

A Conversation with Herman Chernoff

John Bather

Abstract. Herman Chernoff was born in New York City on 1 July 1923. He went to school there and later received the B.S. degree from the City College of New York in 1943, majoring in mathematics with a minor in physics. For a year and a half, he worked as a junior physicist with the U.S. Navy, before joining Brown University for graduate work in applied mathematics. His studies were interrupted by a short period in the U.S. Army, and then his interest in statistics led him to complete his Ph.D. thesis at Columbia University under the supervision of Abraham Wald. At Brown University, Herman met Judy Ullman. They have been married since 1947 and have two daughters, Ellen and Miriam.

Herman worked for the Cowles Commission at the University of Chicago and then spent three years in the Mathematics Department at the University of Illinois before joining the Department of Statistics at Stanford University in 1952, where he remained for 22 years. He moved to M.I.T. in 1974, where he founded the Statistics Center. Since 1985 he has been in the Department of Statistics at Harvard.

Professor Chernoff has been honored for his contributions in many ways. He was President of the Institute of Mathematical Statistics and is an elected member of both the American Academy of Arts and Sciences and the National Academy of Sciences. The book *Recent Advances in Statistics* published in honor of his 60th birthday in 1983 contained papers in the fields where his influence as a researcher and teacher has been strong: design and sequential analysis, optimization and control, non-parametrics, large sample theory and statistical graphics.

The following conversation took place at Chapel Hill at the meeting of the Bernoulli Society and the Institute of Mathematical Statistics in June 1994.

THE EARLY YEARS

Bather: Let us begin with your New York background. You were born in 1923 in New York, where you were brought up and went to high school and college.

Chernoff: At that time New York City had a special high school which was the prep school for CCNY, the City College of New York. When I was in junior high school, I was invited to take a competitive exam to get into Townsend Harris High School, and

I could not resist the challenge. The consequence was that I went to this high school where I did not properly belong: everyone was a lot smarter than me.

Bather: But you survived it and then, because it was a prep school, you went on to CCNY, where you met several people who have figured strongly in statistics and other subjects.

Chernoff: Yes, there is a picture of the Math Club from the CCNY yearbook of 1939 that appeared in an earlier edition of *Statistical Science* for a conversation with Herb Solomon. For some reason Herb was not in the picture. But Milton Sobel, Kenneth Arrow and Oscar Wesler, who is here at North Carolina, and I were in it. The students at Stanford once found this picture and posted it on the notice board with the sign saying “Know Your Faculty.” Most of the people in that picture became professors of mathematics.

Bather: That was a nice touch! You then went to Brown University, but before that you worked for

John Bather is Professor of Statistics, School of Mathematical Sciences, University of Sussex, Brighton, England, BN1 9QH (e-mail: j.a.bather@sussex.ac.uk).



FIG. 1. *Giorgio Dall'Aglio speaking to H. Chernoff (right) while Chernoff shows off the diploma of an honorary degree from La Sapienza (University of Rome) April 1996.*

a time as a junior physicist with the United States Navy.

Chernoff: I graduated in January 1943, and the United States had entered the war. I was thinking of applying to graduate school but a telegram came inviting me to work for the Navy as a junior physicist. I worked for the Navy at the Dahlgren Proving Ground in electronics for about a year and a half, building equipment and fixing radios and counter chronographs used to measure the speed of shells fired by naval guns. On a few occasions I had the opportunity to use some statistical ideas, and that helped later in my decision to go into statistics.

Bather: Your skills were very helpful this morning in sorting out the equipment that I failed to get to work. Let's move on to Brown University. I believe that at this stage you first came across the famous paper by Neyman and Pearson.

Chernoff: It was actually at City College, where I took a couple of courses in statistics. At City College I was given a paper by Neyman and Pearson to read over the weekend. That was quite a traumatic experience. It took me a long time to realize that it was as simple as it seemed to be. It required a complete reorganization of my brain cells to adapt to it and I was quite profoundly impressed.

Bather: Well, so was the whole of the statistics profession at the time, or most of it.

Chernoff: With some resistance.

Bather: At Brown University, you started off by specializing in applied mathematics.

Chernoff: The applied mathematics program there was strong in theory. I enjoyed the theory, but was dissatisfied with my attempts at original research. I did not feel that I had the physical insight that would enable me to do the sort of research that went on in fluid dynamics. I could do the mathemat-

ics once the mathematical problem was formulated, but I did not think that I had the insights which would enable me to formulate the right mathematical problems. I felt that I had much more insight into statistics, and that is why I decided to switch to statistics after I wrote my Master's thesis.

Bather: Going back a little way, I would like to ask you about your military service, which took place after World War II. You went into the U.S. Army for a short period.

Chernoff: I was drafted just after the war ended and, after three days basic training, operated as a clerk in the night shift at the separation center for returning army men. It was in a way the first vacation I had had in a long time, because I worked about six hours in the evening and had the whole day free. On the weekends, I rode to Brown University and I was actually taking a couple of reading courses while I was there. One of the professors asked whether I wanted to do something where my technical skills could be used. I said "Oh yes." He told me that he would arrange it. As a consequence, I was transferred and I had to take real basic training, but I found out that my future would be as a clerk in the quartermaster corps abroad. At that time, I had just obtained an NSF predoctoral fellowship, which made it seem that a good deal would be wasted if I spent the next few years in the army. Much to the horror of my friends, who thought that I had ruined my future military career, I applied for discharge. Strangely enough it was granted.

Bather: That was quite something. So how long were you in the army?

Chernoff: A hundred days, so I did not get to complete my basic training.

Bather: You then went back to Brown but first you attended summer school.

Chernoff: No, actually, I spent a few months at Brown and then I found out about a summer session where they were opening up North Carolina State College and the University of North Carolina to a statistics program. Gertrude Cox had formulated a grand plan of what is now the Research Triangle. She had started the implementation of it and had appointed Harold Hotelling, who was trying to get Wolfowitz also to come from Columbia. But the people at Columbia University saw what was happening and appointed Wald to start a statistics department there, so as not to lose their statisticians. That was the beginning of a big push for statistics in the United States. Until then statistics was a very small field and quite isolated, although Harold Hotelling had been going around all over the country trying to explain how important it was to have more technically trained statisticians.

Bather: Would you say that the experience of the military during the Second World War had a strong effect on the position of statistics?

Chernoff: It had a very big effect. I think the statisticians and the operation researchers had proved to be extremely useful on many occasions and had made a profound impression on the people with influence. Since then I think that effect has died out. While the number of statisticians has grown, physicists tend to have a great deal of influence in the armed forces, and they don't fully appreciate the power of statistical thinking. As a result there is quite a lot of waste in the testing of army equipment.

GRADUATE SCHOOL: BROWN AND COLUMBIA

Bather: You attended the summer course in North Carolina, which was at Raleigh not at Chapel Hill. Had you already come across Wald's work before you met him?

