

A Conversation with David Blackwell

Morris H. DeGroot

David Blackwell was born on April 24, 1919, in Centralia, Illinois. He entered the University of Illinois in 1935, and received his A.B. in 1938, his A.M. in 1939, and his Ph.D. in 1941, all in mathematics. He was a member of the faculty at Howard University from 1944 to 1954, and has been a Professor of Statistics at the University of California, Berkeley, since that time. He was President of the Institute of Mathematical Statistics in 1955. He has also been Vice President of the American Statistical Association, the International Statistical Institute, and the American Mathematical Society, and President of the Bernoulli Society. He is an Honorary Fellow of the Royal Statistical Society and was awarded the von Neumann Theory Prize by the Operations Research Society of America and the Institute of Management Sciences in 1979. He has received honorary degrees from the University of Illinois, Michigan State University, Southern Illinois University, and Carnegie-Mellon University.

The following conversation took place in his office at Berkeley one morning in October 1984.

"I EXPECTED TO BE AN ELEMENTARY SCHOOL TEACHER"

DeGroot: How did you originally get interested in statistics and probability?

Blackwell: I think I have been interested in the concept of probability ever since I was an undergraduate at Illinois, although there wasn't very much probability or statistics around. Doob was there but he didn't teach probability. All the probability and statistics were taught by a very nice old gentleman named Crathorne. You probably never heard of him. But he was a very good friend of Henry Rietz and, in fact, they collaborated on a college algebra book. I think I took all the courses that Crathorne taught: two undergraduate courses and one first-year graduate course. Anyway, I have been interested in the subject for a long time, but after I got my Ph.D. I didn't expect to get professionally interested in statistics.

DeGroot: But did you always intend to go on to graduate school?

Blackwell: No. When I started out in college I expected to be an elementary school teacher. But somehow I kept postponing taking those education courses. [Laughs] So I ended up getting a master's degree and then I got a fellowship to continue my work there at Illinois.

DeGroot: So your graduate work wasn't particularly in the area of statistics or probability?

Blackwell: No, except of course that I wrote my thesis under Doob in probability.

DeGroot: What was the subject of your thesis?

Blackwell: Markov chains. There wasn't very much original in it. There was one beautiful idea, which was Doob's idea and which he gave to me. The thesis was never published as such.

DeGroot: But your first couple of papers pertained to Markov chains.

Blackwell: The first couple of papers came out of my thesis, that's right.

DeGroot: So after you got your degree . . .

Blackwell: After I got my degree, I sort of expected to work in probability, real variables, measure theory, and such things.

DeGroot: And you *have* done a good deal of that.

Blackwell: Yes, a fair amount. But it was Abe Girshick who got me interested in statistics.

DeGroot: In Washington?

Blackwell: Yes. I was teaching at Howard and the mathematics environment was not really very stimulating, so I had to look around beyond the university just for whatever was going on in Washington that was interesting mathematically.

DeGroot: Not just statistically, but mathematically?

Blackwell: I was just looking for anything interesting in mathematics that was going on in Washington.

DeGroot: About what year would this be?

Blackwell: I went to Howard in 1944. So this would have been during the year 1944-1945.

DeGroot: Girshick was at the Department of Agriculture?

Blackwell: That's right. And I heard him give a lecture sponsored by the Washington Chapter of the American Statistical Association. That's a pretty lively chapter. I first met George Dantzig when he gave a lecture there around that same time. His lecture had nothing to do with linear programming, by the way. In fact, I first became acquainted with the idea of a randomized test by hearing Dantzig talk about it. I think that he was the guy who invented a test function, instead of having just a rejection region that is a subset of the sample space. At one of those meetings Abe Girshick spoke on sequential analysis. Among other things, he mentioned Wald's equation.

DeGroot: That's the equation that the expecta-

tion of a sum of random variables is $E(N)$ times the expectation of an individual variable?

Blackwell: Yes. That was just such a remarkable equation that I didn't believe it. So I went home and thought I had constructed a counterexample. I mailed it to Abe, and I'm sure that he discovered the error. But he didn't write back and tell me it was an error; he just called me up and said let's talk about it. So we met for lunch and that was the start of a long and beautiful association that I had with him.

DeGroot: Would you regard the Blackwell and Girshick book (*Theory of Games and Statistical Decisions*. New York, John Wiley & Sons, 1954) as the culmination of that association?

Blackwell: Oh, that was a natural outgrowth of the association. I learned a great deal from him.

DeGroot: Were you together at any time at Stanford?

Blackwell: Yes, I spent a year at Stanford. I think it was 1950–1951. But he and I were also together at other times. We spent several months together at Rand. So we worked together in Washington, and then at Rand, and then at Stanford.

“I WROTE 105 LETTERS OF APPLICATION”

DeGroot: Tell me a little about the years between your Ph.D. from Illinois in 1941 and your arrival at Howard in 1944. You were at a few other schools in between.

Blackwell: Yes. I spent my first postdoctoral year at the Institute for Advanced Study. Again, I continued to show my interest in statistics. I sat in on Sam Wilks' course in Princeton during that year. Henry Scheffé was also sitting in on that class. He had just completed his Ph.D. at Wisconsin. Jimmie Savage was at the Institute for that year. He was at some of Wilks' lectures, too. There were a lot of statisticians about our age around Princeton at that time. Alex Mood was there. George Brown was there. Ted Anderson was there. He was in Wilks' class that year.

DeGroot: He was a graduate student?

Blackwell: He was a graduate student, just completing his Ph.D. So that was my first postdoctoral year. Also, I had a chance to meet von Neumann that year. He was a most impressive man. Of course, everybody knows that. Let me tell you a little story about him.

When I first went to the Institute, he greeted me, and we were talking, and he invited me to come around and tell him about my thesis. Well, of course, I thought that was just his way of making a new young visitor feel at home, and I had no intention of telling him about my thesis. He was a big, busy, important man. But then a couple of months later, I saw him at tea

and he said, “When are you coming around to tell me about your thesis? Go in and make an appointment with my secretary.” So I did, and later I went in and started telling him about my thesis. He listened for about ten minutes and asked me a couple of questions, and then he started telling *me* about *my* thesis. What you have really done is this, and probably this is true, and you could have done it in a somewhat simpler way, and so on. He was a really remarkable man. He listened to me talk about this rather obscure subject and in ten minutes he knew more about it than I did. He was extremely quick. I think he may have wasted a certain amount of time, by the way, because he was so willing to listen to second- or third-rate people and think about their problems. I saw him do that on many occasions.

DeGroot: So, from the Institute you went where?

