Statistical Science
2001, Vol. 16, No. 2, 184–198

A Conversation with Leo Breiman

Richard Olshen

Abstract. Leo Breiman was born in New York City on January 27, 1928. His parents and he migrated five years later to San Francisco where he began school. During Leo's junior high school years, his family moved again, to Los Angeles. Leo graduated from Roosevelt High School in 1945 and entered the California Institute of Technology, from which he graduated four years later with a major in physics. He earned his Master's Degree in mathematics at Columbia University in 1950 and his Ph.D. in Mathematics at the University of California, Berkeley in 1954.

Leo has broad ranging scientific and mathematical interests, including information theory and the theory of gambling. He has been involved in applications coming from studies of automobile traffic, air quality and toxic substance recognition. He is the author of a celebrated graduate text on probability theory, is one of four authors of *Classification and Regression Trees* and its associated CART^{R} software and has also written two other books. With Jerome Friedman, Leo developed the ACE (alternating conditional expectations) algorithm by which nonlinear relationships between the dependent variable and predictor variables in regression are described. He is the originator of "bagging" and "arcing," both computer-intensive approaches to classification that are of much current interest.

Leo's professional positions have included being on the faculty of the Department of Mathematics at UCLA, an independent consultant for 13 years and Professor of Statistics and founding Director of the Statistical Computing Facility at the University of California, Berkeley. In addition, he has had visiting positions at Stanford and at Yale. For his many contributions, Leo has been honored by Fellowship in the Institute of Mathematical Statistics and in the American Statistical Association. He is an elected member of the American Academy of Arts and Sciences and received the Berkeley Citation from the University of California.

The interests and accomplishments of Leo Breiman extend outside the areas of professional statistician and probabilist. He was a waiter in the Catskills, a dishwasher in the Merchant Marine, a trekker into the heart of rainforest Africa, an active father to many children from a small agrarian Mexican village, a member and President of the Santa Monica School Board, the architect of his stunning home and an accomplished sculptor. Leo and his wife, Mary Lou, reside in Berkeley. He is the father of two daughters, Rebecca and Jessica.

Richard Olshen is Professor of Biostatics, Chief, Division of Biostatistics, Associate Chair, Department of Health Research and Policy, Professor (by courtesy) of Electrical Engineering and of Statistics, Stanford University, Division of Biostatistics, Department of Health Research and Policy, Stanford, California 94305-5405 (e-mail: olshen@stat.stanford.edu).

This conversation was held at Leo and Mary Lou Breiman's home on February 19, 1999.

Olshen: Leo, I want to say that it's a privilege to be here. I was an admirer of your work and taught from one of your books before we collaborated, in fact, before I knew you. I understand that you grew up in Boyle Heights in Los Angeles. What was Boyle Heights like?

BOYLE HEIGHTS

Breiman: Well, Boyle Heights was, at that point, a poor Jewish ghetto, working-class Jewish ghetto, which was bordered on one side by the Mexican ghetto. It was basically a first-generation immigrant ghetto. My parents were immigrants; everyone else's parents were immigrants.

And we were all going to Roosevelt High School, which was busy turning offspring of immigrants into doctors and lawyers and scientists. It was a wonderful place to be. You could go into a delicatessen, get a pickle out of a pickle barrel, walk down the street and see salamis hanging and drying. It was something just a little upscale from the lower East Side of New York.

Olshen: And better weather perhaps.

Breiman: Right.

Olshen: When did your interests in mathematics and science begin to emerge?

Breiman: I got turned on by geometry in high school because geometry was the first subject I took that really impressed me. You know, when you hit geometry after having done algebra or added numbers or something, it's so completely different and fascinating that a bunch of us couldn't stop working on it.

Olshen: What more can you tell me about Roosevelt High School, the neighborhood, what got you interested in Caltech and actually got you to Caltech?

Breiman: Well, Roosevelt High School, although it was in one of the poorest socioeconomic neighborhoods of Los Angeles at that time, was also one of the premier academic high schools, probably for the same reason that, say, the Bronx High School of Science was.

It was filled with highly motivated sons and daughters of immigrants, all of whom knew they were somehow going to wind up in college. And what stimulated me to go to Caltech was, I think, my mother's hearing that it was the best scientific school in the West and deciding that her son should go to Caltech.

Olshen: Did you have sibs?

Breiman: No.

Olshen: So it really mattered what happened to the one, the one son?

Breiman: Right.

CALTECH

Olshen: Tell me a little bit about Caltech as it was when you were there.

Breiman: Sure. The entering class at Caltech then was about 250 out of the many thousands who applied. And to get in, you had to take 16 hours of exams. They had no faith in high school grades at all. There was a four-hour exam in physics, four hours in math, four hours in chemistry, and four hours in English. And we sweated it all out.

My first year in Caltech I got a scholarship, and I did extremely well. And then my second and third and fourth years, I got more and more fed up because I got so tired of hearing nothing but science and engineering that it began to turn me off. The place was like a scientific monastery.

So as a result, my grades started going down and going down and going down. And in my last year, I got four D's in my major subject, which was physics.

Olshen: And this is what commended you to the University of California at Berkeley Graduate School?

GRADUATE SCHOOL AT COLUMBIA

Breiman: No. What happened was my grades in physics were terrible, but I kept being interested in math. I did fairly well in math. I applied all over, and the only place that accepted me was Columbia University in New York City. My parents were quite poor and couldn't afford the tuition at Columbia.

But from what I had saved from working, I was able to go to Columbia; and I got my Master's in mathematics in a year. My grades in math were pretty good, but actually I had gone to Columbia determined to become a philosophy major.

When I got there, I talked to the head of the philosophy department, Erwin Edwin. At the time he was quite well known. And he said, patting me paternalistically on the knee, "Look, two of my finest Ph.D.'s right now can't get jobs. Why don't you stay in math and take a few philosophy courses and see how it works out."

So I took a few philosophy courses—for instance, a course in aesthetics, a course in Greek philosophy and so on. I decided this is not at all what I had in mind. So there I was back in math.

Olshen: Tell us exactly when this was.

Breiman: This was in 1950. I got my Master's degree after a year at Columbia, and I applied to the Berkeley Math Department. They said, "Come." To support myself, I worked one summer as a waiter in the Catskills and for another I was in the Merchant Marine.

Olshen: The Merchant Marine sounds good.

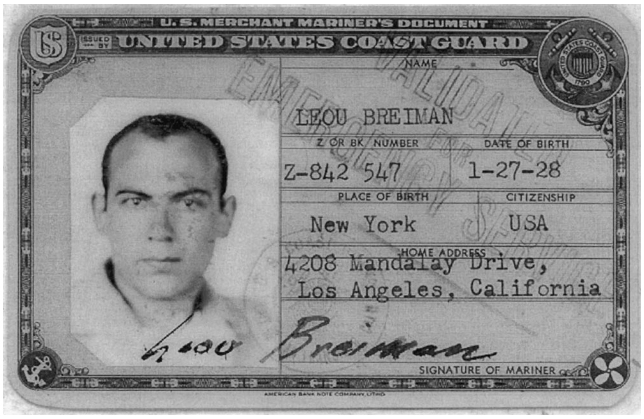

FIG. 1. *Leo's Merchant Marine ID*, 1950.