Chernoff: When I was at Brown, Henry Mann had shown me Wald's 1939 paper on decision theory and that again was another revelation to me, but it was easy to absorb after having had contact with the Neyman–Pearson paper. Before I went to North Carolina, I read the papers that Wald had written on sequential analysis and so I had some exposure to modern statistical theory, mainly from reading. William Feller had been giving an excellent course on probability which was the basis of his first textbook, but he did not seem to have much insight on statistical theory and I was pretty much self-read and rather ignorant: a book by Wilks gave me some help.

Bather: It is very difficult for younger statisticians to appreciate how rapidly things were developing: the time lapse between the Neyman–Pearson paper and Wald's rather sophisticated view of statistical theory was relatively short, only about 10 years I think.

Chernoff: Yes, it was rather short, but I do not think there was that much difference between the two views, but maybe this is hindsight. A lot of things that seem trivial now were not so easy at the time, but I tended to think of Neyman–Pearson theory as leading rather logically to decision theory. Wald's decision theory really must unconsciously have been based on Von Neumann's theory of games. I think Wald had probably been exposed to it, and my suspicion would be that he had heard of the theory of games and completely forgotten it and then rediscovered it with his decision theory.

Bather: I think this happens to us all. You finally met Wald at Columbia and he agreed to take you on for a Ph.D., but only after inspecting a paper you had published in applied mathematics.

Chernoff: He was reluctant to accept another student, especially one who was not registered at Columbia. At Brown, I had already taken all my qualifying examinations and I just had to write a dissertation. Also he was afraid that I had no formal statistics background, but in those days very few had a full statistics background. He insisted that I take some courses at Columbia, which I was quite happy to do. After looking at my proposal for the dissertation subject, he suggested that I was not quite up to it, or it was not quite up to a thesis. I am not sure which. He proposed a couple of topics, one of which was actually a view of the Behrens–Fisher problem which resembled something that Welch did at approximately that time. The other topic that Wald proposed was to solve some special problems in decision theory. I was not particularly interested in doing that, as these problems seemed too routine, but it would have been a helpful addition to the literature at that time. Anyway, I worked on the Behrens–Fisher problem.

Bather: That is a problem that has been remarkably durable and people are still working on it for all I know. This led to your first published paper in statistics, I believe?

Chernoff: That's correct. Part of my thesis was published around 1949: it was called "Asymptotic Studentization in testing of hypotheses." The Welch approach to the Behrens–Fisher problem assumes that there is a solution to the problem and Welch found an approximation. It did not seem clear to me that there was a solution to the problem. I was attempting to get an asymptotic solution. The method that was proposed to me by Wald worked very well asymptotically, but it turns out that a later proposal by Welch and Trickett, and recently implemented by Mark Vangel, gives fantastically good approximations to a problem that has no solution.

Bather: Do you mean, in saying that there is no solution, that the problem was not clearly posed?

Chernoff: No, the problem posed was to find a smooth function of the difference of two sample means and the two sample standard deviations, to use as a test statistic for testing the null hypothesis that the difference of the two population means is zero. This statistic, or rather the critical region based on it, must have Type I error probability exactly 0.05 for all values of the population standard deviations. Linnik proved that this problem has no solution.

You could get a test with significance level 0.05 by using randomization: for example, by rejecting the null hypothesis if a 20-sided die falls on side 1 and accepting it otherwise. That solution would be meaningless because the resulting test would have power equal to 0.05 everywhere. The object was to find a smooth function of the sufficient statistics which tended to be larger than some critical value when the hypothesis is false and which would be less than the critical value with probability independent of the nuisance parameters when the two means were equal.

Bather: Before we leave the subject of Columbia University, perhaps we could mention some of the people that you met there.

Chernoff: Among the regular faculty Howard Levine had just been promoted to Assistant Professor, I think, and Wolfowitz came back from North Carolina at Wald's invitation. Ted Anderson had recently been appointed to the faculty; that was the regular faculty. But they depended heavily on visitors. J. L. Doob was giving a course on stochastic processes, and R. C. Bose was giving a very interesting course in experimental design. While I was there, E. Pitman appeared and gave a course which I never fully attended as I was getting ready to leave. I met Franco Modigliani, who later became a Nobel Laureate in Economics, in Anderson's time series course, and I believe that Hoeffding was attending some of the courses while I was there. It was an exciting time with exciting people.

Bather: Some of the exciting people were at that time students?

Chernoff: Charles Stein had just got out of the army and he was trying to prove the optimality of the sequential probability ratio test. He got some partial results but they were not adequate, and so the department gave him his doctoral degree for the work that he had done in the army on two-stage tests. The other people who were there as graduate students included M. Sobel, L. Weiss, I. Olkin, Jack Kiefer, R. Bechhofer, G. Seth, R. Sitgreaves and H. Teicher. As I was leaving, Bill Kruskal came upon the scene. I had met Bill at the naval proving ground and after the war he had gone back into his father's fur business. But he decided after a year or two that he really wanted to go back to academia and so, as I was leaving, he was arriving. It is difficult to remember everyone. It was interesting in the sense that many people who became substantial figures in probability and statistics were around at that time. Perhaps it was easier then to become prominent. These days there are just as many bright people, but they have a lot tougher time getting tenure.

THE COWLES COMMISSION

Bather: Going back a little time, we must mention that while you were at Brown University you met Judy and you were married in 1947.

Chernoff: I went to Columbia in January 1947. Judy and I were married in September 1947. We lived in New York for about eight months until I finished my dissertation and went as a research instructor to the University of Chicago with the Cowles Commission for Research in Economics. (Since then the Cowles Commission has moved to Yale, where it is called the Cowles Foundation). Ted Anderson had some connection with the Cowles Commission. While I was still a graduate student he proposed to them that they appoint me, and I was working for them while I was at Columbia, using the computer equipment at the Watson Laboratories of I.B.M. that were located right next to Columbia University. At the Cowles Commission at that time were Herman Rubin and Kenneth Arrow. While I was at the Cowles Commission, Kenneth Arrow spent one summer at the Rand Corporation and at the end of the summer came back with two accomplishments which were very impressive. One was the collaboration with Girshick and Blackwell on the proof of the optimality of the SPRT. Wald and Wolfowitz had presented another approach to proving the optimality but their approach had been seriously flawed by measure theoretical problems, among other matters. Kenneth Arrow,



FIG. 2. Photo of H. Chernoff 1948–1949.

Blackwell and Girshick derived, in my opinion, a much sounder variation of the Wald–Wolfowitz idea. Their sophisticated backward induction approach was really the basis of dynamic programming.

Bather: They in effect invented dynamic programming, which was not published by Bellman until 1957.

Chernoff: I have spoken to Blackwell about it. While he regards sequential analysis and dynamic programming as the same subject, he said that he and Girshick had not fully comprehended the implications. Bellman began a thorough exploitation of these ideas a little later, but well before his book appeared. The other important work that Arrow did was the major basis for his Nobel prize. It was the work he did on the question of whether there is a general coherent way of forming a universal preference ordering of choices based on those of the individuals involved.

Bather: In other words, can people agree?

Chernoff: The question is whether there is a general way of forming a reasonably acceptable consensus if they do not agree. To prove that there is no way was a great accomplishment because the question had not been clearly formulated previously.

Bather: That was certainly a powerful result. You also had contact with L. J. Savage during your time at the Cowles Commission?