Blackwell: I went to Southern University in Baton Rouge, Louisiana. That's a state school and at that time it was *the* state university in Louisiana for blacks. I stayed there just one year. Then the next year, I went to Clark College in Atlanta, also a black school. I stayed there for one year. Then I went to Howard University in Washington and stayed there for ten years.

DeGroot: Was Howard at a different level intellectually from these other schools?

Blackwell: Oh yes. It was the ambition of every black scholar in those days to get a job at Howard University. That was the best job you could hope for.

DeGroot: How large was the math department there in terms of faculty?

Blackwell: Let's see. There were just four regular people in the math department. Two professors. I went there as an assistant professor. And there was one instructor. That was it.

DeGroot: Have you maintained any contact with Howard through the years?

Blackwell: Oh yes. I guess the last time I gave a lecture there was about three years ago, but I visited many times during the years.

DeGroot: Do you see much change in the place through the years?

Blackwell: Yes, the math department now is a livelier place than it was when I was there. It's much bigger and the current chairman, Jim Donaldson, is very good and very active. There are some interesting things going on there.

DeGroot: Did you feel or find that discrimination against blacks affected your education or your career after your Ph.D.?

Blackwell: It never bothered me. I'll put it that way. It surely shaped my expectations from the very beginning. It never occurred to me to think about teaching in a major university since it wasn't in my horizon at all.



David Blackwell (lower left), 1930, probably sixth grade.

DeGroot: Even in your graduate-student days at Illinois?

Blackwell: That's right. I just assumed that I would get a job teaching in one of the black colleges. There were 105 black colleges at that time, and I wrote 105 letters of application.

DeGroot: And got 105 offers, I suppose.

Blackwell: No, I eventually got three offers, but I accepted the first one that I got. From Southern University.

DeGroot: Let's move a little further back in time. You grew up in Illinois?

Blackwell: In Centralia, Illinois. Did you ever get down to Centralia or that part of Illinois when you were in Chicago?

DeGroot: No, I didn't.

Blackwell: Well, it's a rather different part of the world from northern Illinois. It's quite southern. Centralia in fact was right on the border line of segregation. If you went south of Centralia to the southern tip of Illinois, the schools were completely segregated in those days. Centralia had one completely black school, one completely white school, and five "mixed" schools.

DeGroot: Well, that sounds like the boundary all right. Which one did you go to?

Blackwell: I went to one of the mixed schools, because of the part of town I lived in. It's a small town. The population was about 12,000 then and it's still about 12,000. The high school had about 1,000 students. I had very good high school teachers in mathematics. One of my high school teachers organized a mathematics club and used to give us problems

to work. Whenever we would come up with something that had the idea for a solution, he would write up the solution for us, and send it in our name to a journal called *School Science and Mathematics*. It was a great thrill to see your name in the magazine. I think my name got in there three times. And once my *solution* got printed. As I say, it was really Mr. Huck's write-up based on my idea. [Laughs].

DeGroot: Was your family encouraging about your education?

Blackwell: It was just sort of *assumed* that I would go to college. There was no "Now be sure to study hard" or anything like that. It was just taken for granted that I was going to go to college. They were very, very supportive.

SOME FAVORITE PAPERS

DeGroot: You were quite young when you received your Ph.D. You were 21 or so?

Blackwell: 22. There wasn't any big jump. I just sort of did everything a little faster than normal.

DeGroot: And you've been doing it that way ever since. You've published about 80 papers since that time. Do you have any favorites in that list that you particularly like or that you feel were particularly important or influential?

Blackwell: Oh, I'm sure that I do, but I'd have to look at the list and think about that. May I look?

DeGroot: Sure. This is an open-book exam.

Blackwell: Good. Let's see . . . Well, my first *statistical* paper, called "On an equation of Wald" (*Ann. Math. Statist.* 17 84-87, 1946) grew out of that origi-

nal conversation with Abe Girshick. That's a paper that I am still really very proud of. It just gives me pleasant feelings every time I think about it.

DeGroot: Remind me what the main idea was.

Blackwell: For one thing it was a proof of Wald's theorem under, I think, weaker conditions than it had been proved before, under sort of *natural* conditions. And the proof is *neat*. Let me show it to you. [Goes to blackboard.]

Suppose that X_1, X_2, \dots are i.i.d. and you have a stopping rule N , which is a random variable. You want to prove that $E(X_1 + \dots + X_N) = E(X_1)E(N)$. Well, here's my idea. Do it over and over again. So you have stopping times N_1, N_2, \dots , and you get

$$S_1 = X_1 + \dots + X_{N_1},$$

$$S_2 = X_{N_1+1} + \dots + X_{N_1+N_2},$$

...

Consider $S_1 + \dots + S_k = X_1 + \dots + X_{N_1+\dots+N_k}$. We can write this equation as

$$\frac{S_1 + \dots + S_k}{k} = \left(\frac{X_1 + \dots + X_{N_1+\dots+N_k}}{N_1 + \dots + N_k} \right) \left(\frac{N_1 + \dots + N_k}{k} \right).$$

Now let $k \rightarrow \infty$. The first term on the right is a subsequence of the X averages. By the strong law of large numbers, this converges to $E(X_1)$. The second term on the right is the average of N_1, \dots, N_k . We are assuming that they have a finite expectation, so this converges to that expectation $E(N)$. Therefore, the sequence

$$\frac{S_1 + \dots + S_k}{k}$$

converges a.e. Then the converse of the strong law of large numbers says that the expected value of each S_i must be finite, and that

$$\frac{S_1 + \dots + S_k}{k}$$

must converge to that expectation $E(S_1)$. Isn't that neat?

DeGroot: Beautiful, beautiful.

Blackwell: So that's the proof of Wald's equations just by invoking the strong law of large numbers and its converse. I think I like that because that was the first time that I decided that I could do something original. The papers based on my thesis were nice, but those were really Doob's ideas that I was just carrying out. But here I had a really original idea, so I was very pleased with that paper. Then I guess I like my paper with Ken Arrow and Abe Girshick, "Bayes and mini-

max solutions of sequential decision problems" (*Econometrica* 17 213-244, 1949).

DeGroot: That was certainly a very influential paper.

Blackwell: That was a serious paper, yes.

DeGroot: There was some controversy about that paper, wasn't there? Wald and Wolfowitz were doing similar things at more or less the same time.

Blackwell: Yes, they had priority. There was no question about that, and I think we did give inadequate acknowledgment to them in our work. So they were very much disturbed about it, especially Wolfowitz. In fact, Wolfowitz was cool to me for more than 20 years.

DeGroot: But certainly your paper was different from theirs.

