realized that this came in beautifully in some design of experiments' situations. If you have two multivariate normal distributions and the difference of covariance matrices is positive definite, then one of them is more informative than the other for estimating the means. So, one experiment can be sufficient for another, and that was going to be the beginning of the treatment.

RK: Isn't this related to Le Cam's work?

SS: Le Cam went further. He came up with approximate sufficiency where he was going to say, what if you can't quite reproduce all of the experiment? Then instead you reproduce part of it. For bounded loss functions, you could come within epsilon of the optimum and things like that. This would be a way toward building a design of experiments approach where you want to have an experiment that is so good that it comes within a certain band of optimality for all different possible questions you might want to answer. But sufficiency is a beautiful idea. Fisher really nailed it early on.

Bayes or Laplace?

SB: So, Laplace didn't come up with sufficiency. But didn't he get Bayes' theorem independently of Bayes?

SS: Yes and no. I think in a certain sense, he did, but I don't think he saw Bayes' work for quite a while, and he came up with a way of doing inverse probability that was really based on an almost fiducial assumption. He only had it for that one prior, and Bayes did not have anything for general priors either. The idea of dealing with equally likely possibilities was unique to Laplace. It was really all flat priors for a very long time in Bayesian inference. I did a translation of Laplace's 1770s paper where he introduced his pseudo-Bayesian stuff and he went on to apply the method to location and scale parameters, too. But he made an error: if you're dealing with a single parameter, things are a lot easier, just as Fisher could do fiducial inference pretty well with one parameter. Laplace had two parameters and he got a multivariate distribution. In the univariate case, you needn't bother calculating the conditional distribution by dividing the joint by the marginal. You simply get something proportional to the bivariate distribution for that particular variable, and you normalize. But with more variables the normalizing constant may depend upon the other variables when conditioning. For some reason, Laplace slipped up when he did a substitution in a particular equation, where the constant of proportionality was involved, and it didn't work! I pointed that out in a paper (Stigler, 1986b) and it reveals a serious limitation to his Bayesian multivariate analysis, which was otherwise going in the right direction. The historical takeaway is that multivariate distributions weren't very well understood then.

SB: So, should it be Bayes's theorem or Laplace's theorem, or does it matter?

SS: I have a little paper somewhere called "Who discovered Bayes's theorem" (Stigler, 1983). An important person in the history of psychology named David Hartley, published a book in 1749 where there is one page

FIG. 6. *Steve at Bayes' grave in London*, 2015.

where he talks about de Moivre's theorem and then he says in the next sentence that an "ingenious friend" had told him about an inverse to this and he describes what we call Bayes' theorem! Who's the friend? I played around with different possible people who it could have been. The blind mathematician, Nicholas Sanderson, could have been there. Then I came to a better understanding of where Bayes' theorem came from by going back into Bayes' paper, and I wrote this in *Statistical Science* about three years ago (Stigler, 2018). A large part of Bayes' paper, including every numerical example, was written by Richard Price. The stuff with the billiard table was Bayes' definitely. But the examples are all Price, and so who is using this for inference? I think Price! Laplace's first pronouncement was totally independent of Bayes. I don't think Bayes' paper had reached his gaze. One reason is his approach is totally different. Laplace's is much more like fiducial. In fact, there's a certain fiducialism in Bayes, too. George Barnard thought that Bayes was really a fiducialist. In all of those early works, the only priors you have are flat priors and Laplace was looking at proportionality in different ways and not developing it in the same sense for the binomial. He learned about Bayes a few years later and recognized it and some other writings later on. A lot of people read the first part of Bayes, but if you try and follow through the mathematics, which is correct, it's written in an ancient form that is really hard to get into, and so it's an unread paper in its totality.

RK: We were at a picnic when I was in graduate school, and you had a T-shirt on that said "Laplace's Best Friend." What was it that made you become his best friend?