Chernoff: Yes. He was very close to Allen Wallis and Milton Friedman. They had worked together at the Statistics Research Group at Columbia University during World War II. Milton Friedman, who is a Nobel Laureate in Economics, was critical of the Cowles Commission approach to econometrics, and there seemed to be some friction between that trio and the people of the Cowles Commission.

At one time Savage felt that he had resolved the choice of criterion problem in decision theory. In the decision theory approach there remained a question of how you select among the various decision rules when there is not a uniformly best choice (which is usually the case). Wald had tentatively suggested the minimax principle. I recall hearing some of the students at Columbia wonder why he did not establish this as the correct way to go. They thought Wald was being too tentative. However, with some thought, it was obvious that the minimax principle had serious shortcomings. If you have a choice between committing suicide or not taking action, in which case you might lead a normal life or you might die a horrible death, the minimax principle tells you to avoid any possibility of a horrible death by committing suicide. That did not seem to be quite right. Savage thought he had resolved that problem by suggesting that the loss could be separated into

an unavoidable loss and a regret for doing worse than we had to if we knew what the state of nature was. Wald's examples always were cases where he was minimizing the regret.

Savage proposed that minimax regret was the resolution to the problem. When I played with that notion, I found that it failed to satisfy one of Arrow's requirements for a universal choice function. That was the principle of irrelevant alternatives. If you had the choice of a , b , c or d , you might decide that a is the best. However, if someone tells you d is not available, it may then turn out that among a , b and c , you prefer b . Minimax regret sometimes behaved this way, and that was a violation of this principle of irrelevant alternatives. I brought this to Savage's attention and, after arguing futilely for a little while that it proved how good his criterion was, he finally agreed that it was wrong. He felt then that perhaps we should be elaborating on De Finetti's Bayesian approach, which he had come across. (He was a voracious reader.) Meanwhile, I decided that I would list the set of principles that I felt an objective statistical decision rule should satisfy. I wrote a discussion paper on rational selection of decision functions which came up with a contradiction. I sat on it for a few years until I finally published it in *Econometrica*. Ultimately, from the point of view of philosophical foundations, I think the Bayesian position has won the day; if there is to be what we now call a coherent procedure, it has to be a Bayesian procedure. The problem in inference, of course, is that we are not capable of carrying out that procedure.

Bather: It seems surprising now that people were not aware of the Bayesian principle at the time. Perhaps, in effect, De Finetti had to reinvent it.

Chernoff: I think Fisher had been very influential in suppressing it and I think the philosophical foundations had not been very clear. It may be that they are still somewhat foggy. Ramsey was supposed to have worked on that at an earlier time, but his work was not well known. A lot of the support for the Bayesian principle was in the nature of rhetoric but, when you look at the fundamental principles of coherence in a systematic fashion, it comes out. In fact, shortly after that conversation with Savage in 1949 about the minimax principle and minimax regret, Herman Rubin wrote a two-page derivation to the effect that a coherent procedure must be a Bayes procedure. That paper has probably been lost to posterity; I thought I had it on file, but it has gone.

Bather: At the Cowles Commission in Chicago, who else was there?

Chernoff: Among the economists there was Jacob Marschak, who I think would have got a Nobel prize if he had lived long enough. Tjalling Koopmans and Franco Modigliani were around at the time. There were several others.

Bather: After your time with the Cowles Commission, you spent three years at the University of Illinois in the Department of Mathematics before you went to Stanford. I believe it was Kenneth Arrow that recruited you to go to Stanford University?

Chernoff: Yes, that's right. He called me after I had been at the University of Illinois for two years and suggested that I come out to visit Stanford. I spent half a year from June 1951 until January 1952 as a visitor to Stanford University, where they had recently started a statistics department. Albert Bowker, who set up the department, was influential with Terman, who was then the provost at Stanford. Stanford University was being very progressive, building itself up to become a top-rate university. When I visited the Stanford Statistics Department there were several other visitors. There had been a political disaster at the University of California, because of the requirement that the faculty sign an oath of allegiance.

Bather: This was the McCarthy era?

Chernoff: That's right.

Bather: I did not know much about that because I was too young, but it was a terrible time for some senior academics like Charles Stein.

Chernoff: Charles Stein had left the University of California because of that oath and had gone to the University of Chicago, and Eric Lehmann apparently was very uncomfortable and was visiting Stanford University at that time. While I was there, Milton Sobel, Ben Epstein and David Blackwell were there as visitors. Herman Rubin and Arrow had gone to Stanford directly from the Cowles Commission. Bowker had also brought in Girshick from the Rand Corporation. Girshick was, I guess, the senior statistician in the department and it was a very lively group.

Bather: So they were building a very strong team and you felt very tempted to join them?

Chernoff: That's right. I liked the University of Illinois very much. Being a provincial from the Bronx, the small town college life at Illinois appealed to my personality. I did not care for my introduction to California that well; the social life did not seem to be quite as nice. The intellectual environment in the Statistics Department at Stanford, however, did appeal to me. Also, in the first few months that I was there, some research problems came to my attention because of connections

with Stanford's ONR (Office of Naval Research) contract. These problems began something that played a very important part in my future research.

MOVING TO STANFORD

Bather: In 1952 you and Judy moved to Stanford, where you were to stay for 22 years. No doubt you settled down happily. When I met you there more than 10 years later, you were extremely well established and everyone thought you would stay there forever.

Chernoff: Yes, I think it was a shock to many people when I left. I used to tell people that after 22 years at Stanford, I felt I had to change my wife or my job; and my wife would not let me go. I told that to one man and he said that his wife did let him go. I didn't have an answer to that one.

Bather: Let us go back to the 1950s and talk about some of the research you did in those early years at Stanford. Tell me about the result that is always known as Chernoff's lemma, but that you attribute to Herman Rubin.

Chernoff: In the work that I did on my visit to Stanford, there were two papers. One was called the "Measure of asymptotic efficiency" and the other "Locally optimal designs," and both of these came out of problems that had important relevance to design of experiments. The first one was, as far as I know, the first application of large deviation theory to statistical problems. I had two simple hypotheses and I noticed that for discriminating between the two, in the range outside the 5% significance level, we were in the case of large deviations. Being ignorant at the time of the beautiful work by Cramér, I derived a slightly overlapping, but not nearly as elegant, result which showed that asymptotically the probability of falling in the tails approaches zero at an exponential rate, under mild assumptions. Rubin claimed that part of my derivation, giving the lower bounds, could be obtained much more easily. After working so hard, I doubted it very much. He showed me the Chebyshev type of proof that gives rise to what's now called the Chernoff bound, but it is certainly Rubin's. When I wrote up the technical report, I mentioned his assistance but when I submitted the paper for publication, I left it out because it was so trivial and it never occurred to me that this would be one of the things that would lead to my fame in electrical engineering circles. That inequality turned out to be a very important result as far as information theory is concerned, and so the lower bound has been called the Chernoff bound ever since. I am very unhappy about the fact that I did not properly credit Rubin at that time because

I thought it was a rather trivial lemma, but many things are only trivial once you know them.