Blackwell: We had things that they didn't have, there was no doubt about that. For instance, induction backward—calculation backward—that was in our paper and I don't think there is any hint of it in their work. We did go beyond what they had done. Our paper didn't seem to bother Wald too much, but Wolfowitz was annoyed.

DeGroot: Did you know Wald very well or have much contact with him?

Blackwell: Not very well. I had just three or four conversations with him.

IMPORTANT INFLUENCES

DeGroot: I gather from what you said that Girshick was a primary influence on you in the field of statistics.

Blackwell: Oh yes.

DeGroot: Were there other people that you felt had a strong influence on you? Neyman, for example?

Blackwell: Not in my statistical thinking. Girshick was certainly *the* most important influence on me. The other person who had just one influence, but it was a very big one, was Jimmie Savage.

DeGroot: What was that one influence?

Blackwell: Well, he explained to me that the Bayes approach was the right way to do statistical inference. Let me tell you how that happened. I was at Rand, and an economist came in one day to talk to me. He said that he had a problem. They were preparing a recommendation to the Air Force on how to divide their research budget over the next five years and, in particular, they had to decide what fraction of it should be devoted to long-range research and what fraction of it should be devoted to more immediate developmental research.

"Now," he said, "one of the things that this depends on is the probability of a major war in the next five years. If it's large then, of course, that would shift the emphasis toward developing what we already know how to do, and if it's small then there would be more emphasis on long-range research. I'm not going to ask



David Blackwell, about 1945.

you to tell me a number, but if you could give me any guide as to how I could go about finding such a number I would be grateful." Oh, I said to him, that question just doesn't make sense. Probability applies to a long sequence of repeatable events, and this is clearly a unique situation. The probability is either 0 or 1, but we won't know for five years, I pontificated. [Laughs] So the economist looked at me and nodded and said, "I was afraid you were going to say that. I have spoken to several other statisticians and they have all told me the same thing. Thank you very much." And he left.

Well, that conversation bothered me. The fellow had asked me a reasonable, serious question and I had given him a frivolous, sort of flip, answer, and I wasn't happy. A couple of weeks later Jimmie Savage came to visit Rand, and I went in and said hello to him. I happened to mention this conversation that I had had, and then he started telling me about deFinetti and personal probability. Anyway, I walked out of his office half an hour later with a completely different view on things. I now understood what was the right way to do statistical inference.

DeGroot: What year was that?

Blackwell: About 1950, maybe 1951, somewhere

around there. Looking back on it, I can see that I was emotionally and intellectually prepared for Jimmie's message because I had been thinking in a Bayesian way about sequential analysis, hypothesis testing, and other statistical problems for some years.

DeGroot: What do you mean by thinking in a Bayesian way? In terms of prior distributions?

Blackwell: Yes.

DeGroot: Wald used them as a mathematical device.

Blackwell: That's right. It just turned out to be clearly a very natural way to think about problems and it was mathematically beautiful. I simply regretted that it didn't correspond with reality. [Laughs] But then what Jimmie was telling me was that the way that I had been thinking all the time was really the right way to think, and not to worry so much about empirical frequencies. Anyway, as I say, that was just one very big influence on me.

DeGroot: Would you say that your statistical work has mainly used the Bayesian approach since that time?

Blackwell: Yes. I simply have not worked on problems where that approach could not be used. For instance, all my work in dynamic programming just has that Bayes approach in it. That is *the* standard way of doing dynamic programming.

DeGroot: You wrote a beautiful book called *Basic Statistics* (New York, McGraw-Hill, 1970) that was really based on the Bayesian approach, but as I recall you never once mentioned the word "Bayes" in that book. Was that intentional?

Blackwell: No, it was not intentional.

DeGroot: Was it that the terminology was irrelevant to the concepts that you were trying to get across?

Blackwell: I doubt if the word "theorem" was ever mentioned in that book. That was not originally intended as a book, by the way. It was simply intended as a set of notes to give my students in connection with lectures in this elementary statistics course. But the students suggested that it should be published and a McGraw-Hill man said that he would be interested. It's just a set of notes. It's short; I think it's less than 150 pages.

DeGroot: It's beautiful. There are a lot of wonderful gems in those 150 pages.

Blackwell: Well, I enjoyed teaching the course.

DeGroot: Do you enjoy teaching from your own books?

Blackwell: No, not after a while. I think about five years after the book was published, I stopped using it. Just because I got bored with it. When you reach the point where *you're* not learning anything, then I think it's probably time to change something.

DeGroot: Are you working on other books at the present time?

Blackwell: No, except that I am *thinking about* writing a more elementary version of parts of your book on optimal statistical decisions because I have been using it in a course and the undergraduate students say that it's too hard.

DeGroot: Uh oh. I've been thinking of doing the same thing. [Laughs] Well, I am just thinking generally in terms of an introduction to Bayesian statistics for undergraduates.

Blackwell: Very good. I really hope you do it, Morrie. It's needed.

DeGroot: Well, I really hope you do it, too. It would be interesting. Are there courses that you particularly enjoy teaching?

Blackwell: I like the course in Bayesian statistics using your book. I like to teach game theory. I haven't taught it in some years, but I like to teach that course. I also like to teach, and I'm teaching right now, a course in information theory.

DeGroot: Are you using a text?

Blackwell: I'm not using any one book. Pat Billingsley's book *Ergodic Theory and Information* comes closest to what I'm doing. I like to teach measure theory. I regard measure theory as a kind of hobby, because to do probability and statistics you don't really need very much measure theory. But there are these fine, nit-picking points that most people ignore, and rightly so, but that I sort of like to worry about. [Laughs] I know that it is not important, but it is interesting to me to worry about regular conditional probabilities and such things. I think I'm one of only three people in our department who really takes measure theory seriously. Lester [Dubins] takes it fairly seriously, and so does Jim Pitman. But the rest of the people just sort of ignore it. [Laughs]

"I WOULD LIKE TO SEE MORE EMPHASIS ON BAYESIAN STATISTICS"

DeGroot: Let's talk a little bit about the current state of statistics. What areas do you think are particularly important these days? Where do you see the field going?

Blackwell: I can tell you what I'd like to see happen. First, of course, I would like to see more emphasis on Bayesian statistics. Within that area it seems to me that one promising direction which hasn't been explored at all is Bayesian experimental design. In a way, Bayesian statistics is much simpler than classical statistics in that once you're given a sample, all you have to do are calculations based on that sample. Now, of course, I say "all you have to do"—sometimes those calculations can be horrible. But if you are trying to design an experiment, that's not all you have to do. In that case, you have to look at all

the different samples you might get and evaluate every one of them in order to calculate an overall risk, to decide whether the experiment is worth doing and to choose among the experiments. Except in very special situations, such as when to stop sampling, I don't think a lot of work has been done in that area.