SS: A friend of mine in Paris sent me a picture of this painting (Steve shows a portrait of Laplace via Zoom) and gave me a URL to check if I wanted to see it. He said, "It is amazing, nobody has seen a young man, a young Laplace before!" The most famous painting of Laplace before this was painted 12 years after he died. But here he is as a young man in 1784. He is 35 years old! So, I went to that URL, and it was an art dealer in Paris, and they were still in a recession there, and things weren't selling well, and no museum had picked up on this and I bought it. This painting is in another room in this apartment I'm in. If you look on my website, people who have come here, I photograph them with Laplace and there is a place on my website where you can look up "Laplace and friends" and you'll see about 30 people posing with Laplace. Laplace has come back into my life there, but he figures so much in the history of statistics, because he really was an amazing scientist. He comes into all sorts of different parts of the philosophy of probability, and he's still readable, but he was singularly focused on science. I reviewed a biography of Laplace a couple of years ago, where the author went to great lengths and came up with some theories about how Laplace was influenced by his teachers, etc. Yet it wasn't a very interesting book because it left out the science, and it seems all Laplace did was science, so hardly anything about his personality came through.

SB: Laplace worked on the so-called Legendre–Gauss least squares problem, he derived the central limit theorem, which I think he worked out better than his predecessors, at least in contrast to de Moivre's version, and he worked in almost every scientific area of his time. I mean he even worked with Lavoisier on certain chemistry problems, which makes me think the word polymath maybe applies to him. But is this because at the time, science was too broad of a field? These days, brilliant people work on a tiny little narrow area of some discipline. But Laplace was moving about in whole areas of knowledge!

SS: You're right in all respects. A whole group of mathematical advances were coming together and allowing people to handle problems that they hadn't been able to handle before. Laplace proved to be a master of doing this. Not everybody could see that way of putting things together. The Newtonian program had sort of stalled by then, partly because it became very difficult if you don't have a mind like Newton's, who could do things that people have trouble reproducing today. I mean, the elegance of his geometric constructions is just amazing. By the late 1700s, mathematics was developing very nicely, and new problems were being discovered and you had all these great problems, and it all came together, and Laplace completed the Newtonian work with his *Mécanique Céleste.* He got interested in probability and wrote this great book on probability that almost nobody has read. (It has some good stuff buried in its different places, including the central limit theorem and asymptotics.) Laplace was called the Newton of France.

SB: So, Laplace was a genius scientist who worked on many problems of interest to the history of statistics, but from the statistical perspective, his major contribution seems to be inverse probability.

SS: Oh, more than that.

I.J. Good

RK: Steve Fienberg's paper *When Did Bayesian Inference Become "Bayesian"?* (Fienberg, 2006) uses this terminological question to trace the rise of Bayesian statistics. The immense influence of Jimmie Savage is apparent. Another really important personality is I. J. Good (Banks, 1996). In his paper where the term "Bayes factor" was introduced, Good referred to the approach he was discussing as founded on a "neo-Bayesian or neo/Bayes-Laplace philosophy."

SS: Jack was an interesting character. I have a foot thick file of correspondence with him now, that's not all letters. He would send me bibliographies of his own work and copies of his papers, and every time he did a new bibliography of his own papers, he would include the past bibliographies in them in that list! There was a danger of getting a critical chain reaction, you know. But he was a thoughtful, and very smart fellow, and I once was invited to go to give a talk at Virginia Tech. This was when Jack Good was there, and I thought, well, I have to tease him a little bit, so I gave the talk with a title "Great Probabilists Publish Posthumously." You know, there's Bayes, there is Bernoulli, both of their great works were published after they were dead. I gave this talk to prod him, with the implication that greatness requires being dead when you publish, and he was absolutely cool and calm and got up and made some intelligent remarks and we engaged in a little bit of banter. He was well grounded, and I had other correspondence with him on different things, but he was somebody who had a good sense of what things are. We almost hired him at Chicago. In the 1950s, Jimmie Savage wanted to hire Jack. He was still in England at the time, and I've got a file on this. He kept making requests and finally he accepted the offer, and then he put it off a little bit, and then we got a letter saying that he was afraid he had to decline or withdraw, because his aged mother would not be able to get along well without him in England, but not too long after that he was hired by Virginia Tech, and as far as I know his mother was still around. They asked Jack, what it would take to get you, and the answer was, well, I could come there, and I could do whatever I want, and I would never be required to teach! I thought that was a monumental mistake on both sides.