Bather: We all benefit from hindsight from time to time. Another topic closely related to this was your interest in optimal design.

Chernoff: The large deviation result I obtained enables one to measure the efficiency of an experiment for deciding between two simple hypotheses. The other paper I wrote then had to do with estimation. If you are interested in estimating a parameter when there are several nuisance parameters, I was able to show that if there are k parameters altogether and you are interested in one function of these parameters, then you need at most k of the experiments that are available to be performed, in certain proportions, to get an optimal design. That overlapped a result by Gustav Elfving which had been applied to regression designs and was very elegant.

Sometimes you have an experiment which is not a regression experiment, but each observation or each experiment that you perform can only give you information about one function of the parameters. That means that the rank of the information matrix is 1. These are not necessarily regression experiments, but they are equivalent, from the point of view of the Fisher information, to regression experiments, and that means that you can use the Elfving geometric solution to solve these nonregression problems. The beauty of this is that it is geometrically clear what is going on. It gives you an insight on how to get useful designs when the optimal design is not necessarily very practical.

Bather: We have talked about problems of estimation, but about this time you had ideas on the testing of hypotheses, which led to you thinking in a sequential way.

Chernoff: The first paper was about the testing of hypotheses, but it was about testing one simple hypothesis against a simple alternative. Because of these papers, which had obvious implications for experimental design, I became interested in the notion of experimental design in a much broader context, namely: what's the nature of scientific inference and how do people do science? The thought was not all that unique that it is a sequential procedure: one carries out an experiment; on the basis of that experiment one learns something, and as one learns something, one is able to design a much sharper and more informative experiment. So the question of sequential experimentation became of interest and, as far as I know, no one had ever proposed a formal theory for sequential experimentation. The solution for a simple problem in sequential estimation required a fixed sample size and it seemed clear that for prob-

lems in estimation there was not much to gain by a sequential theory. As later developments showed, the results are really higher-order results.

Bather: The gains from a sequential approach are secondary rather than primary.

Chernoff: In estimation, but in testing it made a big difference. In locally optimal estimation, the optimal result was a mixture of experiments. It turns out that in sequential testing of hypotheses, when you are testing against composite hypotheses, there is also an incentive to use a mixture of experiments.

The Kullback–Leibler information numbers, which had come up as very useful measures of information in the problem of testing a null hypothesis at a fixed significance level against an alternative, turn out to be extremely important in sequential experimentation. In fact, even in the classic sequential analysis results, the rate at which the log-likelihood ratio statistic approaches a boundary is determined by a number which turns out to be Kullback–Leibler information. It played a key role in sequential experimentation and, ever since then, I seemed to be married to Kullback–Leibler information. It comes up in a lot of my work.

Bather: It could have been called the drift of the log-likelihood ratio sequence, but it was not thought of in that way at the time.

Chernoff: Not by Kullback but I think that is exactly how Wald thought of it, because of the SPRT. I tend to think of it as the rate at which the posterior probability approaches zero.

Bather: Yes, that is another approach.

We have talked about working on estimation, on testing parametric hypotheses, but you also did some work at that time on nonparametric statistics with Richard Savage. Can you tell us about that?

Chernoff: When I got into statistics, I had a relatively strong background in complex analysis and asymptotic theory, and in fact some of the people in the statistics department were afraid that I was a mathematician in disguise. I like mathematics and I like statistics even better, but I had not thought seriously about nonparametric statistics. Around 1956, Myer Dwass and Richard Savage were visiting Stanford University, and Lehmann and Hodges had conjectured that the nonparametric competitor to the t -test not only had full efficiency for translation alternatives if the distributions were normal, but that the asymptotic efficiency relative to the t -test was greater than 1 if the distributions were not normal. Savage and Dwass were very interested in this conjecture and brought it to my attention. One of the things that I had studied as an undergraduate and as a graduate student was calculus of variations,

and it seemed to me that I could easily attack the variational problem. I looked at the problem, assuming that the asymptotic normality of the distribution is easily derived and well known, and, sure enough, the result that the nonparametric t -test has minimal efficiency of 1 at the normal distribution came up very easily, much to my surprise. However, it turned out that the asymptotic normality that was needed when the underlying distribution is not normal was not well established. We had to work on that and it occurred to me then to take an average with respect to the sample cdf. If you replace the sample cdf by the true cdf and apply Taylor expansions and beat the remainder terms to death, that would be a way to prove asymptotic normality. We were quite successful.

Bather: This gave rise to the result known as the Chernoff–Savage theorem. Am I right in thinking that you effectively showed that a nonparametric procedure was somewhat more efficient than the standard normal procedure?

Chernoff: That is correct for translation alternatives, but that was the Lehmann and Hodges conjecture.

SEQUENTIAL DESIGN

Bather: You were giving serious thought to the sequential design of experiments. I think we should explore how these ideas grew in the 1950s.

Chernoff: I had proved the asymptotic optimality of a sequential design procedure using randomized experiments and Kullback–Leibler information numbers and a trivial stopping rule, assuming that there were only a finite number of states of nature and a finite number of distinct experiments available. Then I asked students of mine to generalize these results to the case where there were infinitely many experiments and infinitely many states of nature. Stuart Bessler generalized my result to the case where there were more than two terminal decisions, and Arthur Albert to the case where there were possibly infinitely many states of nature and alternative experiments available. Albert came across a difficulty in the case where there are many states of nature. If the two alternative hypotheses touch one another, the Kullback–Leibler information number then degenerates to zero and the result blows up.

Bather: This was a crucial issue. It led to the introduction of the notion of an indifference zone that some people developed and were led in another direction.

Chernoff: Wald had introduced the notion of indifference zones in his sequential probability ratio

tests when he was thinking of the problem of testing composite hypotheses such as whether the mean of a normal distribution is positive or negative. His approach had been: we will assume an indifference zone and then test the hypothesis that the parameter is at one end of the indifference zone against a simple hypothesis at the other end. His approach to composite hypothesis testing was rather primitive. At that time some of the interpretations of it were definitely wrong. Wald understood the situation better than some of the people who interpreted it, but what happened was that further attempts had been made to try and attack this problem and one was by Kiefer and Weiss. They formulated the problem as a three-hypotheses problem, so you might be willing to test whether the mean is 1 against the alternative that the mean is -1 , allowing for the possibility that the mean is zero, and establish the criterion of minimizing the expected sample size when the mean is zero. I did not care for that formulation at all because it wasn't a proper decision-theoretic formulation. It seemed to me that the proper decision-theoretic formulation was to take the cost of making the wrong decision into account in each case, and that this cost should be zero when the mean is zero. It was also necessary to introduce the cost of sampling and then optimize with respect to the Bayesian criterion.

Bather: So you included both the sampling cost and the terminal decision cost and this led very naturally to the idea of doing it sequentially.

Chernoff: The Kiefer–Weiss proposal was also sequential, but it seemed an unsound formulation from a decision-theoretic point of view, because it neglected consideration of the expected sample size when the true mean is plus or minus 1.