DeGroot: I think the reason there hasn't been very much done is because the problems are so hard. It's really hard to do explicitly the calculations that are required to find *the* optimal experiment. Do you think that perhaps the computing power that is now available would be helpful in this kind of problem?

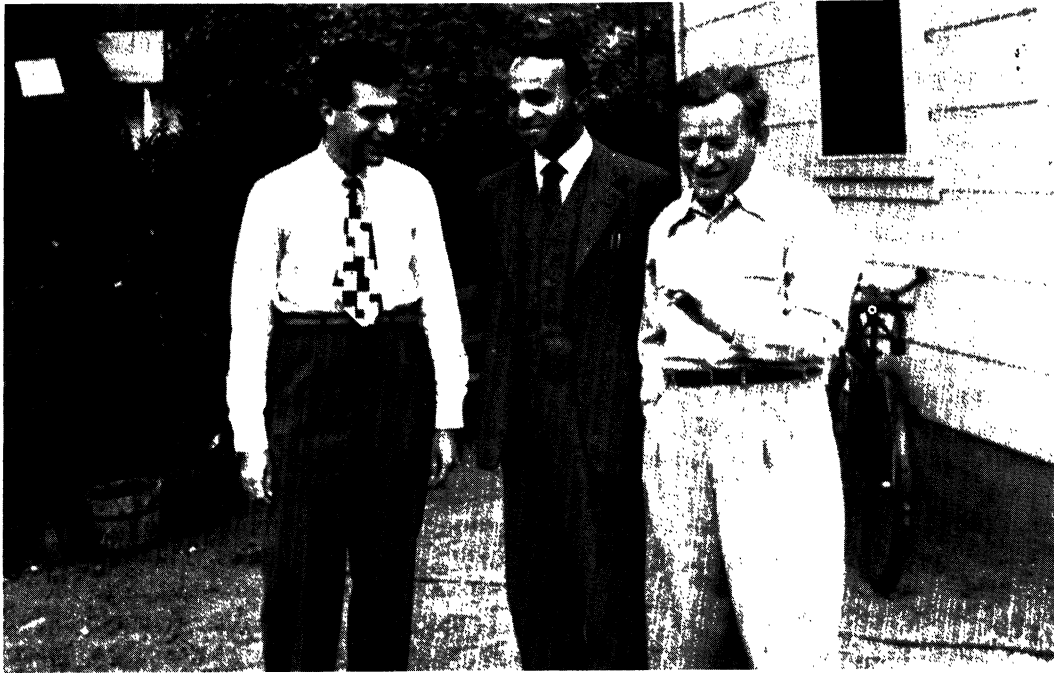
Blackwell: That's certainly going to make a difference. Let me give you a simple example that I have never seen worked out but I am sure could be worked out. Suppose that you have two independent Bernoulli variables, say, a proportion among males and a proportion among females. They are independent, and you are interested in estimating the sum of those proportions or some linear combination of those proportions. You are going to take a sample in two stages. First of all, you can ask how large should the first sample be? And then, based on the first sample, how should you allocate proportions in the second sample?

DeGroot: Are you going to draw the first sample from the total population?

Blackwell: No. You have males and you have females, and you have a total sample effort of size N . Now you can pick some number $n \leq N$ to be your sample size. And you can allocate those n observations among males and females. Then based on how that sample comes out, you can allocate your second sample. What is the best initial allocation, and how much better is it than just doing it all in one stage? Well, I haven't done that calculation but I'm sure that it can be done. It would be an interesting kind of thing and it could be extended to more than two categories. That's an example of the sort of thing on which I would like to see a lot of work done—Bayesian experimental design.

One of the things that I worry about a little is that I don't see theoretical statisticians having as much contact with people in other areas as I would like to see. I notice here at Berkeley, for example, that the people in Operations Research seem to have much closer contact with industry than the people in our department do. I think we might find more interesting problems if we did have closer contact.

DeGroot: Do you think that the distinctions between applied and theoretical statistics are still as rigid as they were years ago or do you think that the field is blending more into a unified field of statistics in which such distinctions are not particularly meaningful? I see the emphasis on data analysis which is coming about, and the development of theory for data analysis and so on, blurring these distinctions



Kenneth Arrow, David Blackwell, and M. A. Girshick, Santa Monica, September 1948.

between theoretical and applied statistics in a healthy way.

Blackwell: I guess I'm not familiar enough with data analysis and what computers have done to have any interesting comments on that. I see what some of our people and people at Stanford are doing in looking at large-dimensional data sets and rotating them so that you can see lots of three-dimensional projections and such things, but I don't know whether that suggests interesting theoretical questions or not. Maybe that's not important, whether it suggests interesting theoretical questions. Maybe the important thing is that it helps contribute to the solution of practical problems.

INFINITE GAMES

DeGroot: What kind of things are you working on these days?

Blackwell: Right now I am working on some things in information theory, and still trying to understand some things about infinite games of perfect information.

DeGroot: What do you mean by an infinite game?

Blackwell: A game with an infinite number of moves. Here's an example. I write down a 0 or a 1, and you write down a 0 or a 1, and we keep going indefinitely. If the sequence we produce has a limiting frequency, I win. If not, you win. That's a trivial game because I can force it to have a limiting frequency just

by doing the opposite of whatever you do. But that's a simple example of an infinite game.

DeGroot: Fortunately, it's one in which I'll never have to pay off to you.

Blackwell: Well, we can play it in such a way that you would have to pay off.

DeGroot: How do we do that?

Blackwell: You must specify a strategy. Let me give you an example. You know how to play chess in just one move: You prepare a complete set of instructions so that for every situation on the chess board you specify a possible response. Your one move is to prepare that complete set of instructions. If you have a complete set and I have a complete set, then we can just play the game out according to those instructions. It's just one move. So in the same way, you can specify a strategy in this infinite game. For every finite sequence that you might see up to a given time as past history, you specify your next move. So you can define this function once and for all, and I can define a function, and then we can mathematically assess those functions. I can prove that there is a specific function of mine such that no matter what function you specify, the set will have a limiting frequency.

DeGroot: So you could extract money from me in a finite amount of time. [Laughs]

Blackwell: Right. Anyway it's been proved that all such infinite games with Borel payoffs are determined, and I've been trying to understand the proof

for several years now. I'm still working on it, hoping to understand it and simplify it.