RK: Jack Good was super influential and I met him in Chicago a couple times. He was also a great conversationalist, and he had a lot to say.

SS: He was in on a lot of things. But on the other hand, he could have done a lot more with some more discipline to his work. He was spinning off small bits and pieces and they're intelligent and they're interesting but to accomplish a lot you've got to really sit down and work at it.

RK: I couldn't agree more. I think the reason you're bringing this up is there's a way in which Good was a seriously major figure in that period and yet there's also a way in which influence could be lost.

John Tukey

RK: I recently read a biography of John Tukey in Statistical Science (Anscombe, 2003). It was a great article and it surprised me with the many things I hadn't known about Tukey. We mentioned Box as a major figure in our lifetime. And I would put Good in that category, though in a somewhat different sense, but certainly Tukey belongs to that category.

SS: Tukey was a genius and in unusual ways. I'll tell you a Tukey story. Late in his life, George Box was turning 80, and we were going to put on a party for him here in Chicago and George Tiao was a part of this operation. We thought, let's invite John Tukey because they were together in Princeton. I called Tukey at home early one evening, and he immediately picked it up and said hello. I remarked that I hadn't heard any ringing before he picked up, saying he must have guessed that I was going to call. He said no, I would have been hearing the ring-back, not the ring, and then he went into a 5–10-minute discussion about exactly how this was working, citing from memory all critical impedances from Bell Telephone technical manuals. He was carrying that around in his head!

RK: Having witnessed him on multiple occasions at Princeton, I would say he was also very odd.

SS: Tukey certainly was an odd person. Fred Mosteller once told me, somewhat in confidence, the way to talk to Tukey to learn something. If you want to ask him a question, you should know you will not get a straight answer. You can learn an awful lot from him about everything, but what you have to do is, you have to talk around the subject until you get to the question, and then he'll tell you the answer and it'll be brilliant. In a famous picture of him with the blackboard behind him, if you look at the blackboard, there's a chalk solid line on the board that seems to come into his head from the left and then goes out the right as a dotted line. That must have been planned by John. He was once at a party at our house, and a neighbor who had never met him, the wife of a neighbor, was there and she looked at Tukey who was actually wearing a tie. She looked at his tie and said, "Oh I get it, two keys—Tukey" as there were two keys crossing on the tie. He commended her observation. He was mystical in some ways, but he was absolutely brilliant. When he gave a preplanned talk, to me it was not his best work. But when he was a discussant, and had just heard something for the first time, he was absolutely brilliant. So, if you could get him to do that off the cuff, you were in for an education.

RK: Tukey stymied me once when I gave a lecture after I arrived at Princeton as a post-doc. He used the word slippage, which you've probably come across, but he was talking about translation families, and I had no idea what he was talking about, and he kept saying it and being brand new there I was too embarrassed to stop him.

SS: He did that to tease people. He was provocative, in that sense, he was hoping you would say, "What's that?" But he liked to name things.

RK: He famously came up with the names "bit" and "software," though many of his names didn't stick.

SS: In his Green Book with Mosteller (Mosteller and Tukey, 1977), the logit transform is given three different names in different parts in the book.

Measurement of Uncertainty Before 1900

SB: Steve, I think your book *The History of Statistics: The Measurement of Uncertainty Before 1900* (Stigler, 1986a) is ground-breaking in many ways. I was reading Ian Hacking somewhere, and he says in the 1980s there was a group formed in Bielefeld, Germany whose members were given tasks and you were going to write on the history of probability and statistics. Is that accurate?