Gideon Schwarz was visiting Stanford at the time and I proposed the problem with what I thought was the right formulation. That is when he derived the procedure that is under his name. Then from the case where there are just three possibilities, it was very easy to generalize to the case where the mean could vary continuously, but there would be a cost of zero for making a wrong decision between -1 and $+1$. Those were beautiful results, but I was not satisfied with them because it seemed to me that a much more natural cost for the wrong decision in testing whether it is positive or negative was the absolute value of the mean.

Bather: It turned out that this view was crucial and you decided to dig a little deeper. I recall that you studied a particular problem in this context, which was the problem of deciding the sign of a normal mean. This was a prototype for a great deal of future work in which I became involved.

Chernoff: You embarrassed me because when I was presenting one of my great breakthroughs at Cambridge, after working on it for a long time, it turned out that you had strongly overlapping results at the same time. In fact you had obtained good bounds from which you could derive some of the asymptotic approximations which I had worked very hard to get.

Bather: The story from my point of view is this: Dennis Lindley had attended the Fourth Berkeley Symposium and he came across your paper on “Sequential tests for the mean of a normal distribution.” When he returned to Cambridge, where I was a research student, he posed the problem and I decided to give it some thought. A little later, when you visited Cambridge, we had our first conversation together.

Chernoff: Maurice Walker had said that you were working on the problem but that you had not made any progress. That is the time that Chambernowne came to my lecture and it turned out that he and Turing had proposed the same problem and got some results on it some years before.

Bather: It was certainly a problem which fascinated me and led me to visit you in Stanford somewhat later, 1964–65, by which time the whole business had grown into a minor industry. I believe you published altogether four papers on that testing problem and I myself published one. You worked with another Englishman at the time, John Breakwell.

Chernoff: I went on leave in 1961 to the London School of Economics. I was going to spend six months at LSE and three months in Rome. Before I went on leave, I had discussed the problem with Breakwell and I told him that I thought the case where you had been observing the data for a long time could easily be solved by power series expansions. He agreed with me and he said he would try to attack that problem. I proposed that I would attack the problem at the beginning of time, as the time t of observation approaches 0, which would be more difficult: I had published a conjecture about the nature of the solution. Breakwell actually succeeded in using confluent hypergeometric functions to solve his problem. I was having trouble with the small- t part of the problem, but managed to simplify some of the analysis. I changed to the (y, s) scale, y being the Bayes estimate of the mean and s its variance.

Bather: So you worked in terms of the parameters of the posterior distribution of the unknown mean.

Chernoff: Which simplified life considerably. While I was in London, I worked hard on it and fi-

nally made my breakthrough. We were planning to go to Rome soon and my wife thought I had gone crazy because I was working day and night writing up my thoughts. She said, “Why don’t you wait until we get to Rome.” I said, “By the time I get to Rome I will have forgotten all this. If I do not write it up now it will be gone.” Not only did I get the asymptotic results for small t , but I also managed to get an approximation to correct for the transition from the continuous time to the discrete time problem. My approach was to take the original discrete time normal problem and replace it with a continuous time problem to permit the use of analytic techniques. Then, after getting properties and characteristics of the solution in continuous time, to approximate by going back to the discrete time problem and doing a numerical solution. The question is: what is the relationship between the discrete time solution and the continuous time solution? I managed to get a very neat description of the relationship. In the meantime, Breakwell had tried a numerical solution and he was very pessimistic because it did not seem to be converging properly, but if you made the correction for discreteness, it fitted exactly. I still have a copy of his letter saying that the method does not seem to converge.

Bather: This question is very closely connected to the problem of allowing for the overshoot in the Wald test. It has been featured, for example, in more recent work of David Siegmund. Let us go back a little and add a footnote to your Cowles Commission experience. There are a couple of papers that we should mention: one was with Rubin.

Chernoff: Herman Rubin had obtained what I would regard as a robustness result on limited information estimates. I did not think he was the world’s best expositor, so we coauthored this paper (where I was mainly his secretary), and wrote up a version of what he had accomplished. The other paper of which I am quite proud is the paper called “Gradient methods of maximization” that I published much later with Jean Crockett. At that time we were using an iterative technique to maximize the likelihood function and it had involved an enormous amount of computation, all done with electric computers. I noticed after five or six iterations that each one was changing the estimate in the same direction as the preceding one and each succeeding step was 9/10 of the previous one. It was obvious that the convergence could be accelerated. The paper was to explain what was going on and how to accelerate the technique. Later on, Fletcher and Powell had a similar idea and did much more than I would have anticipated with it. They used the successive ap-

proximations not only to accelerate the technique, but they estimated the second derivative matrix of the function being maximized so that they could, on the basis of the first derivatives, actually imitate the Newton method and get a much quicker method of iteration. So our paper was a predecessor of the Fletcher–Powell paper.

GROWTH AT STANFORD

Bather: During your time at Stanford you had a number of colleagues, some of whom I remember from my visit in 1964. Is there anything you would like to say about them?

Chernoff: Bowker had started the Statistics Department at Stanford and he brought in Girshick from Rand in the first place. Girshick had Arrow, who had joined the Economics Department, as a joint appointment in the Statistics Department. Rubin had come from the Cowles Commission and was there. Quinn McNemar, in psychology, also had an appointment in the department. When I arrived they had a lot of visitors, but shortly after my appointment, they appointed Lincoln Moses part-time in the medical school. We were quite a small group at first, but grew rapidly. A year after I arrived, Bowker managed to convince Charles Stein to leave Chicago and come back to California, although not at the University of California but with us. They appointed Gerald Lieberman when he finished his dissertation. Then Samuel Karlin visited and they appointed Karlin as a professor. He had been at Caltech, where statistics was not considered of much value at the time, but had become interested in game theory at Rand Corporation and wanted to join us in doing some statistics. About that time, our graduate student enrollment started expanding. From just three or four students, we started letting in about seven or eight. The basis for our expansion was essentially that we had lots of government contracts and we were collaborating with the Mathematics Department. As the years went by we kept expanding. Herb Solomon came to replace Bowker as chairman, and Bowker went on to do bigger and better things as assistant to the provost and dean of the graduate school. Solomon was very expansionist and brought in a lot of people who were not particularly interested in statistics. Scarf and Uzawa, who were mainly interested in economics, were appointed and Suppes was given a joint appointment. Suppes was marginally interested in statistics, but mainly interested in the philosophy of science. Manny Parzen was already there and Solomon also managed to get Olkin and Chung.

In the early 1960s, the department was quite large, but there were a lot of diverse interests. While I was away on sabbatical, the political infighting became very serious, which led to a split. Karlin and Chung moved over to mathematics, so we became a little more compact. Olkin had been appointed while Karlin was away on sabbatical and he came back breathing fire because he had not been consulted. That was part of the reason for the split. When the department became more compact, it seemed to be in pretty solid shape. One of the big advantages of our operation at Stanford, which I do not think the university really appreciated, was that we were a center that people were able to visit. There was a lot of activity in our department and an enormous contribution by visitors and postdocs. I think the experience was very valuable for our visitors. They came, they saw the activity, they engaged in the activity and went back revived. It is an extremely important function of a university. I think our contribution with postdocs was much more valuable than our contribution with Ph.D's.

GENERALIZING THE FREE BOUNDARY PROBLEM AND FACES

Bather: My own experience reflects this. My visit in 1964–65 had a tremendous influence on my career and it gave me a great deal of confidence which I would not otherwise have had. When I came, you were just getting interested in the generalization of the free boundary problem which arose from John Breakwell and his work for Lockheed.

Chernoff: Lockheed was very concerned about the midcourse correction problem. Being interested in sending rockets to the moon or to Mars, they were wondering how to correct the path. They could estimate the miss distance from the target but somehow their attempts on this problem were not working out very well. It seemed to me that the technology that you and I had developed would be applicable and so when Breakwell formulated the problem, I brought it to you. You immediately got some good bounds for the result and those bounds clarified the nature of the solution and showed why these people were having problems with their numerical calculations. We collaborated, combining your bounds with my asymptotic expansions, and then our numerical techniques were able to get good approximations. It worked out very well, I thought.

Bather: Going back to the bounds, I was certainly able to extend my method of producing inner and outer approximations for the boundary of the decision region, but in one particular case of the spaceship problem, the bounds rested on a con-

jecture that you pointed out. I had what I thought constituted a proof, but it was a conjecture and to this day I have not been able to produce a proof of that conjecture.

Chernoff: I think we almost came to blows about it! Here was a case where, even though it seemed obviously true, it was not clear to me how to go about proving it. I think that I would have accepted it without proof if I could see an approach that would have seemed likely to succeed in proving the conjecture. Now I think that the conjecture is meaningful and reasonable in a more general case, but false there. It is probably true in our special case, but I am not convinced yet.

Bather: This is, in fact, one of a list of well-known problems in the field of dynamic programming and decision theory where there are obvious conjectures and nobody can prove them. Another example, which is now well known, is called the bullets problem or the bomber problem.

All this work on the decision problem for the normal mean and the spaceship problem was later collected together in your SIAM monograph published in 1972. There was another topic that came up about that time and this is perhaps the topic for which you will be remembered longest. This is your work in multivariate analysis that involved the use of the human face to display data. Can you say a bit more about this.

Chernoff: I was interested in cluster analysis and pattern recognition. Cluster analysis was a field in which there were about twice as many algorithms



FIG. 3. Photo of H. Chernoff around 1973 (probably at Sequoia Hall, Stanford University), with Ingram Olkin in the background.

as investigators. I was interested in making some sense of it and maybe in using some of my own algorithms. It seemed to me that the difficulty was that there is no general solution to the cluster analysis problem: the appropriate way to analyze the cluster problem depends upon the nature of the data. It would be very important to be able to look at the data before you decided on what the appropriate form of analysis is, but the difficulty was in dealing with multidimensional data when my drawing ability is limited to two dimensions. I was frustrated by how to visualize points in n -dimensional space and it suddenly occurred to me that the idea of drawing a face would in some way resolve that problem. David Hinkley was visiting that year. His wife, Betty, was an ace programmer and I talked her into writing the program to draw this simple face. It took her a couple of days to do it, and lo and behold I was able to run off these faces. In a way it saved my life because during the student revolutions at Stanford they discovered that I had an ONR contract and wanted to find out what terrible research I was doing for the Navy. They sent over a young lady to interview me to see what I was doing. I said, "Well, basically they fund me to do all kinds of research. For example, in response to my needs, I developed this method of faces." "Oh," she said, "that explains it—I work at the computer center and see those faces coming from the plotter and I was wondering what they were all about." So, the students felt that my work was not sufficiently related to the war effort to brand me as a scoundrel, but I hoped that I was helping the Navy.

Bather: I must admit that when I first saw this work on faces, I thought you were joking. In fact you were making use of a very powerful piece of intellectual machinery, namely, that human beings are extremely good at recognizing faces.

Chernoff: In a way, I was having revenge on the artificial intelligence people. I had visited M.I.T. on my sabbatical and spoken to the people interested in pattern recognition. Their attitude seemed to me to be directed to letting the computer analyze data and here I was turning the tables on them. I was the human analyzer. The computer was merely a tool to draw faces.

Bather: It is a very nice idea. You deserve the fame and notoriety that you achieved from that paper. Towards the end of your time at Stanford, your student, John Petkau, was involved with you in the computational problems arising from free boundaries.

Chernoff: That came out of his dissertation. His dissertation was involved in one of the extensions of the early work I had done on free boundary prob-

lems. It was a little more complicated, comparing two computational approaches, and as a result we had to do a numerical analysis. In the past I had done the backwards induction in discrete time using normal distributions in the computations, and it was pretty clumsy. Since his problem involved Bernoulli data, going to continuous time meant transforming from Bernoulli data to a Wiener process, and the correction for going from discrete to continuous time in that case is different. At that point it struck me that we could imitate the Brownian motion just as well by a sum of binomials in doing the numerical calculation for our backwards induction. Because the correction that we used for that case was different from the correction in the normal case, that led us to writing an interesting paper which has been much overlooked: it had to do with the solution of an optimization problem which tells us what the correction is.

Bather: This was a special problem arising from the study of local conditions near the boundary.

Chernoff: The problem that we were interested in related to the solution of a certain optimization problem with a parameter p . Corresponding to each value of the parameter p there is a certain number which is very important for our correction. For the Petkau problem, the number is $1/2$, but for other values of p it is different. Unfortunately, our technique for finding this function worked only for rational values: we solved it for $p = 1/2$, then $p = 1/3$ and $2/3$, then $1/4$, $3/4$, $1/5$, $2/5$ etcetera. As we piled up these points, the function looked extremely discontinuous, so I was in a state of shock, because if our correction was a discontinuous function of p , that would cast a doubt on the value of our correction. So it became very important to check out the continuity. It turned out that the function is continuous but has a cusp at every rational number: at $p = 1/2$, it has the worst possible cusp. So we needed to prove the continuity of our correction. Eventually, one of David Siegmund's students, Hogan, developed a more general view of this problem and he was able to extend our results. At that time, we went through a stage of some uncertainty before we proved the continuity.

Bather: Perhaps the difficulty was because you were very close to optimality, in that region of the boundary where the two functions have to match because of the smooth fit condition.

Chernoff: I am not sure.

Bather: Let me put it another way. What is the relation between this local study of the boundary we were talking about and the extension of the no-overshoot approximation by allowing for the distribution of the overshoot, which came later?

Chernoff: The way I had originally handled it gave rise to an integral equation and originally, in the normal case, that integral equation had been treated by Spitzer.

I forget now how it fits in, but my derivations depended on the relationship between the discrete time problem where we are not allowed to stop in between certain discrete times and the continuous time problem. I was not thinking in terms of overshoot. My derivation was based on the fact that, near the boundary, the continuous and discrete time problems very much resemble a special relatively simple problem that I called the canonical problem.

THE MOVE TO M.I.T.

Bather: All this work was occurring round about the time of your transition across the states from Stanford to M.I.T. Would you like to say a bit more about that and your reasons for moving back east?