GRADUATE SCHOOL AT BERKELEY

Breiman: Merchant Marine. Right. And I made a large amount of money for a kid in his early 20's. I was the dishwasher on board ship. And I got off the ship just in time to proctor an exam that all Berkeley Math T.A.'s were supposed to be at. I took a taxi from the dock and I rushed over to Dwinelle where this big exam with hundreds of freshman was being given. So—

Olshen: So you interleaved your sailing career with your mathematical career?

Breiman: Right. Right.

Olshen: Excellent. Tell me a little bit about the influence of David Blackwell and others with whom you studied at Berkeley. I'm thinking in particular with regard to two subjects: one of them is the Shannon–Breiman–MacMillan theorem (Breiman, 1957) and the other is your work on gambling (Breiman, 1960).

Breiman: Well, I was in the Math Department at Berkeley at that time. There was no separate statistics department. There were a few people at that time who Neyman managed to bring in, such as Michel Loève. And I was taking mathematics courses, and then I took the probability course from Michel Loève, and I loved it. I loved probability theory.

And that was quite a class, a probability course of 10 people or so. Manny Parzen was in that class, Howard Tucker, too, and one or two other people who went on to become well known in statistics. Then I wrote a thesis. But Michel Loève was a perfectionist, kept me working on the thesis, and working on the thesis.

Olshen: I can't help but be reminded that Michel Loéve was my adviser when I was a freshman at Berkeley. And the first words he spoke to me when I walked in his office were, "You better make all A's or I'm going to throw you out of this window." That was his speech to a frightened freshman!

Breiman: That sounds like Michel. So I had written this thesis, and I thought it was pretty good. Michel kept saying, "Well, I don't know." Then, Harold Cramér came to Berkeley for a semester and organized a seminar. And I gave my thesis results in his seminar.

Afterward he said, "I'll have to talk to Michel and tell him I think this is pretty good." But the other important thing that happened was I got a notice from my draft board saying, "This is it. You've used up all of your deferments. We expect to see you here in 30 days."

And so I handed Michel a copy of the latest version of my thesis, with the draft notice on top. And he took one long look, and he said, "Well, we have to get you out of here pretty fast." So he organized the final defense of the thesis in about two weeks and put a stamp of approval on it. I was thrilled. Then I served almost two years in the Army.

Olshen: Wait. Let's back up. Was your thesis related to the Shannon–Breiman–MacMillan theorem or to gambling?

Breiman: No. What happened was that in the Army then, you could get out two months early if you had a job. I wrote to Professor Neyman and he gave me a job as acting assistant professor. I came back and at that time I got to know David Blackwell pretty well.

We used to sit in seminars, and he would send me these little notes saying, "Can you prove this and this and this?" And one of the notes he sent me was, "Can you prove that this thing converges almost surely to the following?" And I thought about it for the next couple of days. And then I thought, "Yes. I can do that. I can see how to do that."

That was how the Shannon–Breiman–McMillan Theorem (Breiman, 1957) came about. Dave Blackwell was, for me, an incredibly inspirational person to work with. To see the way his mind worked was terrific.

Olshen: Did he get you into this whole business of getting rich quickly on favorable games (Breiman, 1960)?

Breiman: No. How that happened was pretty much on my own. I was thinking about the Kelly criterion.

Olshen: Tell me about that.

Breiman: The Kelly criterion roughly says that if you do a certain kind of betting in a simple favorable game, you can make your winnings go to infinity. It was not almost surely, but that the expected values would go to infinity. I thought about Kelly's results for a while. Then I realized that they might lead to an optimum strategy for playing favorable games.

And that's what set me off. It was an accidental thing. David was not involved. I don't think that at that point he was interested in gambling at all.

SELLING ICE IN MEXICO

Olshen: It's become clear already that you've always had wide-ranging interests. One that's fascinated me, although I can't pretend to understand it yet, is the scheme for selling ice in Mexico. Can you tell us about that?

Breiman: Right. Well, that's actually due to a couple of friends of mine who went down to Puerto Vallarta in, I think, 1964 or '5, the year after the movie *Night of the Iguana* was made. This was when Puerto Vallarta was just this little town and you had to boil all of your water.

The other thing you had to be careful about was if they put ice in your drink, don't drink it. All the local ice was made out of local water and was swarming with bacteria. So my friends got this idea, which also interested me, of getting a purified ice machine down to Mexico and setting up a factory for making purified ice.

I thought that was great and I said, "I'll invest." What happened was, sure enough, they bought this old ramshackled building called El Bucanero, the bucaneer. They set up an ice plant in it and it was a great success. And the ice was known throughout Puerto Vallarta because they sucked the impurities out of the middle so that the ice looked like little donuts. And wherever you went, if you saw a drink with little ice donuts in it, you knew you were okay. After a while, it was so successful that the mayor's son decided to go into the purified ice business. And sure enough, all our good customers somehow started patronizing the mayor's son's ice factory. And business dried out. But it was great while it lasted.

Olshen: Sounds good.

Breiman: Uh-huh.

THE UCLA YEARS

Olshen: Now I'm going to change the subject somewhat radically in order to segue into your career at UCLA. And, in particular, I want to ask you something that I've always wondered very much about. And that is, how did you come to write that wonderful book on probability (Breiman, 1968)?

Breiman: When I went to UCLA I went because I only had an acting appointment at Berkeley. And those days the rule was pretty strict that if you were a Berkeley Ph.D, you definitely couldn't get your first job at Berkeley. The next best thing looked to me like UCLA.

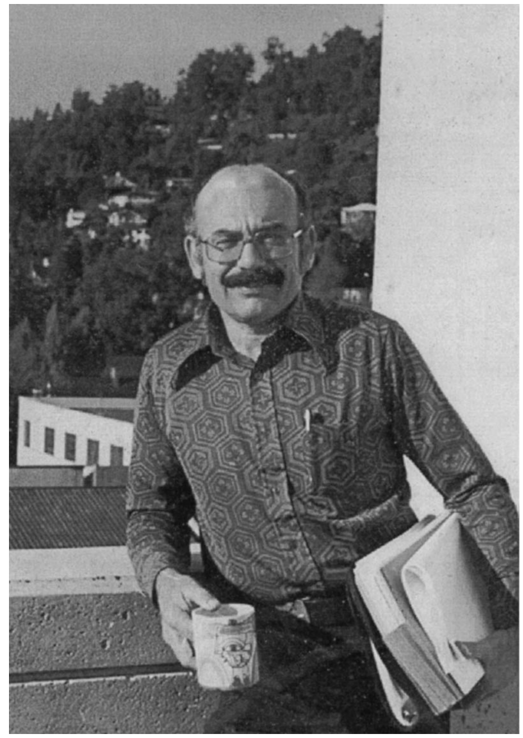

FIG. 2. *Leo as a youngish probabilist at UCLA.*

So I walked in off the street and went to the Math Department and said, "I would like to get a job." They got some letters about me, and I got a job. I was there for about seven years, and I actually got tenure fairly quickly. I was their only probabilist. So I was teaching the graduate course in probability theory.

