TUKEY, JOHN WILDER

Abstract. John Tukey was one of the great statistical scientists of the twentieth century. He introduced algorithms, concepts, language, philosophy, and techniques. He made important contributions to government, to industry and to science.

Keywords and Phrases. Bell Labs, Computational mathematics, Computers, Cryptography, Data analysis, Exploratory data analysis, History, Information science, Mathematics, Neologisms, Princeton University.

AMS Subject Classification. Primary: 62-07, 62-09, 62M15

Secondary: 54A05, 6205, 62F40, 62G35, 62H99, 62J05, 62J10, 62J15, 65T50, 97A20, 97B10

Born: June 16, 1915, in New Bedford, Massachussets.

Died: July 26, 2000, in New Brunswick, New Jersey.

Contributed to: algorithms, analysis of variance, confirmatory data analysis, exploratory data analysis, environmental science, graphics, military analysis, multiple comparisons, multivariate analysis, psephology, recreational mathematics, regression, robustness, sampling, statistical philosophy, statistical practice, statistical theory, scientific education, time series, topology.

John Tukey was a professor and an executive. He merged the scientific, governmental, technological and industrial worlds more seamlessly than, perhaps, anyone else in the 1900's. His scientific knowledge, creativity, experience, calculating skills and energy were prodigious. He was renowned for conceptualizing problems, creating statistical concepts and coming up with both new words for new ideas and new uses for old words. There are few branches of statistics that his ideas have not impinged upon.

Tukey was born in New Bedford, Massachusets. After being schooled at home, he entered Brown University by way of College Board Exams and earned bachelor's and master's degrees in chemistry. For a Ph.D. he went to Princeton. On arrival there he quickly switched to mathematics. During World War II Tukey moved on to statistics. He often commented on being influenced by E. Anderson and C. P. Winsor andf that he learned statistics by reading the *Journal of the Royal Statistical Society Supplement*.

FAMILY

Tukey was the only child of Adah M. Tasker and Ralph H. Tukey. His parents had graduated 1 and 2 in the Bates College class of 1898. Ralph Tukey obtained a Ph.D. in Latin from Yale University and later taught at New Bedford High School. Tukey's mother was a substitute teacher there. Tukey was a prodigy learning to read at a very early age. Perhaps, in consequence, he was schooled at home. His wife Elizabeth Rapp was born March 2, 1920 in Ocean City, New Jersey. She went to Temple University and later was valedictorian in the 1944 class in Business Administration at Radcliffe College. Elizabeth and John married on July 19, 1950. Elizabeth died January 6, 1998. The Tukeys had no children, but they had a large extended family.

The statistician Francis J. Anscombe was Tukey's brother-in-squared-law. They had married to sisters.

CAREER

Brown University

Brown University was very important to John Tukey. He began his formal education there obtaining bachelors and masters degrees. Later he was a member the

Brown Corporation, the group that is responsible for the university. During other periods he was a member of the Computer and Library Committees. He was also involved in Mathematics Department personnel affairs to an extent.

Princeton University - early years

Tukey's residence was the Graduate College. There and at Fine Hall he came into close contact with the likes of R. Feynman, A. Turing, O. Morgenstern, M. Morse, and J. von Neumann. As Princeton surely had the best mathematics department in the late thirties Tukey made important contacts, including ones with young researchers from Europe. He obtained a doctorate in 1939 with a thesis on topology supervised by S. Lefschetz.

World War II

Tukey remained at Princeton for his whole academic career. He was an Instructor from 1939 to 1941 then an Assistant Professor from 1941 to 1948. With the advent of the war in Europe Tukey came to be a Research Associate of the Fire Control Research Office (FRCO). There he worked on problems of stereoscopic rangefinders, gunfire control, particularly from a B29 bomber, and computing leads for machine guns aiming at fighter planes. He was also involved in the Enigma code breaking operations.

Others working at the FRCO at that time were F. Mosteller, W. J. Dixon, and A. M. Mood. Tukey's future years were much influenced by his colleague Charles P. Winsor. Tukey later said that *"It was Charlie and the experience of real data that converted me to statistics."*

Princeton University - postwar

After the war Tukey took up the professional life of a half-time appointment in the Mathematics Department of Princeton and the other half as a member of Technical Staff at Bell Telephone Laboratories. This split appoinment continued until his retirement in 1985. He became Associate Professor in 1948 and Professor in 1950.

In the late fifties Tukey was instrumental in setting up the Statistical Techniques Research Group and the Institute of Defense Analysis Branch at Princeton. In 1966 he became the first Chair of the Statistics Department.

Bell Labs

Tukey went to Bell Telephone Laboratories, Murray Hill in 1945 as a computational topologist. He became part of a team developing the Nike missile, in particular aerodynamics and trajectory. He worked his way up the managerial ladder of the Labs ending with the position of Associate Executive Director Information Sciences.