Chernoff: There were two reasons. One, there was a lot of politics going on at Stanford University and I did not care too much for the politicians at the university at that time. I found, as chairman of the department for one year, that dealing with the dean and the associate dean in particular was a trial and tribulation. Also, it became clear to me that, at that time, theoretical statistics was spinning its wheels. Theoreticians were going into more and more elaborate generalizations. We had reached a period where we had to confront the computer much more intensively and we also had to do much more applied work. The Statistics Department at Stanford was very healthy, in the sense that half of our appointees were joint appointments with other departments. However, from the point of view of most of our students, they saw very little of the applied work that was going on. I thought, for the future, the field needed a lot more contact with real applications in order to provide insights into which way we should go, rather than concentrating on further elaborations on theory. I found that most of my colleagues were not terribly interested in changing their way of operation. I felt that M.I.T. would be a wonderful opportunity, being a school of technology, to develop a statistics program with a strong emphasis on applications. After I left Stanford, the statistics department became much more applied and at M.I.T. everybody thought they could do their own statistics, so it was difficult for me to develop a large enough group to be self-sufficient.

Bather: There wasn't a department at M.I.T.?

Chernoff: At M.I.T., I was a subset of the applied mathematics subset of the mathematics department. The other applied mathematicians felt

that, as only 1 out of 12 senior professors, I was demanding more than one-twelfth of the assets of the group in order to get my program going. I can sympathize with their feelings that I was being too aggressive, but on the other hand the thought of having a statistics program with only two professors struck me as absurd, because it was necessary for students to learn a lot more than one or two professors could teach them. I could not appoint a senior professor without being able to find someone outstanding, and outstanding people did not want to come to M.I.T., where they would be part of a Mathematics Department.

Bather: You did succeed in one of your ambitions and that was moving towards applications, because you set up the Statistics Center in Cambridge.

Chernoff: My history in statistics, surprisingly enough, was relatively applied, in the sense that a lot of my ideas came out of the work I did on a contract funded by ONR, much of which involved applications. I don't think I am very good as an applied statistician and so I find it difficult to find standard approaches to use in such problems. On the other hand, when you look at these problems honestly, it often turns out that novel theoretical issues get raised. That was my *modus operandi* during my period at Stanford and certainly since then. That explains a little bit why, except for the experimental design and sequential analysis, my work tends to scatter in all sorts of directions.

Bather: There is a more positive way of describing this. It indicates a man with the confidence to use his ability in many directions, being not so afraid that he has to specialize on one. Maybe it is also a generation thing. It always seems remarkable to me that, through the generations in this century, we have moved away from scientists who knew a lot of science, including mathematics, people like the 19th-century physicists. Now even very clever people can only cope with one branch of statistics.

Chernoff: These fields became larger and, necessarily, more specialized. I remember a conversation I had with Harold Hotelling in 1960, who mentioned that, on his honeymoon, he spent a year going around the world and that was the year he fell behind. Until then, he knew all of statistics. When statistics was a smaller field, it was easier to know a great deal more and to take part in a wider range of activities than it is now.

MEDICAL TRIALS

Bather: We have talked about a wide variety of problems, mostly theoretical, but let me remind you that one of the main motivations for studying se-

quential testing was the medical one, because of the ethical considerations. You did some work with John Petkau in the 1970s, concerned with sequential medical trials.

Chernoff: What happened was that my first involvement was due to S. N. Ray, who had raised the subject of sequential rectified sampling. He had done his dissertation, I think, with Jack Hall at North Carolina and came as a research associate to Stanford. He raised this problem and we applied the techniques that had been used for sequential sampling. When it was finished, it was clear to me that it could be regarded as a one-armed bandit problem. You have a fixed horizon of observations and you keep pulling the arm as long as you think you are making a profit. When you decide you are not making a profit, you just stop pulling the arm, but you only have a finite amount of time in which to pull the arm if it is profitable. Shortly after, there was a Berkeley Symposium and I was invited to give a talk. I applied that solution to the idea of having clinical trials, but in a somewhat unrealistic fashion because I did not consider controls.

This version of the problem had some relevance to the ethical aspects of the problem, because the major contribution to the costs was the number of successes and failures during the horizon of plausible treatment rather than the cost of taking observations. In the meantime, Frank Anscombe and Ted Colton, who I think was Anscombe's student, had stated more realistic versions of these problems from the medical point of view. Colton had done a minimax version of the problem, but I think not sequentially. Anscombe had proposed a sequential version which later turned out to be very efficient.

Bather: That was a remarkably inspired guess.

Chernoff: It was more than a guess; he had done some calculations. In the meantime, Petkau had followed me to M.I.T. and started working on a dissertation which was along these lines. Later on, we adapted the Anscombe formulation and tried to get the optimal procedure and actually computed the optimal procedure from this Bayesian point of view. It turned out, as David Siegmund had suggested, that Anscombe's procedures were remarkably efficient. Siegmund somehow has always ignored the approach we developed because he thinks that it is too complicated, and I have tended to avoid Siegmund's formulation because it tends to be non-Bayesian. Although I regard myself as non-Bayesian, I feel in sequential problems it is rather dangerous to play around with non-Bayesian procedures. Siegmund is very capable of determining when and where he is in danger so does not run a risk but, in the hands of less skilled people, non-

Bayesian procedures are very tricky in sequential work.

Bather: In my view, the approach taken by Siegmund and Michael Woodroofe has been largely concerned with evaluating and getting good approximations to the performance characteristics of established tests without really taking very seriously the question of optimality. Optimality is, of course, implicit in the Bayesian approach.

Chernoff: I don't think that is quite accurate. Siegmund and Woodroofe do worry about optimality, in fact sometimes higher-order optimality. I tend to work from a rather rough point of view; I like to get first-order optimality without worrying about higher orders. But it is true that they like to use certain tests. For example, Siegmund has done repeated significance tests. One of the nice things about the one-armed bandit problem is its easy interpretation of the significance level at which you should stop, as a function of the proportion of the information accumulated compared to the total amount of information you will eventually have if you continue.

Bather: On that topic, there has been a remarkable change of attitudes in the medical field in recent years. For the first 20 to 30 years after Wald, very little of the theory found its way into practice, but there are signs now that people are actually using sequential procedures. They may not be exactly the procedures that you proposed, but I am sure the influence of your ideas has been quite strong.

Chernoff: I think the ethical issue has made it seem more important to them to consider stopping trials only when they feel they have enough information.

THE MASSACHUSETTS NUMBER GAME

Bather: One of the most attractive titles among your later publications had to do with the Massachusetts Numbers Game. You produced a paper in 1981, while you were at M.I.T., that was concerned with how to beat the system.

Chernoff: While I was at Boston they had started the Massachusetts lottery and one of the key games was called the numbers game, in which they had four-digits. A public relations (PR) man had an interview with a newspaper. The newspaper man asked if any of the four digit numbers had ever been repeated in the first 500 trials of this game. The PR man said "No." After all there were 10,000 possibilities and only 500 games had been played. A graduate student did some calculations and found that, in this variation on the birthday problem there almost certainly should have been a

duplication. So I figured that if there had not been, it was obviously fixed. I asked Harvey Greenspan, the chairman of the M.I.T. applied math group, what I should do about it. On the one hand, I wanted to be public spirited and announce that the game was fixed. On the other hand, I figured if the game was fixed, it was because the Mafia was involved and they might not like it if I complained. His suggestion was, if it was fixed, I should try and find out how to take advantage of it. So I sent away for the publicly available information and, in the meantime, some foolhardy student wrote a letter to the editor stating that the event of no duplications was very unlikely. The PR man replied that he had just assumed that there had been no duplications and, looking over the past information, there had been seven, which was more or less reasonable.