SS: I agreed to go and spend the year in Bielefeld and Ian Hacking was there and Donald Mackenzie from Scotland and a lot of other people. I had written a couple of chapters of my book, no more than that, and then I thought maybe I could finish it there. But as we got closer to the time to move there, I got cold feet. My wife and I had two children in the 70s. We decided to have some more, and she got pregnant, and somehow the idea of going to Germany and giving birth there was not a practical one and so I bailed and didn't go. But a lot of interesting things came out of the Bielefield group including a two-volume book called *The Probabilistic Revolution*, and there's another called *The Empire of Chance*, which has five authors. I think also Lorraine Daston's book *Classical Probability in The Enlightenment*, which is a very good intellectual history. Ian Hacking was going to write a book, which he ended up writing much later called *The Taming of Chance.* I wrote somewhere recently that chance has never been tamed: it's always waiting around the corner coming to get you. But I know Hacking reasonably well. I've enjoyed learning from him on a lot of occasions. He's a very interesting character and he wrote a wonderful review of my second book, *Statistics on the Table* (Stigler, 1999) in *Nature*. Initially, my 1986 *History of Statistics* was going to go up to 1925. It stopped at 1900 because I knew I would have to include early Fisher and I realized that was not going to happen in finite time.

Concepts That Stick and Stigler's Law of Eponymy

RK: One thing I wanted to ask you about is that important new concepts, concepts that stick, often get absorbed into the mainstream of our discipline only after some period in which a lot of detailed work is done, yet over time a lot of those details seem to fade away. I'm thinking about some old mathematical developments, but

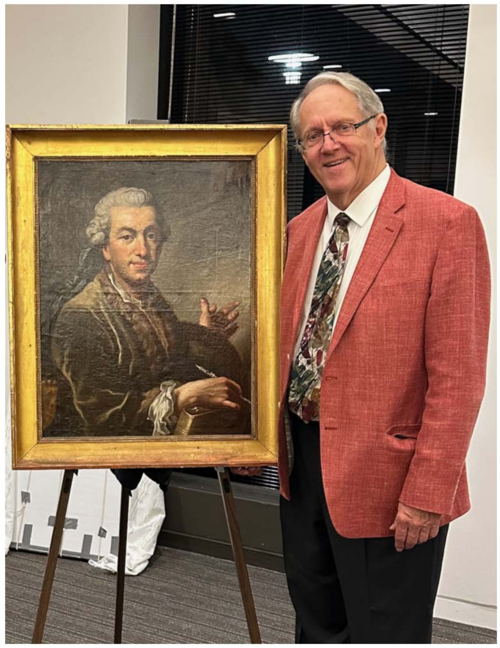

FIG. 7. *Steve with the portrait of Laplace in* 2022. *The portrait was painted in* 1784 *when Laplace was* 35.

I can point to recent things like the bootstrap. There was this whole battery of bootstrap-related research just after Efron published his initial paper, things like the second-order properties and so on, but no one talks about them anymore. I wonder if you can see that historically.

SS: Well, the first instinct when you're presented with a new idea is to say, new ideas are usually wrong. But when you get past that stage, you start saying let's poke at it a little bit. Let's see what its properties are. This is the opposite of what happens in most journal articles, where the author has carefully selected the examples to make their new procedure look better than brands C, D and F that had been in practice for a while. People like to prod at ideas and in the case of the bootstrap there were a number of interesting pokes. One was something that David Wallace had mentioned in one of his classes or lectures and then Nat Shanker followed up on it. The only thing I contributed to the article was part of the title: *Qualms About Bootstrap Confidence Intervals* (Schenker, 1985). It involved the fact that in terms of certain skewed distributions, the bootstrap does exactly the wrong thing; without correction, it gets the interval backwards. Then Efron introduced the bias correction, and that was a kind of progress. There were other things that people began to come up with, too. Peter McCullagh discovered that in certain structured designed data sets, it was impossible to get the bootstrap to give you the answer that you wanted. I remember that this had a great effect on a number of people. Art Owen at Stanford was calling attention to this.