I taught it about three or four years. And through this whole teaching, I kept trying to figure out really what was going on in probability theory—going on in the sense of working on proofs until I really understood or figured I could understand what made the thing work.

Well, after about seven years at UCLA I said to myself, "Look. Leo, you're not cut out to be an abstract mathematician. It's been fun, but this isn't going to work for you." So I resigned.

Olshen: Now, the book had been written by then?

Breiman: No.

Olshen: Oh, it hadn't?

Breiman: No.

Olshen: Okay.

Breiman: So I took out the money that I had put into the retirement system, and I spent six months doing nothing but writing the book. This was quite a period of time because it took strong discipline to write it. But again it was also a lot of fun, and it was my first born.

Thus, I was tickled pink when SIAM approached me just three or four years ago and said, “We want to republish it in our classic series” (Breiman, 1968).

Olshen: I’ve heard that you went to Africa during your time at UCLA.

Breiman: Yes. This is the way it happened. I wanted a sabbatical that would be really different. So I went to UNESCO and I said, what have you got for me? I don’t want to teach in a university. And UNESCO finally said, “We’ve got a place for you as an educational statistician in Liberia.”

So I went to Liberia as an educational statistician and what did they need? They needed to find out how many kids they had in the schools. Why was this a problem? They only had 50 miles of paved road in the whole country. The schools were in the rain forest jungles that were virtually inaccessible. So we formed teams. We formed about 20 teams to ride or paddle or walk all over the country, that would go into little villages, call out the school children, and just count them.

And so I lived there, I went on a number of these treks into the backwoods, into tunnels in the rain forest jungles that came to villages, and what was amazing when I got there was that little children would walk up to me and rub my skin to see if the white would come off. That was an altogether fascinating experience. I just loved the life in Africa and I was sorry to leave.

Olshen: Tell me now about your interactions with Bill Meisel and about how you came to Technology Services Corporation and your position there.

AN INDEPENDENT CONSULTANT

Breiman: After I resigned from UCLA and I used up my retirement money, it became pretty clear I needed to do something to support myself. So I asked around. And at the time, SDC—Systems Development Corporation—which was an offshoot of Rand, had a lot of Federal money for research into freeway traffic.

I had written an article about how traffic approached a Poisson process (Breiman, 1963) and I knew someone who worked in traffic research. And he said, “Sure. You know, why don’t you consult for SDC.” So I consulted for them for about a year on traffic. It was a lot of fun trying to figure out what freeway traffic looked like and what its statistical characteristics were.

Then that money dried up. So I was asking around again, and Pete Payne, who was another consultant, said, “Well, I think TSC” —this firm, Technology Services Corporation— “is looking for consultants.” So I went over there and Bill Meisel interviewed me. We talked for a while. And he said, “Okay. You’re on. When can you start?” Bill was a very interesting guy because he had been on the faculty of USC Engineering.

Olshen: Did you know him at Caltech?

Breiman: No. Not at all. And he had written a book on pattern recognition. TSC was essentially a radar house, a defense house, not a very big outfit, maybe 50 or 60 technical people. And Bill was running this sort of odds and ends division that he decided to focus upon its becoming an Environmental Division.

So I started consulting for the Environmental Division. Between the two of us, we started writing proposals to the EPA, to the California State Air Pollution Board and so on. And we began winning a lot of these proposals. I started doing a lot of work that was EPA sponsored on air pollution, water pollution and a variety of things of that sort.

A lot of the problems that we dealt with in those days involved large amounts of data. For instance, in one big project we had seven years of hourly and daily data on over 450 variables relevant to air pollution. We were trying to predict next day ozone levels in the Los Angeles Basin.

Bill was very instrumental in getting me to think about problems such as classification and regression. For EPA, and many of the other places around that were funding research, so many of the problems were prediction and classification that this was a hot area. And this was an area that Bill knew something about. And he pushed me into this area.

Olshen: Smells like the beginnings of CART^{R} to me. (Breiman et al., 1984).

Breiman: Well, that’s right. We were working on prediction problems like next day ozone in the Los Angeles basin, carbon monoxide levels on freeways, but also things such as could we recognize the sender of handset Morse code—this was something we were doing for the spook agencies—or could we recognize from sonar returns whether the other submarine was Russian or American? Or could we recognize whether that battleship we were getting radar returns from was Russian or American? So a lot of this stuff was fascinating classification stuff. At the same time, Bill went for almost every interesting statistical request for proposals he could see.

For instance, one thing we did was a study of delay in criminal courts in Colorado. We did some other work for the criminal court system. I designed some surveys and helped supervise surveys for the EPA.

Thus, I was doing a whole bunch of things that taught me about statistics—what you did with statistics and data while you were trying to solve problems. Does that answer your question?

SANTA MONICA, MEXICO, THE SCHOOL BOARD AND TEACHING KIDS MATHEMATICS

Olshen: That's just fine. Now, somehow all of this or a lot of it seems to have taken place in Santa Monica. I'm curious about your other involvements in Santa Monica, for example, with the Santa Monica School Board, how you came to be not only on it, but also its leader.

Breiman: Well, let me back up a little bit on that one. My friends and I got to know this little village about 30 miles north of Puerto Vallarta called Mezcales. At the time, I was taking care of two or three kids even though I was a bachelor. But they were kids of friends of mine who had gone to Saudi Arabia to work and decided that was not a good place for kids to be.

So they sort of turned them over to me. And one of the kids was a young boy who was getting picked on by two older girls. I wanted to find another boy to even the score, but I couldn't find another boy whose parents were willing to let go of him. Then somebody asked why not take up a couple of young Mexican boys from this village of Mezcales.

I went through an incredible hell with the bureaucracies—both Mexican and the United States—but I managed to get these two kids up to live in my house with me. At that time it was in Topanga Canyon. They went to the local grammar school along with my American kids, and it was a great success.

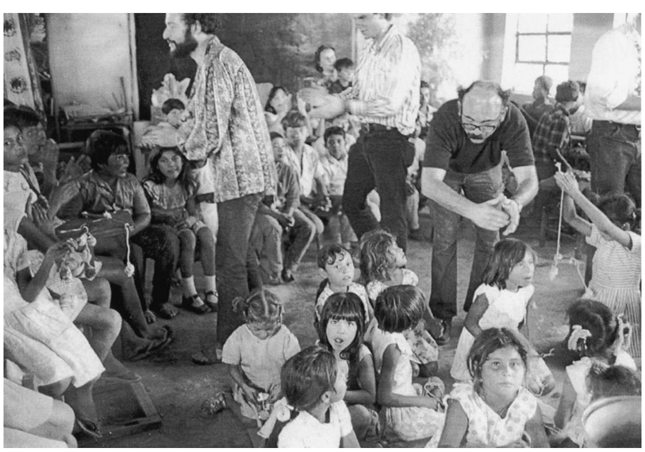

FIG. 3. *Leo with school children in Mezcales, Mexico in* 1971.

One reason for bringing them up is that the people in that village were dirt farmers, and the only way these kids would even get away from being dirt farmers was to learn English, because the only big industry around was the tourist industry. By the end of the year, those kids were chattering in English.