At the time, I was interested in pattern recognition and I thought that this would be a good way of finding out the most favorable numbers to bet on. So I did a study but when I presented the paper at Georgia Tech I got a call from a newspaper man who was interested, so I sent him a copy. He wrote an article which appeared in *Esquire*, which started out as follows: "I do not know whether it is a good idea for the National Science Foundation to fund Professor Chernoff, whose main object in life seems to be to beat the Massachusetts lottery." What frightened me was that this article appeared immediately below an article criticizing Senator Proxmire. He would certainly see the article and he was famous for giving out "Golden Goose" awards.

Bather: So you were afraid that he may set his investigators on you?

Chernoff: Absolutely. In fact, I had a lot of newspaper men calling me up. The National Science Foundation called me up and wanted to know what was going on. I said, "It came out as a technical report and I sent it to you." They said, "Well, you know what we do with these technical reports, we put them in the pile. Send another copy." The reason I wrote the report was that it was an excellent way of showing why some of our graduate students had lost money on this game. They had failed to understand a couple of issues which were known to statisticians and unknown to the rest of the community, and it was a good example to illustrate these principles to people who are not statistically sophisticated. These were the gamblers ruin problem and regression to the mean.

Bather: No doubt the proportion taken off the top by the state of Massachusetts had an effect on their chance of making a profit.

Chernoff: That's right. The state took 40% and it turned out that the chances of making a profit

were nontrivial at the very beginning but became essentially negligible shortly afterwards.

Bather: You might be interested to know that Great Britain has just introduced a similar national lottery. This indicates to me that our economy is so bad, we need to imitate the state of Massachusetts.

Chernoff: Well I hope that they allocate some of the profit from the lottery to treat people who become addicted to gambling. One of the problems of the Massachusetts lottery is the attractive advertising that they do. Fortunately, they have now decided to cut the budget for this advertising.

MOVING TO HARVARD

Bather: On a more serious note, about this time, you decided to move from M.I.T. to Harvard. Was this because you had struggled long and hard to set up a viable statistics department or a statistics group at M.I.T. and failed?

Chernoff: Basically, yes. While I was at M.I.T., I managed to attract Bill Du Mouchel to come and join me. We had some rather good people; postdocs came with the title of Instructor or sometimes Assistant Professors spent a few years with us and those years I thought were very pleasant. I particularly enjoyed working side-by-side with Du Mouchel, though we didn't collaborate. I found that he was absolutely brilliant, doing black magic on problems which I had found impossible. I enjoyed the M.I.T. environment, but as a place to set up a program for training doctoral students in statistics, it was seriously limited. We could only hope to get a few students who were brilliant enough to learn everything they needed very much by themselves because we would not have the capacity to do anything more



FIG. 4. Photo of H. Chernoff around 1985, about which he says: "People say this is a good picture. I don't like it. They say that it looks just like me. That's why I don't like it."

than the most primitive sort of teaching. At one point, I wrote a letter which my wife said was suicidal. I wrote to the president of M.I.T. telling him that he should look at Purdue University to see how a statistics program should be carried out in a technical university. My wife explained to me that you do not tell the president of M.I.T. to follow in the footsteps of Purdue University. It was clear that the program would not go anywhere, at which point Harvard was aware of my difficulties and invited me to join them. I must admit that Harvard also has a problem. In my opinion it is much too small a department to maintain the robustness necessary to survive. If a couple of serious problems came up, the department would be in a very difficult shape.

Bather: You have now been at Harvard 10 years. Some might think that this is a state of Nirvana for a statistician and the only thing you could now do is to run for president. I take it you have no intention of doing that.

Chernoff: No, it is reminiscent of the people in the army who felt that they could be promoted to become civilian. I guess in three years I will be promoted to professor emeritus at Harvard. I have made plans to retire in 1997 on my 74th birthday, at which point Harvard may provide a closet in which I can continue my work. I have enjoyed my years at Harvard and we have a variety of students who are most successful in spite of the limited faculty. I should say that at M.I.T. I tried to build up a program which would be very applied oriented and, in particular, we had quite a few masters students until the depression hit in 1980. The cost of paying tuition to get a masters degree became exorbitant and we lost our masters degree program. One aspect of that program of which I was very proud was that they had to write a dissertation. Originally, I felt that this dissertation could be a short report on an applied topic, but it turned out that the students there were not satisfied with writing a short report and wrote real theses, many of which were quite worthwhile. Even those students who I do not think were qualified to go on to do a Ph.D did excellent theses, and most of these were applied to real problems that came up in the medical or engineering areas. At Harvard we do not require a thesis for a masters degree, and my feelings are that the students lose a very important part of their training that is necessary to give them a feeling of self-sufficiency and an ability to operate on their own.

One of the innovations that we developed in the Harvard program was that Don Rubin introduced the idea that the students, who had been spinning their wheels when I first arrived at Harvard,

should be forced after they had passed their qualifying exam, to report orally on their progress at the end of each semester. I think that this program of having all the students present their material at the end of each semester has been a very positive influence. On the one hand it becomes evident when they are not making progress, whereas on the other hand they learn to present material reasonably well. Another aspect of my experience at Harvard was that the teaching assistants there have been superb. There have been very few teaching assistants that I have had in other schools who could compare with the vast majority of the assistants that I had at Harvard and I never understood why. In fact, one of my first teaching assistants, Schafer, was uniquely gifted, so much so that once, when the students evaluated us, one of the students wrote: "Professor Chernoff is extremely disorganized. I wish he would get out of the way and let the TA's handle the class." It might have been that the others tried to imitate him and maybe Harvard University makes a special effort to get teaching assistants to do a good job. I have always been very pleased with their performance.

Bather: Going back to your plans to retire in 1997, this will give you the advantage that they will

not ask you to chair the department again, which you did quite recently, I understand.

Chernoff: When I arrived at Harvard, Don Rubin had come one year earlier and he became chairman. He served nine years. Carl Morris, who is part-time in the medical school, will be chairman next year. One year when Rubin was on leave, I was acting chairman. For a department of our size it wasn't exactly an onerous job, but I would rather not do it.

Bather: You are still extremely active, more active than many younger professors, and I very much hope that you will continue even after you retire.

Chernoff: A few years ago I developed an interest in molecular biology. I had heard about the human genome project, which will have accumulated an enormous amount of data. If statisticians cannot do something with this, what good are they? So now I have an opportunity to return some of the unkindnesses that others visit upon statisticians. Since everybody in the world thinks he can teach statistics even though he does not know any, I shall put myself in the position of teaching biology even though I do not know any.

Bather: That is a nice thought to end our discussion. I wish you well for the future.

Chernoff: Thank you.