DeGroot: Have you published papers on that topic?

Blackwell: Just one paper many years ago. Let me remind myself of the title [checking his files], "Infinite games and analytic sets" (*Proc. Natl. Acad. Sci. U.S.A.* **58** 1836–1837, 1967). This is the only paper I've published on infinite games; and that's one of my papers that I like very much, by the way. It's an application of games to prove a theorem in topology. I sort of like the idea of connecting those two apparently not closely related fields.

DeGroot: Have you been involved in applied projects or applied problems through the years, at Rand or elsewhere, that you have found interesting and that have stimulated research of your own?

Blackwell: I guess so. My impression though is this: When I have looked at real problems, interesting theorems have sometimes come out of it. But never anything that was helpful to the person who had the problem. [Laughs]

DeGroot: But possibly to somebody else at another time.

Blackwell: Well, my work on comparison of experiments was stimulated by some work by Bohnenblust, Sherman, and Shapley. We were all at Rand. They called their original paper "Comparison of reconnaissances," and it was *classified* because it arose out of some question that somebody had asked them. I recognized a relation between what they were doing and sufficient statistics, and proved that they were the same in a special case. Anyway, that led to this development which I think is interesting theoretically, and to which you have contributed.

DeGroot: Well, I have certainly used your work in that area. And it has spread into diverse other areas. It is used in economics in comparing distributions of income, and I used it in some work on comparing probability forecasters.

Blackwell: And apparently people in accounting have made some use of these ideas. But anyway, as I say, nothing that I have done has ever helped the person who raised the question. But there is no doubt in my mind that you do get interesting problems by looking at the real world.

"I DON'T HAVE ANY DIFFICULTIES WITH RANDOMIZATION"

DeGroot: One of the interesting topics that comes out of a Bayesian view of statistics is the notion of randomization and the role that it should play in statistics. Just this little example you were talking about before with two proportions made me think

about that. We just assume that we are drawing the observations at random from within each subpopulation in that example, but perhaps basically because we don't have much choice. Do you have any thoughts about whether one should be drawing observations at random?

Blackwell: I don't have any difficulties with randomization. I think it's probably a good idea. The strict theoretical idealized Bayesian would of course never need to randomize. But randomization probably protects us against our own biases. There are just lots of ways in which people differ from the ideal Bayesian. I guess the ideal Bayesian, for example, could not think about a theorem as being probably true. For him, presumably, all true theorems have probability 1 and all false ones have probability 0. But you and I know that's not the way we think. I think of randomization as being a protection against your own imperfect thinking.

DeGroot: It is also to some extent a protection against others. Protection for you as a statistician in presenting your work to the scientific community, in the sense that they can have more belief in your conclusions if you use some randomization procedure rather than your own selection of a sample. So I see it as involved with the sociology of science in some way.

Blackwell: Yes, that's an important virtue of randomization. That reminds me of something else though. We tend to think of evidence as being valid only when it comes from random samples or samples selected in a probabilistically specified way. That's wrong, in my view. Most of what we have learned, we have learned just by observing what happens to come along, rather than from carefully controlled experiments. Sometimes statisticians have made a mistake in throwing away experiments because they were not properly controlled. That is not to say that randomization isn't a good idea, but it is to say that you should not reject data just because they have been obtained under uncontrolled conditions.

DeGroot: You were the Rouse Ball Lecturer at Cambridge in 1974. How did that come about and what did it involve?

Blackwell: Well, I was in England for two years, 1973–1975, as the director of the education-abroad program in Great Britain and Ireland for the University of California. I think that award was just either Peter Whittle's or David Kendall's idea of how to get me to come up to Cambridge to give a lecture. One of the things which delighted me was that it was named the Rouse Ball Lecture because it gave me an opportunity to say something at Cambridge that I liked—namely, that I had heard of Rouse Ball long before I had heard of Cambridge. [Laughs]

DeGroot: Well, tell me about Rouse Ball.

Blackwell: He wrote a book called *Mathematical Recreations and Essays*. You may have seen the book. I first came across it when I was a high school student. It was one of the few mathematics books in our library. I was fascinated by that book. I can still picture it. Rouse Ball was a 19th century mathematician, I think. [Walter William Rouse Ball, 1850–1925] Anyway, this is a lectureship that they have named after him.

DeGroot: I guess there aren't too many Bayesians on the statistics faculty here at Berkeley.

Blackwell: No. I'd say, Lester and I are the only ones in our department. Of course, over in Operations Research, Dick Barlow and Bill Jewell are certainly sympathetic to the Bayesian approach.

DeGroot: Is it a topic that gets discussed much?

Blackwell: Not really. It used to be discussed here but you very soon discover that it's sort of like religion; that it has an appeal for some people and not for other people, and you're not going to change anybody's mind by discussing it. So people just go their own ways. What has happened to Bayesian statistics surprised me. I expected it either to catch on and just sweep the field or to die. And I was rather confident that it would die. Even though to me it was the right way to think, I just didn't think that it would have a chance to survive. But I thought that if it did, then it would sweep things. Of course, neither one of those things has happened. Sort of a steady 5–10% of all the work in statistical inference is done from a Bayesian point of view. Is that what you would have expected 20 years ago?

DeGroot: No, it certainly doesn't seem as though that would be a stable equilibrium. And maybe the system is still not in equilibrium. I see the Bayesian approach growing, but it certainly is not sweeping the field by any means.

Blackwell: I'm glad to hear that you see it growing.

DeGroot: Well, there seem to be more and more *meetings* of the Bayesians, anyway. The actuarial group that met here at Berkeley over the last couple of days to discuss credibility theory seems to be a group that just naturally accepts the Bayesian approach in their work in the real world. So there seem to be some pockets of users out there in the world, and I think maybe that's what has kept the Bayesian approach alive.

Blackwell: There's no question in my mind that if the Bayesian approach does grow in the statistical world it will not be because of the influence of other statisticians but because of the influence of actuaries, engineers, business people, and others who actually like the Bayesian approach and use it.

DeGroot: Do you get a chance to talk much to

researchers outside of statistics on campus, researchers in substantive areas?

Blackwell: No, I talk mainly to people in Operations Research and Mathematics, and occasionally Electrical Engineering. But the things in Electrical Engineering are theoretical and abstract.

“THE WORD ‘SCIENCE’ IN THE TITLE BOTHERS ME A LITTLE”

DeGroot: What do you think about the idea of this new journal, *Statistical Science*, in which this conversation will appear? I have the impression that you think the IMS is a good organization doing useful things, and there is really no need to mess with it.