RK: I was editor of *Statistical Science* when we ran a bootstrap article by Alistair Young with discussion

(Young, 1994), which I know Mark Schervish contributed to. Mark's comment was just as you say, poking around. I know he had this counterexample, where bootstrap failed, but Efron said something along the lines of in real life, problems are not so much pathological as they are clunky, and I thought it was very apt. The other aspect of this is that it may take a long time for people to even understand what the core idea is.

SS: I agree with that. People who in 1600 were taking means, were not thinking deep thoughts, they were not articulating the difficulties that they were overcoming. It's true that you learn a lot more, as you go forward, and some things become less important.

RK: So, you know I asked about this once before and you referred me to Merton on this, but I couldn't find anything by Merton that was relevant by searching.

SS: Well, he has a lot of stuff on multiple discoveries, and maybe that's what I had in mind. Bob was a wonderful sociologist of science who was at Columbia University. He was the father of the modern sociology of science, and among other things. He was the one who saw the sociological relevance in the fact that big discoveries often have multiple discoverers and become priority disputes. This points in part to the value of intellectual capital of discovery in science, and the importance that people put on priority. But it's more than that. The earlier discoveries may not stick, which reminds me of a comment that I've heard attributed to Larry Shepp. When somebody came up to him and said, "You know your such-and-such result, I have found that someone discovered it before you," Shepp replied, "Oh yeah? When I discovered it, it stayed discovered."

SB: That's a good one.

SS: A related thing I once wrote about (Stigler, 1992) has to do with a doctor named Archibald Pitcairn in the 1680s. He wrote about the problem of inventors, and what he had in mind was the discovery of the circulation of blood by William Harvey in 1628. Harvey's early book has an experiment where he's tallying how much blood could go through a dog's system. He was saying it's being pumped out so fast it can't be diffused, it has to be circulating. That's a mathematical argument, and it's done with great rigor, with experimental animals that are hard to deal with. When that came out, and it was so convincing, there were two reactions to it. There were people who said it couldn't be true. And the others who said we knew that all along: the other great reaction to a new discovery. With Harvey, critics were pointing back to Hippocrates. They went back and cherry-picked comments out of Hippocrates and said Hippocrates' work is absolutely consistent with knowing about the circulation of blood. Then Pitcairn said, in order for priority to be accorded to a particular inventor, it must be true that they said the discovery was true more often than they said it was false! And, in fact, you can find lots of places in Hippocrates, which contradict circulation of blood. There is a quote from Alfred North Whitehead: "Everything in science has been said before by someone who did not discover it. To come very close to saying something and fully realizing its consequences are very different things" (Whitehead, 1917, p. 127).

SB: This is a kind of flip side to Stigler's law of eponymy. Steve, could you tell us how you came to that?

SS: I'll tell you how it began. In 1980, I had been in correspondence for quite a while with Merton. He was known for his work on the importance and function of priority in science. He had a festschrift prepared for him a couple of years before that was primarily by sociologists and, unlike for most people, the sociologists were going to do a second festschrift for him, and I got invited to write for the second festschrift (Stigler, 1980). I think at the moment that I wrote it, I had never met him. I'd been corresponding with him and in those days you corresponded by letter. And these letters were daunting because I would write him something and I would get a two- or three-page, single-spaced letter back with all sorts of references, all over the place, wonderful stuff but that meant that when I went to answer it, I had to work really hard. We had a huge correspondence. It was absolutely wonderful. So, I started thinking about what I could do, and I got this idea: Stigler's law of eponymy. It is related to the question of multiple discoveries, but it's not exactly the same as it says that no scientific discovery is named after its first discoverer.

SB: So, you wrote this up as a scholarly paper?