At Tukey's 1985 retirement event, retired Chairman of the Board W. O. Baker said: *"John has had an incisive role in each major frontier of telecommunications science and technology: uses of transistors and solid state; digital coding and computers; ... ; evolution of software and operations support systems; earth satellite and other microwave techniques; electronic switching; laser-based photonics; topology of integrated circuits; ...",*

Advising and Consulting

Throughout his career Tukey was called upon for advice and invited to be a consultant. In 1954 he was a Member of the Killam Committee charged by President Eisenhower to study the possibility of a surprise nuclear attack. That committee recommended the construction of the U-2 airplane. In 1959 JWT was a U.S. Delegate to Technical Working Group 2 of the Conference on the Discontinuance of Nuclear Weapon Tests. His expertise, in part, concerned the time series problem of distinguishing earthquakes from nuclear explosions. He sought improved estimates of the directions of first motion of seismic vibrations and estimates of the depth of the event that was causing the vibrations.

He was a Member of the President's Scientific Advisory Committee (PSAC) starting in 1960. All told he advised five U.S. presidents. From 1965 until his death Tukey was a consultant to the Educational Testing Service (ETS), Princeton. He also had a long involvement with the National Assessment of Educational Progress (NAEP). In 1970-1 he was a member of the President's Commission on Federal Statistics. In 1972 he was a Member of the U.S. Delegation to the U.N. Conference on the Human Environment that took place in Stockholm. Tukey advised the Census Bureau for many years. In 1980 the Census Bureau's step of adjusting the raw values to obtain "improved" estimates became a political issue. Tukey testified in support of adjustment in the court case that resulted.

In addition to ETS he consulted for Merck, RCA, Bellcore, Xerox Palo Alto Research Center, the Health Effects Institute, and many others.

CONTRIBUTIONS

Mathematical

Tukey's name lives on in topology: "Galois-Tukey connections", "Tukey equivalence", "Tukey reducibility", "the Tukey theory of analytic ideals", "Tukey ordering" as well as "Tukey's Lemma". His name is also associated with the so-called ham-sandwich theorem. He did some recreational mathematics in the late 1930's (flexagons, catching lions) and later contributed to numerical analysis. One can mention H. Trotter's and Tukey's development of conditional Monte Carlo for example.

The discovery that brought Tukey's name to world-wide prominence was the Fast Fourier Transform (FFT). A particular case was presented in a course at Princeton

1963 and developed into a clear algorithm by Cooley [4]. The impact of the work was astonishing. Almost overnight engineers switched from analog to digital signal processing.

Statistical

Sampling. In 1950, following the substantial criticism of A. C. Kinsey's book, *Sexual Behavior in the Human Male*, the American Statistical Association assembled a committee to review the statistical problems arising. The Committee members were W. G. Cochran, F. Mosteller and J. W. Tukey. They were were concerned particularly with the sampling methods and the absence of controlled randomness in Kinsey's work. The committee's efforts led to substantial advances in the theory and understanding of sampling methods.

Tukey also developed statistical properties of certain polynomials in data, the so-called "polykays". They are generalizations of the sample mean and variance and are useful when the data are not close to normal.

Robust techniques. Tukey had become interested in the problem of robustness in the last stages of his work at the FRCO. The distributions turning up in the scatter of measurements appeared to be longer-tailed than the normal. This began a continuing interest in appropriate distributions to model long-tailed data and a consideration of the properties of estimates such as trimmed means.

Time series analysis. Through his work at Bell Labs Tukey got interested in the frequency analysis of time series. Hamming and he introduced a viable family of estimates of the spectral density of a time series and Tukey developed their statistical properties leading to improved estimates. He also introduced the method of complex demodulation that has proved so useful for the frequency analysis of nonstationary time series.

Confirmatory data analysis (CDA). Tukey named the jackknife procedure and introduced it as a general estimate of variance. It provides an indication of the uncertainty of an estimate by judiciously combining estimates based on subsets of the full data set. Tukey was also a strong proponent of the use of randomization distributions in obtaining p-values and confidence intervals.

Exploratory data analysis (EDA). After the FFT Tukey is probably most widely known for re-energizing descriptive statistics. He created the field of EDA. His 1962 paper, [7] changed the language and paradigm of statistics. The work relied in part on using statistics that were not unduly influenced by outlying observations.

Statistical graphics. Many of Tukey's graphical techniques were developed for EDA. His boxplots and stem-and-leaf diagrams now appear throughout scientific presentations and high school texts. The dynamic graphics methods, such as PRIM-9, allowed examination of moderate dimensional multivariate data sets.

ANOVA and regression. The methods of these fields are the workhorses of statistics. Tukey made many important contributions to each, partly through an emphasis on the consideration of residuals. An assumption of additivity is basic in many developments. Tukey was concerned with how to examine that assumption in practice and developed the one degree of freedom for nonadditivity statistic. Further he contributed to fractional replication and components of variance analysis.

Multiple comparisons. Tukey struggled with the problem of how to control the error rate when many inferential statements are being made. The difficulties arise if one gives into temptations associated with carrying out several analyses of the same data set. Tukey proposed a method based on the studentized range. This is popularly known as the "honestly significant difference" (HSD) method.

Multivariate. