Rubin, D. R. (1993), "The Future of Statistics," *Statistics and Computing*, 3, 204.

Tanur, J. M., et al. (1987), *Statistics: A Guide to the Unknown* (2nd ed.), San Francisco: Holden-Day.

Tukey, J. W. (1962), "The Future of Data Analysis," *The Annals of Mathematical Statistics*, 33, 1–67.

——— (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Watts, D. G. (1968), *The Future of Statistics*, New York: Academic Press.

Discussion

David R. Brillinger

University of California
Berkeley, CA 94720
(*brill@stat.Berkeley.edu*)

1. INTRODUCTION

It is a total pleasure to be invited to comment on Colin's timely paper. In it Colin refers to Bell Labs and AT&T several times. Further, the Tukey (JWT) paper lists his affiliations as Princeton University and Bell Telephone Laboratories, so I seize an opportunity to celebrate the Labs of the early 1960s as well as comment on his ideas.

Colin's paper brings back so many memories of the 1960–1964 period: anecdotes, FFTs, lunches, seminars, Hamming, Tukey, Hamming–Tukey, golf, learning, visitors, computing, books, history, open doors, pink paper drafts, technical reports, rides between Princeton and Murray Hill, shared offices, AMTSs, chiding, support (personal and financial), opportunities (both seized and missed), blackboards, air-conditioning, freedom, confidence, pranks, Tukey anecdotes, gossip, conferences, unpublished memos, and people who are no longer with us. Pursuit of excellence was the order of the day. I could write a page or more on each of these topics, but this is not the place.

I was at Bell Labs for the summers of 1960, 1961, and then for the years 1962–1964. I was a summer student at first and next a Member of Technical Staff (MTS). These were magic years at a magic place. None of the involved persons with whom I have used the term have ever disagreed. I can say that everything important about statistics that I ever learned, I learned at lunch at Murray Hill. The rest of my career has been applying what I learned.

Colin reviews a place (University College London, 1948–1958) and people (Fisher, Hotelling, Tukey) in his paper. I will do the same.

2. THE PEOPLE

Colin is, of course, one of the key influences, drivers, critics, and contributors to the development of modern data analysis. He is a problem solver with few if any peers. At the Labs he used to be in his office (with door wide open), at lunch, always available and always interruptible. The others in the group with wide-open doors and a thirst for discovery included Martin Wilk, Ram Gnanadesikan, Bill Williams, Roger Pinkham, and a stream of visitors. Of course, John Tukey dropped in/appeared steadily from the management wing of the buildings. The fields of expertise included sampling, multivariate analysis, time series, analysis of variance, and the newly defined field of data analysis. [Gnanadesikan (2001) reminded me that JWT came up with the term "data analysis" at a party at my house in 1960. Ram's paper contains many reminiscences about the Labs and comments on data analysis.]

Martin Wilk went on to become a Vice President of AT&T and then Director of Statistics Canada. He was one of the few people who could cause John Tukey to really focus on the topic at hand. (JWT was one of the great multiprocessors and typically focused on several things at a time.) In particular, Martin could sum up mighty ideas in a pithy phrase or sentence. To give an example, there was a scorn for significance tests at the Labs. Martin remarked: "Significance tests are things to do while your are thinking about what you really want to do." Both Colin and Martin went on to write influential papers with Tukey on exploratory data analysis.

3. THE RESEARCH

The Labs' researchers' directions then were not specifically laid out by the higher-ups, rather various management and engineering types would drop in with problems. It seemed that few, if any, in the statistics group could resist these problems, puzzles, or datasets. There were expected and unexpected discoveries. Terminology was created, graphic displays were basic, residuals were fodder, engineering and chemical science were ever present. Gnanadesikan (2001) used the word "synergy" to describe the milieu.

A theme of my discussion is that the Labs of the early 1960s were magic years for data analysis. They were also magic years for the digitization of the engineering sciences. The FFT (fast Fourier transform) has been mentioned, but also seismic records and speech were being digitized and an analysis sometimes culminated with an analog record. I mention this because a great talent that Colin brought to the Statistics Group was skills in combinatorics and discrete mathematics.

TECHNOMETRICS, AUGUST 2006, VOL. 48, NO. 3
DOI 10.1198/004017006000000200

4. TUKEY'S PAPER

"Tukey's paper" was the first article of the first number of the *Annals of Mathematical Statistics* of 1962. The editor at that time was J. L. Hodges Jr., who was renowned for both theoretical and applied statistics work. No thanks are given in the paper to referees, so perhaps the editor published it on his own authority. The paper had been received by the *Annals* on July 1, 1961 and was presented at the IMS Meeting in Seattle in 1961, so it was out in public.