SS: I wrote this as a serious paper with data on the sociology of science and among other things, did a study of all the ways the normal distribution has been referred to as Gaussian distribution, and various other things, but never as the de Moivre distribution (Stigler, 1980). The question is why was this the case? I mean, everybody knows that there are lots of examples of this, but if there are enough examples you've got a phenomenon. The answer, I postulated, was that all of these names are honorifics and are given out honorably. But what is also true is that they have to be given at some distance. If you name something after yourself, it doesn't catch on. And, if you name it after your best friend or your office mate or something like that it's unlikely to catch on. But if somebody in one country names something after someone in another, it's taken more seriously and it's more likely to catch on, and if distance is playing a role in this, the further you are from the fact, the less accurate your assessment of it may be. And so, it's sort of built into the system. I did emphasize that all of these assignments are made to honor people, and I had some good examples like Laplace transform. The Laplace transform can be found in Lagrange. The Fourier transform can be found before Fourier in Laplace. When you look at the different appearances of these things, you find that they aren't quite the same, or maybe they're more developed. But in Laplace's case, he was really doing inversion of Fourier transforms, taking full advantage of the fact that you could expand the complex exponential in sine and cosine terms and then when you multiply it out and integrate it, you can actually see the integration wipe out the cosine terms or sine terms because it's so symmetric. That gave me an understanding that I'd never achieved before. I could prove the theorem in graduate school, but I could never understand why it worked and once I saw how Laplace was doing it and how this worked, I understood. You could argue a little bit about the inaccuracy of the claim of the law of eponymy. I called it Stigler's law of eponymy to call particular attention to it. There's a Wikipedia page on this, but I have never touched that page.

RK: So, Steve, in a certain sense you're a counterexample to your own law!

SS: Well, that's a question, but I give a lot of credit to Robert Merton. Some have said that Stigler's law is also true about Stigler's law, because Merton discovered it!

RK: So, Merton had already written about priority. Can you summarize either what his perspective was or what are the important things to take away from that?

SS: He was talking in particular about the reward system in science. What is it that drives people to work in science? Why do they try to do things? If you ask mathematicians and many scientists, they say they do it because of the beauty of science, but that's not all. They also want to publish, and they also want to get credit for their work. And what is true is that when you have a priority dispute people don't fight unless they're fighting over something and what you're fighting over with priority is ego, associated with coming up with some important idea. There were other people who would call attention to priority fights and multiple discoveries earlier than Merton did, and there are several of them and they're all listed in his paper. He knew the literature immensely well, going back centuries.

SB: And sometimes theorems turn into verbs. We have things like Rao–Blackwellization!

SS: There's a study of that phenomenon by Herb Clark, a psycholinguist at Stanford University. But absolutely so yes. Other examples are Chebyshev's inequality. And then, lemmas like Neyman and Pearson's lemma. Well, these are really important things but, these are real, serious honorifics and they stand the test of time.

RK: It's interesting because some of them are, I think, genuine honorifics and some of them actually aren't so much. It's the convenient way to name something, is how it feels to me.

SS: You remember them better than you do theorem such and such. Think about central limit theorem. That's a lousy name, because it's taken as a translation from the German, which should have been translated as the fundamental limit theorem!

BOOK COLLECTOR

SB: The cliche image is that a lot of the resources required to write a history of statistics book are still in print format, and certainly when you started out there was no alternative. I can picture you in some esoteric bibliotheque in Paris, looking for books with a lot of dust on them, trying to figure out what Laplace wrote to Lagrange! Did that lead you to your passion for collecting such books?