Blackwell: That is the way I feel. On the other hand, I must say that I felt exactly the same way about splitting the *Annals of Mathematical Statistics* into two journals, and that split seems to be working. So I'm hoping that the new journal will add something. I guess the word “science” in the title bothers me a little. It's not clear what the word is intended to convey there, and you sort of have the feeling that it's there more to contribute a tone than anything else.

DeGroot: My impression is that it *is* intended to contribute a tone. To give a flavor of something broader than just what we would think of as theoretical statistics. That is, to reach out and talk about the impact of statistics on the sciences and the interrelationship of statistics with the sciences, all kinds of sciences.

Blackwell: Now I'm all in favor of that. For example, the relation of statistics to the law is to me a quite appropriate topic for articles in this journal. But somehow calling it “science” doesn't emphasize that direction. In fact, it rather suggests that that's *not* the direction. It sounds as though it's tied in with things that are supported by the National Science Foundation and to me that restricts it.

DeGroot: The intention of that title was to convey a broad impression rather than a restricted one. To give a broader impression than just statistics and probability, to convey an applied flavor and to suggest links to all areas.

Blackwell: Yes. It's analogous to computer science, I guess. I think *that* term was rather deliberately chosen. My feeling is that the IMS is just a beautiful organization. It's about the right size. It's been successful for a good many years. I don't like to see us become ambitious. I like the idea of just sort of staying the way we are, an organization run essentially by amateurs.

DeGroot: Do you have the feeling that the field of statistics is moving away from the IMS in any way? That was one of the motivations for starting this journal.



David Blackwell, 1984.

Blackwell: Well, of course, statistics has always been substantially bigger than the IMS. But you're suggesting that the IMS represents a smaller and smaller fraction of statistical activity.

DeGroot: Yes, I think that might be right.

Blackwell: You know, Morrie, I see what you're talking about happening in mathematics. It's less and less true that all mathematics is done in mathematics departments. On the Berkeley campus, I see lots of interesting mathematics being done in our department, in Operations Research, in Electrical Engineering, in Mechanical Engineering, some in Business Administration, a lot in the Economics Department by Gerard Debreu and his colleagues; a lot of really interesting, high class mathematics is being done outside mathematics departments. What you're suggesting is that statistics departments and the journals in which they publish are not necessarily the centers of statistics the way they used to be, that a lot of work is being done outside. I'm sure that's right.

DeGroot: And perhaps *should* be done outside statistics departments. That used to be an unhealthy sign in the field, and we worked hard in statistics departments to collect up the statistics that was being done around the campus. But I think now that the field has grown and matured, that it is probably a healthy thing to have some interesting statistics being done outside.

Blackwell: Yes. Consider the old problem of pattern recognition. That's a statistical problem. But to the extent that it gets solved, it's not going to be solved by people in statistics departments. It's going to be solved by people working for banks and people

working for other organizations who really need to have a device that can look at a person and recognize him in lots of different configurations. That's just one example of the cases where we're somehow too narrow to work on a lot of serious statistical problems.

DeGroot: I think that's right, and yet we have something important to contribute to those problems.

Blackwell: I would say that we *are* contributing, but indirectly. That is, people who are working on the problems have studied statistics. It seems to me that a lot of the engineers I talk to are very familiar with the basic concepts of decision theory. They know about loss functions and minimizing expected risks and such things. So, we have contributed, but just indirectly.

DeGroot: You are in the National Academy of Sciences . . .

Blackwell: Yes, but I'm very inactive.

DeGroot: You haven't been involved in any of their committees or panels?

Blackwell: No, and I'm not sure that I would want to be. I guess I don't like the idea of an official committee making scientific pronouncements. I like people to form opinions about scientific matters just on the basis of listening to individual scientists. To have one group with such overwhelming prestige bothers me a little.

DeGroot: And it is precisely the prestige of the Academy that they rely on when reports get issued by these committees.

Blackwell: Yes. So I think it's just great as a purely honorific organization, so to speak. To meet

just once a year, and elect people more or less at random. I think everybody that's in it has done something reasonable and even pretty good, in fact. But on the other hand, there are at least as many people *not* in it who have done good things as there are in it. It's kind of a random selection process.

DeGroot: So you think it's a good organization as long as it doesn't do anything.

Blackwell: Right. I'm proud to be in it, but I haven't been active. It's sort of like getting elected to Phi Beta Kappa—it's nice if it happens to you . . .

"I PLAY WITH THIS COMPUTER"

DeGroot: Do you feel any relationship between your professional work and the rest of your life, your interests outside of statistics? Is there any influence of the outside on what you do professionally, or are they just sort of separate parts of your life?

Blackwell: Separate, except that my friends are also my colleagues. It's only through the people with whom I associate outside that there's any connection. It's hard to think of any other real connection.

DeGroot: It's not obvious what these connections might be for anyone. One's political views or social views seem to be pretty much independent of the technical problems we work on.

Blackwell: Yes. Although it's hard to see how it could *not* have an influence, isn't it? I guess my life seems all of a piece to me but yet it's hard to see where the connections are. [Laughs]

DeGroot: What do you see for your future?

Blackwell: Well, just gradually to wind down, gracefully I hope. I expect to get more interested in computing. I have a little computer at home, and it's a lot of fun just to play with it. In fact, I'd say that I play with this computer here in my office at least as much as I do serious work with it.

DeGroot: What do you mean by play?

Blackwell: Let me give you an example. You know the algorithm for calculating square roots. You start with a guess and then you divide the number by your guess and take the average of the two. That's your next guess. That's actually Newton's method for finding square roots, and it works very well. Sometimes doing statistical work, you want to take the square root of a positive definite matrix. It occurred to me to ask whether that algorithm works for finding the square root of a positive definite matrix. Before I got interested in computing, I would have tried to solve it theoretically. But what did I do? I just wrote up a program and put it on the computer to see if it worked. [Goes to blackboard]

Suppose that you are given the matrix M and want to find $M^{1/2}$. Let G be your guess of $M^{1/2}$. Then your

new guess is $\frac{1}{2}(G + MG^{-1})$. You just iterate this and see if it converges to $M^{1/2}$. Now, Morrie, I want to show you what happens. [Goes to terminal]

Let's do it for a 3×3 matrix. We're going to find the square root of a positive definite 3×3 matrix. Now, if you happen to have in mind a particular 3×3 positive definite matrix whose square root you want, you could enter it directly. I don't happen to have one in mind, but I do know a theorem: If you take any nonsingular 3×3 matrix A , then AA' is going to be positive definite. So I'm just going to enter any 3×3 nonsingular matrix [putting some numbers into the terminal] and let $M = AA'$. Now, to see how far off your guess G is at any stage, you calculate the Euclidean norm of the 3×3 matrix $M - G^2$. That's what I call the error. Let's start out with the identity matrix I as our initial guess. We get a big error, 29 million. Now let's iterate. Now the error has dropped down to 7 million. It's going to keep being divided by 4 for a long time. [Continuing the iterations for a while] Now notice, we're not bad. There's our guess, there's its square, there's what we're trying to get. It's pretty close. In fact the error is less than one. [Continuing] Now the error is really small. Look at that, isn't that beautiful? So there's just no question about it. If you enter a matrix at random and it works, then that sort of settles it.