They were absolutely fluent. Okay. So the next summer I returned the first two, whose parents were very happy, and brought up three more. This went on for seven years until altogether I brought up something like 21 kids. By this time I had gotten to know the Santa Monica School District pretty well.

And by the way, the local teachers really loved my Mexican kids. And—

Olshen: Even to get from Topanga Canyon to Santa Monica schools is a trek.

Breiman: It is. And I was a bachelor at the time and consulting. So I had to organize things pretty carefully. But anyhow I finally moved down to Santa Monica because what happened is the Santa Monica system caught me. What I mean by that is that they discovered I was living in Topanga, which was in the Los Angeles School District. So the kids should have been going to Los Angeles schools, which were not very good. I had them going to Santa Monica schools under the address of a friend of mine, who lived in Santa Monica. They visited that house one day and discovered the truth. And they said to me—

Olshen: They being?

Breiman: They being the Santa Monica School District officials.

Olshen: Okay.

Breiman: That, "You either move to Santa Monica within 30 days or the kids are out." So I moved to Santa Monica. I got to know that school district really well.

When I was at UCLA, I had decided that the way kids were being taught mathematics was all wrong because they wound up thinking of mathematics as an awful boring subject that had nothing to do with the everyday world around them. So I decided I would teach two fifth grade classes mathematics. One I taught was at UCLA—their experimental school for emotionally disturbed kids—and then at another school in the Valley, which was just straight middle class and –

Olshen: Were you paid for this?

Breiman: No. No. But I had a wonderful time. And the kids had a wonderful time. We played all kinds of games. They didn't know they were doing mathematics, but they were learning things like how to play battleship games. They were learning

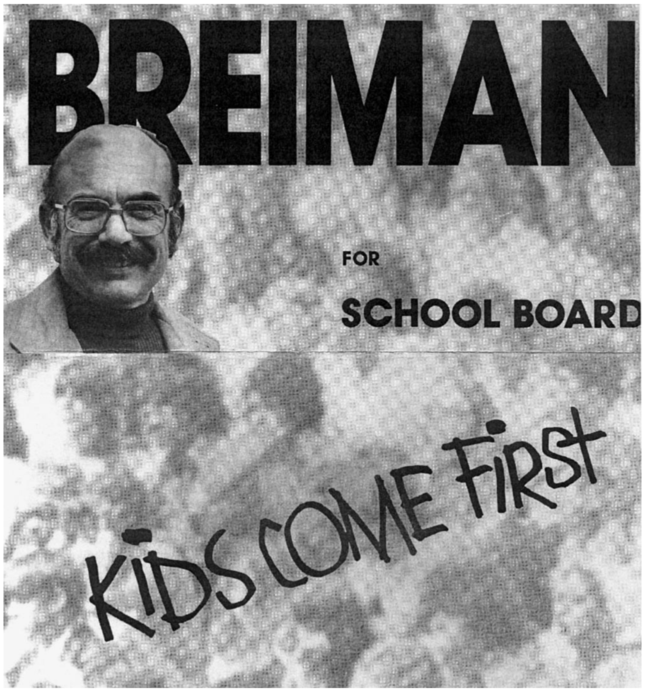

FIG. 4. *Campaign leaflet from the race for the Santa Monica School Board*, 1976.

Cartesian geometry, they were learning algebra, they were learning all sorts of sophisticated things under the guise of playing games.

And, you know, we would play a game like, "Okay. How far is it? How tall is that wall?" And the kids would guess and then I would sort of beat around, "How can we really find out how tall it is without climbing it?," and things like that.

So between these two things, that is, having all of my kids wander in and out of Santa Monica schools and being really interested in how mathematics should be taught in grammar schools, I decided that the Santa Monica School District needed me. All right. So I ran for the School Board.

And it's interesting—nobody in Santa Monica knew me. When I went to the teacher's union for an endorsement and I talked to them, they said, "Well, we love what you're saying, but we can't back you because you don't have a chance of winning." And the Democrats wouldn't back me because they said, "Well, you haven't paid your dues here."

But I worked hard and I put on a good media campaign and talked all over the place and walked door to door and stood in supermarkets and handed leaflets to people. And, to my surprise and everybody else's, I got elected.

Olshen: How did you become President?

Breiman: At that time, the Santa Monica School Board consisted mainly of local businessmen. This was in the days when Santa Monica was still a fairly conservative city. And the local businessmen respected the fact that I was really interested in education. And so they nominated me first as Vice President, and the Vice President automatically got to be President the next year. But I have to say those businessmen on the Board were really sincerely and genuinely interested in the quality of education in Santa Monica—much more so than some of the politicized and more radical people that followed them on the Board.

Olshen: With your writing and your teaching at two universities and your career in a private sector, you've had to teach lots of people. Do you find that there's a common philosophic theme or a common spirit to teaching, whether it's a fifth grader in Santa Monica or a brilliant graduate student in Berkeley?

Or is there some abiding theme that you have that permeates your teaching at all levels and also your writing? You talked about making mathematics fun and about relating it to the world around you—

Breiman: Right.

Olshen: —as examples.

Breiman: Let me think out loud about this. With the kids, I sort of got convinced that the crux of their failure to learn mathematics was the translation process, namely, that they were being taught multiplication or whatever in this little shut-off world and they couldn't translate it into their real world.

And that the thing that was symptomatic of this was their inability to work word problems. In word problems, what you have to do is take the English and utter those magic words "Let x equal." In other words, you have to translate the English text into mathematics.

My sense of it is that the people who turn out to be good at mathematics are those who can translate from the real world into a mathematical world and can translate from a spoken language into the language of mathematics. I think that permeates my teaching at all levels. We start with a certain problem, and I find myself drawing pictures of the problem before we ever get abstract. You're trying to make the problem concrete before you translate it into the abstract.

BACK TO BERKELEY AND THE STATISTICAL COMPUTING FACILITY

Olshen: Thank you. To move on, chronologically, how and why did you come to Berkeley?

Breiman: In my consulting, I had gotten various offers, to my surprise, to come back (as my friends called it, "come back in out of the cold") to a university. And I didn't particularly want to. I was having a wonderful time consulting. There I was solving real problems, working with data, things that I loved.

But then one day I got a phone call from Peter Bickel and he said, "Would you like to come up to Berkeley for a semester just to teach and give a seminar?" And I said "Sure." Did that, went back and that was a good time.

Breiman: Afterward, I went back down to LA and about a year later, I got another phone call—I think this time from Dave Brillinger—saying, "Okay. Would you like to come up and be a faculty member?"

Now at that time there was only one place in the world that could have convinced me to come back into academia and that was Berkeley. I had done my graduate work here, I still had a lot of friends here and I loved it here, so I said, "Okay."

Olshen: When you came here, was it with a view to founding the Statistical Computing Facility or did that just evolve? Tell me about it, the struggles for resources, your leadership and what you see of it in the future.

Breiman: Let me back up just one minute if I can, Richard.

Olshen: Please.

Breiman: Because, you see, when I left the university, in a way I knew I was burning my bridges in terms of ever coming back to the university again because for all the years I was consultant, let's say for 13 years, I think I had maybe one publication. I mean, you're not paid as a consultant to publish; you're paid to solve problems.

So I think it was rather extraordinary that Berkeley managed to get me hired back as a senior faculty member. And, you know, I've never known the full story of how they got it through. Nobody has ever seen fit to tell me. I think I'll have to ask at some point.

Olshen: There doesn't seem to have been a lot of bad news associated with the appointment. Maybe it's just best to let it go. But now tell me about the Statistical Computing Facility.

Breiman: What happened was this. When I got to Berkeley, all the computing in the department was essentially done downstairs in the Central Computing Facility.

Olshen: Downstairs in Evans Hall?

Breiman: In Evans Hall. They had a big multiprocessor computer main frame, and we had a little PDP-11 in the department, which was virtually worthless, had 32 kilobytes of main memory and so on. I remember my first visit down to the Central Computing Facility. I asked them, "What statistical packages do you have?"

And they said, "Well, I think we have BMDP. No. We don't have the whole thing. We only have parts of it."

And I said, "Well, do you know what parts you have?"

And they said, "Well, we'll try to find out."

I was supposed to teach this course in multivariate analysis. And it was clear that for the department to ever get into any kind of relationship with data, they had to have a decent computing facility. And the basement of Evans was not the answer. So I said, "Okay, the first thing I'm going to try to do is to get a decent computing facility here."

This was 1980, and the best minicomputer we could get was the VAX 750. It was sort of within our price range. It had a 16-bit system, a lot more memory and it cost about $75,000 at a discount price.

Olshen: Did you have the support of the University administration in all this or was this all out of the Department of Statistics?

Breiman: All out of the Department of Statistics. We had no money.

Olshen: Uh-huh.

Breiman: And all I had was a tin cup that I was rattling.

Olshen: And an ice machine in Mexico.

Breiman: Right. So I started calling around, trying to find out if we could get the Federal agencies to possibly fund the VAX. And I hit up Ed Wegman at ONR. And I said, "Ed, if we're going to do any decent amount of any applied research, we've got to get a computer."

He said, "How much?"

I said, "$75,000."

He said, "Okay. Send me a proposal."

And Ed sure enough came through. And that was our first VAX. We got that installed and hired a graduate student as our first system manager. And from then on it was easier—once you break the ice, it gets easier. After many donations and grants, we are now running 2,000 accounts per year with four full time high level staff members and substantial computing power. We also support computing in economics and biostatistics.

Olshen: In parallel with your activities in forming the SCF, there was also the rise of the Berkeley UNIX Group. Was there any articulation between the statisticians and the Berkeley UNIX Group, and did you run UNIX or VMS on your VAX?

Breiman: There are some wonderful stories about this. When I came to Berkeley, David Freedman was chairman. And Dave, I think, had a lot to do with my appointment in Berkeley because he and I had known each other for quite a while. He was also running our PDP-11. Dave got fascinated by computing.

The PDP-11 was running UNIX, except there was some guy in the Math Department that had some programs that would only run under the RISTUS operating system (which was the other operating system you could run on the PDP-11). So we had an agreement. For four days we would run UNIX and three days RISTUS. And Dave Freedman, every three or four days, had to go in, change the operating systems and reboot.

Olshen: Amazing!

Breiman: Right.

CART, ACE, PIMPLE, THE LITTLE BOOTSTRAP, BAGGING, BOOSTING AND ARCING

Olshen: Let's move on now. I want to hear about your more recent scientific interests. In no particular order, I can think of ACE (Breiman and Friedman, 1985), PIMPLE (Breiman, 1991), the little bootstrap (Breiman, 1992), bagging (Breiman, 1996a, 1996b), boosting (Freund and Schapire, 1996, 1997), arcing (Breiman, 1998), and CART^{R}. Whereas perhaps in time they will be, for now they are not necessarily all statistical household words.

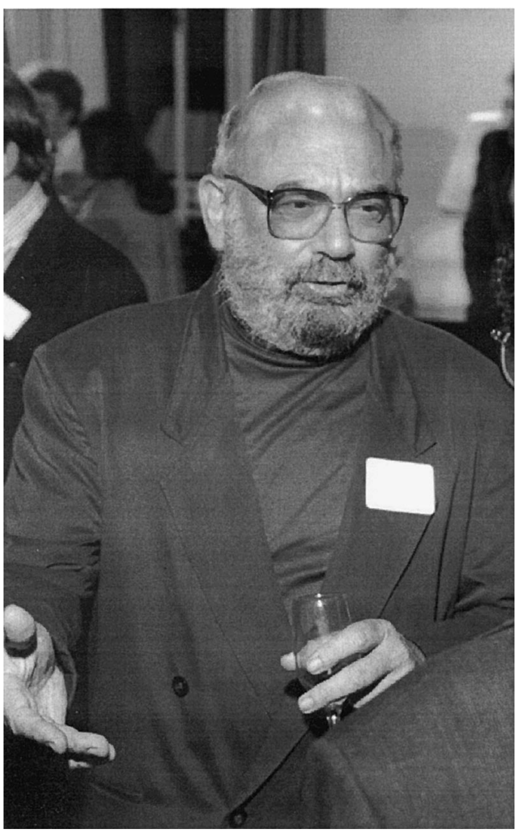

FIG. 5. *A happy Leo in receipt of a grant from AT&T for computing to the Statistical Computing Facility at University of California, Berkeley in* 1995.

Breiman: Probably for the part of my life that began after I resigned from UCLA, I think the most significant thing was CART, of which you are a part. And as you know, that was a very exciting period. All those ideas going back and forth among you and me and Jerry Friedman and Chuck Stone.

Even so, after the CART book (Breiman et al., 1984) was published, I think all of us—maybe you not as much as the rest of us—were completely fed up with thinking about trees. We just had had too heavy a dose. So our interest turned elsewhere. But Jerry and I kept talking.

Jerry and I had both been hired by Bill Meisel as consultants to TSC. So I got to know Jerry fairly well a good number of years before I came up to Berkeley. Jerry and I kept talking because I think we're two of the very few statisticians around who are actually interested in how to analyze high-dimensional data.

And one of the things we were talking about, one of the outstanding problems–this was in '86, '87—that John Tukey kept talking about was, "How do you transform variables in ordinary linear regression to get more effective prediction? Should you be using log X? Should you be using X? What?"

So Jerry and I chewed on this problem for a while. And then this thought hit us of doing this alternating smoothing technique. And I got more and more

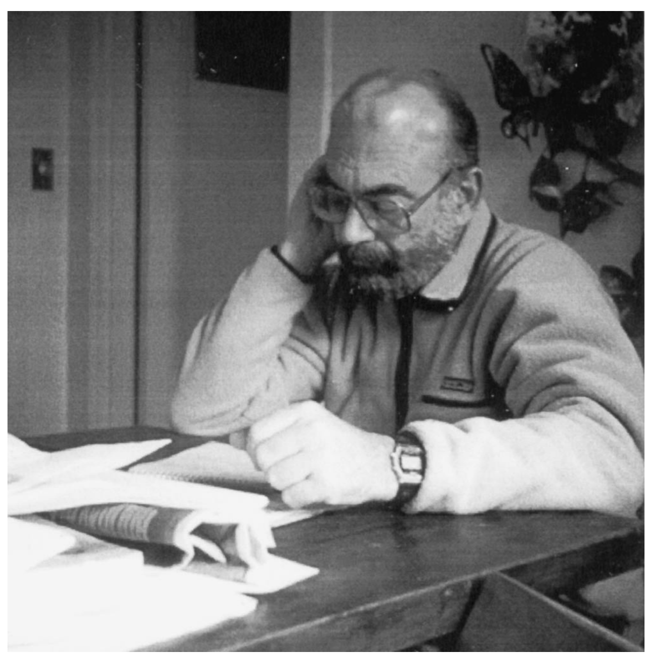

FIG. 6. *Leo working in a prior Berkeley residence*, 1985.

excited about it. Now, I had one of the old Apples, one of the first desk-top computers with 64 kilobytes of memory. But it had a color screen that you could program. So I called up Jerry and I said, "Jerry, come on up. I'll program it on the Apple and we'll see if it works."

So we put in simulated data like Y equals A log X plus B and we ran the early version of ACE on the Apple. And every time there was iteration on the Apple screen, we would see the trace of the transformation.

The Apple was so slow we would see one trace, and then Jerry and I would go have a beer and come back and look at the next trace. By midnight it was clear that the damn thing was converging toward log of X. So we knew we had it. From then on in, all we needed to do was nail it down with some theory and more experiments.

Now the story how the technique got named ACE was this: Jerry and I were drinking in a Shattuck Avenue bar one day and we were discussing what to call it. And Jerry was all for ACE because it was a snazzy name and an acronym for alternating conditional expectations. I was reluctant. And I said, "Jerry, that's a little too much. How can we call it ACE?"

And all of a sudden Jerry pointed across the street. And there was the word "ACE." He said, "Look at that" as though it were a sign from Heaven. And sure enough, what we were looking at was the Ace Hardware store sign. Seeing that convinced me that Jerry was right. It ought to be called ACE (Breiman and Friedman, 1985).

Olshen: Now, is this informative as to how PIMPLE (Breiman, 1991) got its name?

Breiman: No. Not quite with PIMPLE. I—

Olshen: Tell me a little bit about PIMPLE first.

Breiman: Okay. I was interested in functional approximation, because you can look at a lot of multivariate regression as really functional approximation with some noise added. So I was doing this reading into functional approximation in higher dimensions.

I came across this method where they approximated functions by sums of products of univariate functions. That is, if you had a kernel function $K(x, y)$, you approximated it by a sum of products $F_i(x)$ times $G_i(y)$. That rang a bell and I thought, "Why not try an approximation in regression by expanding the function as a product of simpler functions?"

Okay. And the product sign is a pi. So that's where pi came in. And the next word after pi was "implementation." Well, what's the natural acronym? PIMPLE!

Olshen: What about arcing, bagging and boosting?

Breiman: Okay. Yeah. This is fascinating stuff, Richard. In the last five years, there have been some really big breakthroughs in prediction. And I think combining predictors is one of the two big breakthroughs. And the idea of this was, okay, that suppose you take CART, which is a pretty good classifier, but not a great classifier. I mean, for instance, neural nets do a much better job.

Olshen: Well, suitably trained?

Breiman: Suitably trained.

Olshen: Against an untrained CART?

Breiman: Right. Exactly. And I think I was thinking about this. I had written an article on subset selection in linear regression. I had realized then that subset selection in linear regression is really a very unstable procedure. If you tamper with the data just a little bit, the first best five variable regression may change to another set of five variables. And so I thought, "Okay. We can stabilize this by just perturbing the data a little and get the best five variable predictor. Perturb it again. Get the best five variable predictor and then average all these five variable predictors." And sure enough, that worked out beautifully. This was published in an article in the *Annals* (Breiman, 1996b).

Then, CART also had the same sort of feature. You know, if you altered the data a little, you might get a much different tree. And so I thought, "Well, why can't I try the same thing with CART? If I alter the data for one tree, and then alter the data, grow another tree and then begin averaging them, or letting them vote for the most popular class, maybe I can increase the accuracy."

So the question was how to perturb the data? And then I realized, from some theoretical considerations that probably the best way was to start with the original training set, take a bootstrap sample from it, grow a new tree on the bootstrap sample, draw another bootstrap sample, grow another tree on it and, in the case of regression, just average them all. In the case of classification, have them vote for the most popular class.

I called this "bagging" for "bootstrap aggregation." And it worked out beautifully in terms of increasing prediction accuracy. There was a fair amount of excitement about it. And then Yoav Freund and Robert Schapire (Freund and Schapire, 1996, 1997) came up with an algorithm they called Adaboost. Adaboost was designed to combine classifiers in such a way as to drive the training error to zero as rapidly as possible. And sure enough, it did so very rapidly. With most data sets that I've looked at, it drives the training error to zero in three or four combinations of classifiers.

But it was interesting that even after the training error was zero, if you kept combining classifiers as per the Adaboost algorithm, the test set error kept decreasing long after the training set error was zero. And it kept decreasing even after you had combined, say, a hundred classifiers.

When you look at Adaboost, what it is doing in a fairly complex way is putting increased weight on those members of the training sample that had been misclassified the last time around. Then a new training set is formed by sampling according to these weights, instead of equiprobable as in bagging, and this new training set is presented to the classification algorithm. And this algorithm did marvelously well.

On most data sets that people have looked at, Adaboost did quite a good deal better than bagging did. This was a startling discovery because you could take a sow's ear and transform it into a silk purse. That is, you could take a classifier like, say, everyday vanilla CART, which was good but not a great classifier, and by using this Adaboost algorithm, which was almost trivial to program, just iterated calls to CART, turn it into a world-class classification algorithm that, by almost any standards, had accuracy as good as anything else out there, and better than almost everything else out there.

And so then the question became why: why was this complex algorithm working so well?

Olshen: Well, let's just back up. So we were talking about boosting now?

Breiman: That's right. Adaboost. When people talk about boosting, they're usually talking about Adaboost. This was a wonderful and strange sort of not very well understood algorithm.

Olshen: What about arcing (Breiman, 1998)?

Breiman: Now, arcing came about this way. When I first thought about Adaboost and looked at it carefully, it became very clear what's going on. You're putting additional weight on those things that are hard to classify, those observations that are near the boundaries between classes.

So, if you think about it this way, then there's a whole class of algorithms that can do the same thing. You can write down many algorithms that put increased weight on those cases more likely to be misclassified, and it will work pretty well. I called those algorithms arcing, for adaptive resampling or adaptive reweighting, and combining. So, for instance, I gave an ad hoc arcing algorithm that did as well as Adaboost.

And incidentally, the combination methods, boosting and bagging, don't have to be used with trees. Almost any classifier could be used with them. These were universal procedures to combine classifiers or regressions. There was a nice paper where the classifiers that were used were just simple hyperplanes, random hyperplanes (Ji and Ma, 1997). They combined these random hyperplanes using an arcing algorithm and got marvelous results.

So the whole thing became more and more intriguing.

THE CENSUS

Olshen: Continuing in the spirit of large data sets, but very different from what we've discussed thus far, can you talk about your involvement in work on the Census adjustment?

Breiman: As I recall, in about 1987 or '88, David Freedman said to me rather unexpectedly one day, "Come on. Let's go have lunch together."