SS: Well, when I started taking an interest in the history of statistics in the early 70s, by getting interested in odd things that at first didn't make sense, yes, I started collecting books. I discovered that I could buy books for 10 or 20 dollars that were 200 years old with leather bindings, written by really good people. So, I started collecting what I now have, and you're looking at it just a corner of what I have (Steve shows a portion of his massive library via Zoom.) This is my study which at the very bottom right, above my head, is an entire set of the 14 editions of Fisher's *Statistical Methods for Research Workers*. At the very top, I think, is Borel's work and some other stuff, and this is one wall. All four walls are like that. I have built a large library, and I did this by hanging out, as you said, in bookstores in Europe, and finding things. I've also hung out in the libraries, and I visited archives. In the 1970s, I went in and got some correspondence of de Moivre in the Bernoulli archives in Basel. I visited the University College library in London and read papers of Pearson and Galton. I visited Adelaide and read the papers of Fisher. But really what got me going was I had a sort of a scholarly plan: The question I'd started with was how do you get from least squares to regression and why was it taking nearly a century, and what was the intellectual development that led to this and how did this happen? One of the first books I purchased was essentially the first edition Legendre on least squares. It was $300, which was incredibly cheap considering its historical importance. It is worth a lot more now.

RK: Steve, I have to ask you. Have you considered picking some chunk of what you have sitting in your house and getting someone to convert stuff into PDFs?

SS: A lot of it already exists in PDF. When Google digitized the major libraries of this country, that meant that everything before 1900, or before 1923 maybe, is sort of open source now. Google Books is not terribly well curated but it has an immense amount of material.

RK: What about the correspondence though?

SS: Correspondence is a different thing. The entire French libraries, by the way, are online, including some correspondence. It's called Gallica by Bibliotheque Nationale. But the correspondence is a huge thing. I don't own a lot of correspondence. I have couple of letters of Quetelet, a couple of letters of Galton, a couple of theses of Pearson and other things. But it doesn't amount to a hill of beans, compared with what the university libraries have. Putting all of this online with good finding tools is very expensive and so they aren't going to do it either.

A Lottery in France

SB: Before we started the interview, you mentioned that you were almost done with a project on the history of lottery in France, and now it's finished (Stigler, 2022). Why a lottery in France, and not in, say, Japan?

SS: I started collecting books in the 1970s, and I have a huge collection, but in the 1990s, I spotted something in a French bookseller's list that looked interesting. It was an almanac about a lottery in France, and sight unseen, I bought it for a couple of hundreds of dollars. It was falling apart, but I noticed what's in it was a lot of data. The lottery would draw five numbers from among 1 to 90 and then you win or not according to what you were betting on, much like a modern lotto. I'll skip the details, but all five winning numbers were given for each of 6606 drawings between 1758 and 1834. It happens that the lottery went out of existence in January of 1836 so it's almost the entire list of all the drawings. I was looking at that and saying wow! Data! I wanted to find out whether this was being drawn fairly, and so I started to type the data into a file, and well I'll tell you it's in an old font and the papers are dirty; this was not easy. Optical character recognition was then not the way to go, and after a while I got tired, you know. So, I put it aside and a couple years later, a young student by the name of Teresa Ging came into my office and says, do you have any ideas for an honors paper for me? A light goes on. So, she typed all of this in and wrote a nice paper on the French lottery. Then I ran all sorts of tests on those numbers. I learned how to make the right correction for the fact that you're drawing five numbers without replacement. You can now find that in a problem in McCullagh and Nelder (1989). So anyway, I tried a lot of tests and it passed wonderfully, and I got interested in the subject and wrote a paper that was published in 2003 in a French journal (Stigler, 2003). But I kept gathering more information, including three huge scrapbooks of old tickets from these lotteries and others. I now have a whole shelf of books about lotteries, going back to 1619. And I learned an immense amount and I wrote the book *Casanova's Lottery*. It has about 50 or 60 pictures of different old documents and things like that.

SB: Thank you, Steve. This was wonderful!

REFERENCES

ANSCOMBE, F. R. (2003). Quiet contributor: The civic career and times of John W. Tukey. *Statist. Sci.* **18** 287–310. MR2056571 https://doi.org/10.1214/ss/1076102417

BANKS, D. L. (1996). A conversation with I. J. Good. *Statist. Sci.* **11** 1–19. MR1437124 https://doi.org/10.1214/ss/1032209661

EATON, M. L. (2007). William H. Kruskal and the development of coordinate-free methods. *Statist. Sci.* **22** 264–265. MR2408962 https://doi.org/10.1214/088342306000000367

FIENBERG, S. E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Anal.* **1** 1–40. MR2227361 https://doi.org/10.1214/06-BA101

HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon–Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.* **20** 225–241. MR0030730 https://doi.org/10.1214/aoms/1177730032

KASS, R. E. (1989). The geometry of asymptotic inference. *Statist. Sci.* **4** 188–219.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models. Monographs on Statistics and Applied Probability*. CRC Press, London. 2nd ed. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

MOSTELLER, F. and TUKEY, J. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, Reading, MA.

SCHENKER, N. (1985). Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **80** 360–361. MR0792734

STIGLER, S. M. (1969). Linear functions of order statistics. *Ann. Math. Stat.* **40** 770–788. MR0264822 https://doi.org/10.1214/aoms/1177697587

STIGLER, S. M. (1973a). Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *J. Amer. Statist. Assoc.* **68** 872–879. MR0362575

STIGLER, S. M. (1973b). Studies in the history of probability and statistics. XXXII. Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika* **60** 439–445. MR0326872 https://doi.org/10.2307/2334992

STIGLER, S. M. (1976). Contribution to discussion of "On Re-reading R. A. Fisher," by L. J. Savage. *Ann. Statist.* **4** 498–500.

STIGLER, S. M. (1980). Stigler's law of eponymy. In *Science and Social Structure: A Festschrift for Robert K. Merton. Transactions N.Y. Acad. Sci. Series* **2** 147–157.

STIGLER, S. M. (1983). Who discovered Bayes's theorem? *Amer. Statist.* **37** 290–296.

STIGLER, S. M. (1986a). *The History of Statistics: The Measurement of Uncertainty Before* 1900. The Belknap Press of Harvard Univ. Press, Cambridge, MA. MR0852410

STIGLER, S. M. (1986b). Laplace's 1774 memoir on inverse probability. *Statist. Sci.* **1** 359–378. MR0858515

STIGLER, S. M. (1992). Apollo mathematicus: A story of resistance to quantification in the seventeenth century. *Proc. Amer. Philos. Soc.* **136** 93–126.

STIGLER, S. M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard Univ. Press, Cambridge, MA. MR1712969

STIGLER, S. M. (2003). Casanova, 'Bonaparte', and the loterie de France. *J. Soc. Fr. Stat.* **144** 5–34.

STIGLER, S. M. (2007). The epic story of maximum likelihood. *Statist. Sci.* **22** 598–620. MR2410255 https://doi.org/10.1214/07-STS249

STIGLER, S. M. (2008). Karl Pearson's theoretical errors and the advances they inspired. *Statist. Sci.* **23** 261–271. MR2446501 https://doi.org/10.1214/08-STS256

STIGLER, S. M. (2016). *The Seven Pillars of Statistical Wisdom*. Harvard Univ. Press, Cambridge, MA. MR3585675 https://doi.org/10.4159/9780674970199

STIGLER, S. M. (2018). Richard Price, the first Bayesian. *Statist*. *Sci*. **33** 117–125. MR3757508 https://doi.org/10.1214/17-STS635

STIGLER, S. M. (2022). *Casanova's Lottery*: *The History of a Revolutionary Game of Chance*. Univ. Chicago Press, Chicago, IL.

VON NEUMANN, J. (1955). *The Mathematical Foundations of Quantum Mechanics*. Princeton Univ. Press, Princeton, NJ. Translated from the 1932 German edition.

WHITEHEAD, A. N. (1917). *The Organization of Thought*. Williams and Norgate, London.

WICHURA, M. J. (2006). *The Coordinate-Free Approach to Linear Models*. *Cambridge Series in Statistical and Probabilistic Mathematics* **19**. Cambridge Univ. Press, Cambridge. MR2283455 https://doi.org/10.1017/CBO9780511546822

YOUNG, G. A. (1994). Bootstrap: More than a stab in the dark? *Statist*. *Sci*. **9** 382–415. MR1325434