Tukey's Foreward to the *Collected Works* (Jones 1986) is worth a read. For example, one finds at the beginning: "Besse Day (Mauss), who spent a year with R. A. Fisher, once told me that he told her that 'all he had learned he had learned over the (then hand-cranked) calculating machine'." I record this quote to lead into the remark that JWT was involved in more than pencil-and-paper data analyses. Tukey's paper presents an example. There are several analyses of one particular dataset, a 36 × 15 table of the values of some particular multiple regression coefficients. JWT presents a robust/resistant row/column fitting procedure. The Foreward is also interesting for JWT's comments on Bayesian statistics.

5. COLIN'S PAPER

Colin asks a sequence of questions:

- "How do we attract the brightest students to our subject?"
- "How to convey this to a bright student, who has some analytical attitude, but who is attracted to the glamour of pure science (or math), or the promise of riches in Wall Street?"
- "Is statistics a science?"
- "If statistics is a science, what is its subject matter?"
- "What do statisticians study?"
- "The question remains, is statistics a science?"
- "But is statistics itself a science?"
- "So is statistics, or data analysis if you prefer, a science?"
- "Surely each of these applications areas is not completely different from all the others?"
- "How does one choose an appropriate methodology?"

6. SOME ANSWERS TO THE QUESTIONS

First off, I am not going to get into the "is it a science?" discussion, because I just do not think that it matters much. I am happy to view "statistics/data analysis" as a fine endeavor that provides much amusement and contributions of insight and understanding to scientific researchers. I leave the question to others, but note that Colin mentions his "sincere admiration for engineers, who have to make things work in the real world" (I have heard this sentiment phrased as "every engineering problem has a solution"), and engineering statistics is one of our subfields (see *Technometrics*).

However, "how to involve students" is a question dear to my heart. I do have suggestions:

- Get them to read books like the Hoaglin–Mosteller–Tukey (1983, 1985, 1991) series. (I note Colin's chiding of JWT's EDA book with "his 1977 EDA book discusses the methods of exploratory data analysis, but says nothing about how to use these methods.")
- Get them to attend pertinent courses.
- Teach pertinent courses.
- Get them to attend talks, and get talks presented.
- Pay them well.
- Raid the computer science departments. (There are lots of straight computing problems, like how to work out bagplots and how to speed up computations, that can lure students in.)

My own serious attempt at an original course was Statistics 215a, taught in the fall semesters of 2003 and 2004 here at Berkeley. The syllabus, book list, and readings are provided in the Appendix.

Another attempt I made was to use the book of De Veaux, Velleman, and Bock (2006) as text in a third-year undergraduate course. In it many EDA techniques are illustrated, there is a chapter on "Regression Wisdom," and one finds the stricture "Make a picture. Make a picture. Make a picture." repeated many times. (This was a Labs mantra.) Students from a broad group of departments registered for the course and appeared to grasp the EDA concepts almost immediately.

I am sure others teach such courses. It strikes me that one does not have to yearn for a reincarnation of that 1960s Labs environment, because the ideas are out and Tukey-type data analysis is now the order of the day.

7. SUMMARY

I call this 1960–1964 period "magic years" because the seeds for high-quality statistical analyses were sown then, and analyses in which electronic computers, graphics, and residuals became paramount. Sadly, one cannot say the same about the Labs; how the mighty have fallen.

I end with the following note. There was talk at the 1960s lunches of forming a Society of Data Analysis. My contribution was to suggest that Tukey could be called "soda pop."

APPENDIX: STATISTICS 215A "APPLIED STATISTICS AT AN ADVANCED LEVEL," UNIVERSITY OF CALIFORNIA BERKELEY 2003, 2004

Syllabus

Week 1. Stem-and-leaf, 5-number summary, boxplot, parallel boxplots, examples

Week 2. EDA vs. CDA vs. DM, magical thinking, scatter plots, pairs(), bagplot(), spin()

Week 3. Summaries of location, spread vs. level plot, empirical Q–Q plot, smoothing scatterplots, smoothing types

Week 4. The future of data analysis, linear fitting, OLS, WLS, NLS, multiple OLS, robust/resistant fitting of straight line

Week 5. Optimization methods, the psi function, residual analysis, fitting by stages, the *x*-values

Week 6. Wavelets, NLS, robust/resistant variants, smoothing/nonparametric regression, sensitivity curve, two-way arrays

Week 7. Residuals analysis for two-way array, L1 approximation, median polish, diagnostic plot, data analysis and statistics: an expository overview

Week 9. Exploratory analysis of variance: terminology, overlays, ANOVA table, rob/res methods, examples

Week 10. Some principles of data analysis

Week 11. $r-2$, $R-2$, Simpson's paradox, lurking variables

Week 12. Exploratory time series analysis (ETSA), plotting time series, methods

Week 13. Data mining, definitions; contrasts with statistics

Week 14. Data mining for time series, for association rules, market basket analysis.