As we were having lunch together, he said, "How would you like to be on this committee that's looking over plans for the 1990 Census?"

And I said, "No. I don't want that. But I'll tell you what I'll do. I will get involved later on. But I don't want to sit in on this committee." And then he got Ken Wachter to sit in on the committee.

Ken did a wonderful job. Now, as the beginning of the 1990 Census rolled around in 1989, Dave called me up and said, more or less, "Okay. I'm claiming my pound of flesh. I want you to get involved in the Census." So I said, "Okay."

And Dave and I began taking trips to various Census offices as they became set up for the year 2000. He actually walked the streets with some of them [the census takers] to see how it was being done. And then it was over and that Census produced two different sets of figures.

One was just a straight enumerative count and the other was an effort to adjust the Census. The people that the Census missed were estimated to be about 2% of the population, or five million people. The way that the adjustment was made was basically in terms of the capture–recapture idea. Shall I give a brief description of this? Okay.

It's a simple idea. Suppose you want to know how many fish are in a lake. All right. So you go there and you fish and you fish and you fish, and each time you pull out a fish, you put a red mark on its back.

You finally finish and you've counted, say, 10,000 fish. And you figure that you think you've got them all, but you're not sure. So what you're going to do is a come back a week later and you say, "Okay. Now I am going to catch a 100 fish and see how many of them have got marks on the back." All right.

So you catch 100 fish and see how many have got marks on the back. Now, if originally you had captured all the fish in the lake, every one of those 100 would have a mark on its back. But you look at them, and only 98 have marks on the back. So you say, "Oh, I didn't get 2% of them the first time around."

There are more than 10,000 fish in the pond. I actually underestimated by 2%. But here is the big problem: Somebody notices that of the two fish that didn't have marks on their backs, maybe one of them had a mark on its back, but it got rubbed off. So the undercount is no longer 2%. It's gone down to 1%.

Maybe somebody else examines the other fish very carefully and notices that there may have been a mark on that fish's back, too, in which case there's no undercount. So notice then that you have to be able to match these fish to within 1 in 100 to get an accurate estimate of the undercount.

Now, to estimate the undercount in the 1990 Census—they did the Census first. Then they followed it by a mini-Census consisting, I think, of something like 50,000 households. I'm not sure of the exact number. Then they tried to match the people in the mini-Census to the original Census to see what kind of agreement they got.

And, for instance, if 2% of the people in the mini-Census were not found in the main Census, they would just assume that the main Census had undercounted by 2%. Also, by knowing the characteristics and location of the people in the mini-Census that weren't found by the main Census, they would know the characteristics and location of the people the Census had missed.

Now, so the question was how accurate was this adjustment? This became a big court case because many states wanted an adjustment. They felt they had been undercounted, and Congressmen in the House of Representatives are apportioned on the basis of Census population.

So adjustment could mean that a few congressional seats might swing from one state to another. Plus, a lot of the billions of dollars of federal money that goes out to cities and states are based on Census figures, so which set of figures to adopt—the original or the adjusted—was a very political, big money question.

So I said, "Okay, Dave. Here is what I'm going to do." The Census Bureau had done a large evaluation of the adjustment with thousands of pages of documentation. And I said, "I'm going to look carefully at this evaluation and go through it page by page and see what it says about the accuracy of this adjustment procedure."

My experience in industry had been, if you get a data set, the first thing you do is you try to find out how good the data is—you know, how much is missing, how much is in error and so on. That's one thing you really learn by hard-knocks experience. And I began going through, and I was amazed.

More and more of this supposed adjustment was clearly due to errors than anything else. Realize this: Here on one hand you've got, say, 100 slips of paper about 100 Americans from the mini-Census; on the other hand, you've got 100 slips of paper from the main Census, and the question is how many of them can you match?

Well, many times people give different, say, middle names to the Census. They give different ages, different sexes. You can't match people that well. You need 99% accuracy in matching to get 50% accuracy in a 2% undercount. So the whole problem is that Americans, because of their mobility and immigrant status and so on, cannot be matched to within 99%.

And so if you look at it and evaluate, the errors pile up. As you know, I finally wrote a big paper in *Statistical Science* (Breiman, 1994), summarizing all the research I had done based on Census documents, and I concluded that easily over 80% of the estimated Census adjustment was due to bad data, mainly errors in matching.

Now, here is the thing that floors me, Richard. The Census Bureau admits in its own internal publications that at least 57% of its proposed 1990 adjustment was due to bad data. You'd think that after getting stung so badly by the data and the evaluation of the adjustment in 1990, they wouldn't try to do it again. Because if something fails, you just don't repeat it. Well, what they're doing instead in 2000 is most interesting. They're going to do it again, but not evaluate it!

ADVICE

Olshen: Leo, you have a record of wide-ranging interest in and contributions to statistics and statistical computing and probability and pedagogy. What advice would you give to a young person today who wants to continue in your traditions? What should he or she study and why?

Breiman: Well Richard, I'm torn in a way because what I might even tell them is, "Don't go into statistics." My feeling is, to some extent, that academic statistics may have lost its way. When I came, after consulting, back to the Berkeley Department, I felt like I was almost entering

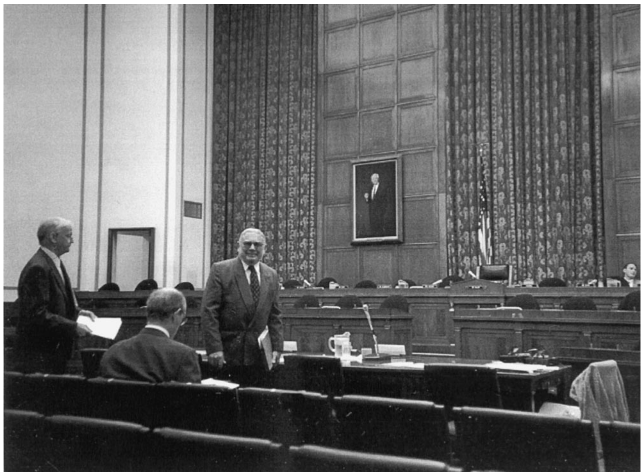

FIG. 7. *Leo after testmony before a Congressional committee on the census in* 1997.

Alice in Wonderland. That is, I knew what was going on out in industry and government in terms of uses of statistics, but what was going on in academic research seemed light years away. It was proceeding as though it were some branch of abstract mathematics. One of our senior faculty members said a while back, “We have to keep alive the spirit of Wald.” But before the good old days of Wald and the divorce of statistics from data, there were the good old days of Fisher, who believed that statistics existed for the purposes of prediction and explanation and working with data.

Before you came this morning, I pulled out Webster’s dictionary and looked for the definition of statistics, and here is how it goes: “Statistics, facts or data of the numerical kind assembled, classified, and tabulated so as to present significant information about a given subject.” When used with a singular verb, it is, quote, “The science of the assembling, classifying, tabulating, and analyzing such facts or data.”

Now, little of that is going on in the academic world of statistics. For instance, I was looking at *The Annals of Statistics* and I estimate that maybe 1 paper in 20 had any data in it or applied the methods there to any kind of data. The ratio is not much higher in the *Journal of the American Statistical Association*. So my view of what’s fascinating in the subject of statistics and the common academic view are pretty far apart.