But now wait a minute, the story isn't quite finished yet. Let me just continue these iterations . . . Look at that! The error got bigger, and it keeps getting bigger. [Continuing] Isn't that lovely stuff?

DeGroot: What happened?

Blackwell: Isn't that an interesting question, what happened? Well, let me tell you what happened. Now you can study it theoretically and ask, should it converge? And it turns out that it will converge if, and essentially only if, your first guess commutes with the matrix M . That's what the theory gives you. Well, my first guess was I . It commutes with everything. So the procedure theoretically converges. However, when you calculate, you get round-off errors. By the way, if your first guess commutes, then all subsequent guesses will commute. However, because of round-off errors, the matrices that you actually get don't *quite* commute. There are two ways to do this. We could take MG^{-1} or we could have taken $G^{-1}M$. Of course, if M commutes with G , then it commutes with G^{-1} and it doesn't matter which way you do it. But if you don't calculate G exactly at some stage, then it will not quite commute. And in fact, what I have here on the computer is a calculation at each stage of the noncommutativity norm. That shows you how different MG^{-1} is from $G^{-1}M$. I didn't point those values out to you, but they started out as essentially 0, and then there was a 1 in the 15th place, and then a 1 in the 14th

place, and so on. By this stage, the noncommutativity norm has built up to the point where it's having a sizable influence on the thing.

DeGroot: Is it going to diverge or will it come back down after some time?

Blackwell: It won't come back down. It will reach a certain size, and sometimes it will stay there and sometimes it will oscillate. That is, one G will go into a quite different G , but then that G will come back to the first one. You get periods, neither one of them near the truth. So that's what I mean by just playing, instead of sitting down like a serious mathematician and trying to prove a theorem. Just try it out on the computer and see if it works. [Laughs]

DeGroot: You can save a lot of time and trouble that way.

Blackwell: Yes. I expect to do more and more of that kind of playing. Maybe I get lazier as I get older. It's fun, and it's an interesting toy.

DeGroot: Do you find yourself growing less rigorous in your mathematical work?

Blackwell: Oh yes. I'm much more interested in the ideas, and in truth under not-completely-specified hypotheses. I think that has happened to me over the last 20 years. I can certainly notice it now. Jim MacQueen was telling me about something that he had discovered. If you take a vector and calculate the squared correlation between that vector and some permutation of itself, then the average of that squared correlation over all possible permutations is some simple number. Also, there was some extension of this result to k vectors. He has an interesting algebraic identity. He told me about it, but instead of my trying to prove it, I just selected some numbers at random and checked it on the computer. Also, I had a conjecture that some stronger result was true. I checked it for some numbers selected at random and it turned out to be true for him and *not* true for what I had said. Well, that just settles it. Because suppose you have an algebraic function $f(x_1, \dots, x_n)$ and you want to find out if it is identically 0. Well, I think it's true that any algebraic function of n variables is either identically 0 or the set of x 's for which it is 0 is a set that has measure 0. So you can just select x 's at random and evaluate f . If you get 0, it's identically 0. [Laughs]

DeGroot: You wouldn't try even a second set of x 's?

Blackwell: I did. [Laughs]

DeGroot: Getting more conservative in your old age.

Blackwell: Yes. [Laughs] I've been wondering whether in teaching statistics the typical set-up will be a lot of terminals connected to a big central computer or a lot of small personal computers. Let me turn the interview around. Do you have any thoughts

about which way that is going or which way it ought to go?

DeGroot: No, I don't know. At Carnegie-Mellon we are trying to have both worlds by having personal computers but having them networked with each other. There's a plan at Carnegie-Mellon that each student will have to have a personal computer.

Blackwell: Now when you say each student will have to have a personal computer, where will it be physically located?

DeGroot: Wherever he lives.

Blackwell: So that they would not actually use computers in class on the campus?

DeGroot: Well, this will certainly lessen the burden on the computers that are on campus, but in a class you would have to have either terminals or personal computers for them.

Blackwell: Yes. I'm pretty sure that in our department in five years we'll have several classrooms in which each seat will be a work station for a student, and in front of him will be either a personal computer or a terminal. I'm not sure which, but that's the way we're going to be in five years.

"I WOULDN'T DREAM OF TALKING ABOUT A THEOREM LIKE THAT NOW"

DeGroot: A lot of people have seen you lecture on film. I know of at least one film you made for the American Mathematical Society that I've seen a few times. That's a beautiful film, "Guessing at Random."

Blackwell: Yes. I now, of course, don't think much of those ideas. [Laughs]

DeGroot: There were some *minimax* ideas in there . . .

Blackwell: Yes, that's right. That was some work that I did before I became such a committed Bayesian. I wouldn't dream of talking about a theorem like that now. But it's a nice result . . .

DeGroot: It's a nice result and it's a beautiful film. Delivered so well.

Blackwell: Let's see . . . How does it go? If I were doing it now I would do a weaker and easier Bayesian form of the theorem. You were given an arbitrary sequence of 0's and 1's, and you were going to observe successive values and you had to predict the next one. I proved certain theorems about how well you could do against every possible sequence. Well, *now* I would say that you have a probability distribution on the set of all sequences. It's a general fact that if you're a Bayesian, you don't have to be clever. You just *calculate*. Suppose that somebody generates an arbitrary sequence of 0's and 1's and it's your job after seeing each finite segment to predict the next coordinate, 0 or 1, and we keep track of how well you do. Then I

have to be clever and invoke the minimax theorem to devise a procedure that asymptotically does very well in a certain sense. But now if you just put a prior distribution on the set of sequences, any Bayesian knows what to do. You just calculate the probability of the next term being a 1 given the past history. If it's more than $\frac{1}{2}$ you predict a 1, if it's less than $\frac{1}{2}$ you predict a 0. And that simple procedure has the corresponding Bayesian version of all the things that I talked about in that film. You just know what is the right thing to do.

DeGroot: But how do you know that you'll be doing well in relation to the reality of the sequence?