Book List

Cleveland, W. S. (1994), *The Elements of Graphing Data*, Belmont, CA: Wadsworth.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983), *Graphical Methods for Data Analysis*, Duxbury.

Hand, D., Mannila, H., and Smyth, P. (2000), *Principles of Data Mining*, Cambridge, MA: MIT Press.

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning*, New York: Springer-Verlag.

Hoaglin, D., Mosteller, F., and Tukey, J. (1983), *Understanding Robust and Exploratory Data Analysis*, New York: Wiley.

——— (1985), *Exploring Data Tables*, *Trends, and Shapes*, New York: Wiley.

——— (1991), *Fundamentals of Exploratory Analysis of Variance*, New York: Wiley.

Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression*, Reading, MA: Addison-Wesley.

Rao, C. R. (2002), *Linear Statistical Inference and Its Applications*, New York: Wiley.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics With S–PLUS*, New York: Springer-Verlag.

Readings

Breiman, L. (2001), "Statistical Modeling: The Two Cultures," *Statistical Science*, 16, 199–231.

Diaconnis, P. (1985), "Theories of Data Analysis: From Magical Thinking Through Classical Statistics," in *Exploring Data Tables*, *Trend, and Shapes*, eds. D. Hoaglin, F. Mosteller, and J. Tukey, New York: Wiley, pp. 1–36.

Friedman, J. H. (2001), "The Role of Statistics in the Data Revolution," *International Statistical Review*, 29, 5–10.

Hand, D. J. (1998), "Data Mining: Statistics and More," *The American Statistician*, 52, 112–118.

Mallows, C., and Pregibon, D. (1987), "Some Principles of Data Analysis," in *Proceedings of the 46th Session ISI*, Tokyo, pp. 267–278.

Mannila, H. (2001), "Theoretical Framework for Data Mining," *SIGKDD*, 1, 30–32.

——— (1980), "We Need Both Exploratory and Confirmatory," in *The Collected Works of John W. Tukey*, ed. L. V. Jones, Monterey, CA: Wadsworth & Brooks/Cole, pp. 811–817.

Tukey, J. W. (1962), "The Future of Data Analysis," in *The Collected Works of John W. Tukey*, ed. L. V. Jones, Monterey, CA: Wadsworth & Brooks/Cole, pp. 391–484.

Tukey, J. W., and Wilk, M. B. (1966), "Data Analysis and Statistics: An Expository Overview," in *The Collected Works of John W. Tukey*, ed. L. V. Jones, Monterey, CA: Wadsworth & Brooks/Cole, pp. 549–578.

ADDITIONAL REFERENCES

DeVeaux, R. D., Velleman, P. F., and Bock, D. E. (2006), *Introductory Statstics*, Boston: Pearson, Addison-Wesley.

Gnanadesikan, R. (2001), "A Conversation With Ramanathan Gnanadesikan," *Statistical Science*, 16, 295–309.

Hoaglin et al., see the Appendix.

Jones, L. V. (1986), *The Collected Works of John W. Tukey*, Vols. III and IV, Monterey, CA: Wadsworth & Brooks/Cole.

Mallows, C., and Tukey, J. W. (1982), "An Overview of Techniques of Data Analysis, Emphasizing Its Exploratory Aspects," in *The Collected Works of John Tukey*, ed. L.V. Jones, Monterey, CA: Wadsworth & Brooks/Cole, pp. 891–968.

Discussion

Andreas Buja

Department of Statistics
University of Pennsylvania
Philadelphia, PA 19104
(*buja@wharton.upenn.edu*)

Colin Mallows discussion of Tukey's paper gives us an opportunity to clarify our thoughts about the state of the field. Before I enter into a debate with Colin, I will follow his lead by reminiscing about the past—a more recent past than his, however.

It used to be that self-identification as a statistician, at parties, say, produced rambling responses about "the worst class I had to take in college." The confession "I'm in statistics" was not exactly a conversation stopper, but it did not move the conversation in a desirable direction either. This I remember from the 1980s. Did we have a problem back then, and, if so, do we still have it today?

TECHNOMETRICS, AUGUST 2006, VOL. 48, NO. 3
DOI 10.1198/004017006000000192