In the past five or six years, I’ve become close to the people in the machine learning and neural nets areas because they are doing important applied work on big, tough prediction problems. They’re data oriented and what they are doing corresponds exactly to Webster’s definition of statistics, but almost none of them are statisticians by training.

So I think if I were advising a young person today, I would have some reservations about advising him or her to go into statistics, but probably, in the end, I would say, “Take statistics, but remember that the great adventure of statistics is in gathering and using data to solve interesting and important real world problems.”

THE ARTIST

Olshen: Thank you. The room in which we are having this conversation, even the house itself, is full of reminders of the breadth of your interests, not least your life as an artist. What about that side of Leo Breiman?

Breiman: Well, during my consultant time in Los Angeles, I could actually make enough money in a fairly small number of hours that I had a fair amount of leisure time. And I remembered that when I first came up to Cal, my first roommate, of whom I was very fond, was an art major. He encouraged me to sculpt and I made some things out of clay. I didn’t think much of them.

Then when I had this leisure time in Los Angeles much later on, I took a sculpting class at UCLA Extension. There was a wonderful teacher who encouraged us to do all sorts of wild things. I

FIG. 8. *Leo paying homage to statistical predecessors in Hampstead, London in* 1992.

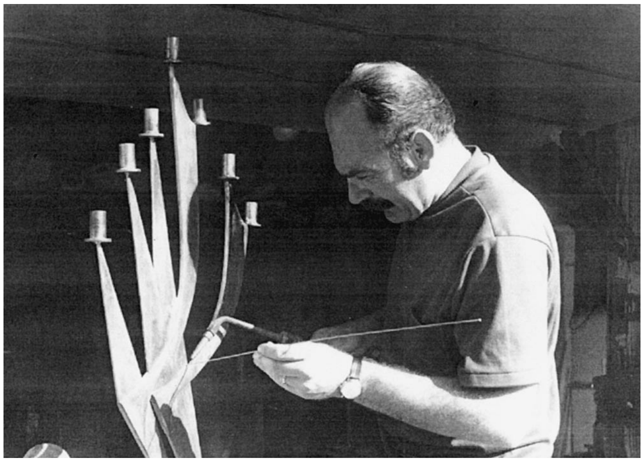

FIG. 9. *Early in Leo's career as a sculptor*, 1973.

FIG. 10. *Notice of an exhibit of Leo's works in Berkeley in* 1998.

also met this fellow, who came and talked to us once, who had a foundry for bronze casting in Venice—Alf Peterson. Alf was a marvelous guy And he said, "If you want to practice, come down to the foundry. I'll show you how to do things."

I made a figure out of clay and wax and chicken screen. That was my first angel over there (indicating) in the corner. And I took it down to Alf's place and he said, "Okay. I'm going to show you how to cast bronze."

He showed me how to make molds of the angel, how to make the negatives, how to put it in the ovens, how to pour the bronze and so on. I was intrigued and absorbed. I thought that it was terrific And so I've kept going ever since then and have loved it.

Lately, though, I've begun working in stone, which is very interesting. So you'll see some stone heads also drifting around. And that's slow work. You've got to keep tinkling, tinkling with the hammer and chisel.

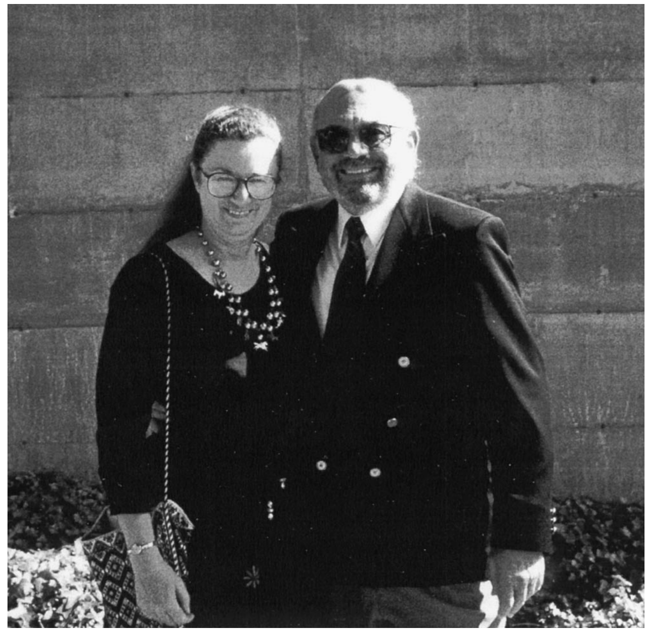

FIG. 11. *Leo and Mary Lou Breiman at a Berkeley Department of Statistics graduation ceremony*, 1995.

Olshen: I guess I'm done with my questions. Have you any brief parting thoughts? Has this been fun?

Breiman: This has been a lot of fun, Richard. You know, sometimes I feel sad about statistics. There are so many smart people in it and I hope it gets better before it gets worse.

Olshen: Perhaps your own example will have a positive effect on others, Leo. Thank you.

Breiman: Richard, it's been a pleasure.

REFERENCES

BREIMAN, L. (1957). The individual ergodic theorem of information theory. *Ann. Math. Statist.* **28** 809–811. [Correction (1960). *Ann. Math. Statist.* **31** 809–810.]

BREIMAN, L. (1960). Optimal gambling systems for favorable games. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 60–77. Univ. California Press.

BREIMAN, L. (1963). The Poisson tendency in traffic distribution. *Ann. Math. Statist.* **34** 308–311.

BREIMAN, L. (1968). *Probability Theory.* Addison-Wesley, Reading, MA. [Republished (1991) in *Classics of Mathematics*. SIAM, Philadelphia.]

BREIMAN, L. (1991). The II-method for estimating multivariate functions from noisy data (with discussion). *Technometrics*. **33** 125–160. (Awarded the Youden Prize as the best expository paper of the year in *Technometrics*.)

BREIMAN, L. (1992). Submodel selection and evaluation in regression—The X-fixed case and little bootstrap. *J. Amer. Statist. Assoc.* **87** 734–751.

BREIMAN, L. (1994). The 1990 Census adjustment—Undercount or bad data? (with discussion). *Statist. Sci.* **9** 458–475.

BREIMAN, L. (1996a). Bagging predictors. *Machine Learning* **26** 123–140

BREIMAN, L. (1996b). The heuristics of instability in model selection. *Ann. Statist.* **24** 2350–2383.

BREIMAN, L. (1998). Arcing classifiers (with discussion). *Ann. Statist.* **26** 801–849.

BREIMAN, L. and FRIEDMAN, J.H. (1985). Estimating optimal transformations in multiple regression and correction (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619. (Theory and Methods Paper of the Year.)

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA. (Since 1993 this book has been published by Chapman and Hall, New York.)

FREUND, Y. and SCHAPIRE, R. (1996). Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference* 148–156.

FREUND, Y. and SCHAPIRE, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139.

JI, C. and MA, S. (1997). Combinations of weak classifiers. *IEEE Trans. Neural Networks* (Special Issue) **8** 32–42.