Blackwell: Well, the theorem of course says that you'll do well for all sequences except a set of measure zero according to your own prior distribution, and that's all a Bayesian can hope for. That is, you have to give up something, but it just makes life so much *neater*. You just know that this is the right thing to do.

I encountered the same phenomenon in information theory. There is a very good theory about how to transmit over a channel, or how to transmit over a sequence of channels. The channel may change from day to day, but if you know what it is every day, then you can transmit over it. Now suppose that the channel varies in an arbitrary way. That is, you have one of a finite set of channels, and every day you're going to be faced with one of these channels. You have to put in the input and a guy at the other end gets an output. The question is, how well can you do against all possible channel sequences?

You don't really know what the weather is out there, so you don't know what the interference is going to be. But you want to have a code that transmits well for all possible weather sequences. If you just analyze the problem crudely, it turns out that you can't do *anything* against all possible sequences. However, if you select the code in a certain random way, your overall error probability will be small for each weather sequence. So you see, it's a nice theoretical result but it's unappealing. However, you can get exactly the same result if you just put a probability distribution on the sequences. Well, the weather could be any sequence, but you expect it to be sort of this way or that. Once you put a probability distribution on the set of sequences, you no longer need random codes. And there is a deterministic code that gives you that same result that you got before. So either you must behave in a random way, or you must put a probability distribution on nature.

[Looking over a copy of his paper, BLACKWELL, D., BREIMAN, L. and THOMASIAN, A. J., "The capacities of certain channel classes under random coding," *Ann. Math. Statist.* **31** 558-567, 1960] I don't think we did

the nice easy part. We behaved the way Wald behaved. You see, the minimax theorem says that if for every prior distribution you can achieve a certain gain, then there is a random way of behaving that achieves that gain for every parameter value. You don't need the prior distribution; you can throw it away. Well, I'm afraid that in this paper, we invoked the minimax theorem. We said, take any prior distribution on the set of channel sequences. Then you can achieve a certain rate of transmission for that prior distribution. Now you invoke the minimax theorem and say, therefore, there is a randomized way of behaving which enables you to achieve that rate against every possible sequence. I now wish that we had *stopped* at the earlier point. [Laughs] For us, the Bayesian analysis was just a preliminary which, with the aid of the minimax theorem, enabled us to reach the conclusions we were seeking. That was Wald's view and that's the view that we took in that paper. I'm sure I was already convinced that the Bayes approach was the right approach, but perhaps I deferred to my colleagues.

DeGroot: That's a very mild compromise. Going *beyond* what was necessary for a Bayesian resolution of the problem.

Blackwell: That's right. Also, I suspect that I had Wolfowitz in mind. He was a real expert in information theory, but he wouldn't have been interested in anything Bayesian.

DeGroot: What about the problem of putting prior distributions on spaces of infinite sequences or function spaces? Is that a practical problem and is there a practical solution to the problem?

Blackwell: I wouldn't say for infinite sequences, but I think it's a very important practical problem for large finite sequences and I have no idea how to solve it. For example, you could think that the pattern recognition problem that I was talking about before is like that. You see an image on a TV screen. That's just a long finite sequence of 0's and 1's. And now you can ask how likely it is that that sequence of 0's and 1's is intended to be the figure 7, say. Well, with some you're certain that it is and some you're certain that it isn't, and with others there's a certain probability that it is and a probability that it isn't. The problem of describing that probability distribution is a very important problem. And we're just not close to knowing how to describe probability distributions over long finite sequences that correspond to our opinions.

DeGroot: Is there hope for getting such descriptions?

Blackwell: I don't know. But again it's a statistical problem that is not going to be solved by professors of statistics in universities. It might be solved by people in artificial intelligence, or by researchers outside universities.

“JUST TELL ME ONE OR TWO INTERESTING THINGS”

DeGroot: There’s an argument that says that under the Bayesian approach, you have to seek the optimal decision and that’s often just too hard to find. Why not settle for some other approach that requires much less structure and get a reasonably good answer out of it, rather than an optimal answer? Especially in these kinds of problems where we don’t know how to find the optimal answer.

Blackwell: Oh, I think everybody would be satisfied with a reasonable answer. I don’t see that there’s more of an emphasis in the Bayesian approach on optimal decisions than in other approaches. I separate Bayesian inference from Bayesian decision. The inference problem is just calculating a posterior distribution, and that has nothing to do with the particular decision that you’re going to make. The same posterior distribution could be used by many different people making different decisions. Even in calculating the posterior distribution, there is a lot of approximation. It just can’t be done precisely in interesting and important cases. And I don’t think anybody who is interested in applying Bayes method would insist on something that’s precise to the fifth decimal place. That’s just the conceptual framework in which you want to work, and which you want to approximate.

DeGroot: That same spirit can be carried over into the decision problem, too. If you can’t find the optimum decision, you settle for an approximation to it.

Blackwell: Right.

DeGroot: In your opinion, what have been the major breakthroughs in the field of statistics or probability through the years?

Blackwell: It’s hard to say . . . I think that theoretical statistical thinking was just completely dominated by Wald’s ideas for a long time. Charles Stein’s

discovery that \bar{X} is inadmissible was certainly important. Herb Robbin’s work on empirical Bayes was also a big step, but possibly in the wrong direction.

You know, I don’t view myself as a statesman or a guy with a broad view of the field or anything like that. I just picked directions that interested me and worked in them. And I have had fun.

DeGroot: Well, despite the fact that you didn’t choose the problems for their impact or because of their importance, a lot of people have gained a lot from your work.

Blackwell: I guess that’s the way scholars *should* work. Don’t worry about the overall importance of the problem; work on it if it looks interesting. I think there’s probably a sufficient correlation between interest and importance.

DeGroot: One component of the interest is probably that others are interested in it, anyway.

Blackwell: That’s a big component. You want to tell somebody about it after you’ve done it.

DeGroot: It has not always been clear that the published papers in our more abstract journals did succeed in telling anybody about it.

Blackwell: That’s true. But if you get the fellow to give a lecture on it, he’ll probably be able to tell you something about it. Especially if you try to restrict him: Look, don’t tell me everything. Just tell me *one or two* interesting things.

DeGroot: You have a reputation as one of the finest lecturers in the field. Is that your style of lecturing?

Blackwell: I guess it is. I try to emphasize that with students. I notice that when students are talking about their theses or about their work, they want to tell you everything they know. So I say to them: You know much more about this topic than anybody else. We’ll never understand it if you tell it all to us. Pick just one interesting thing. Maybe two.

DeGroot: Thank you, David.





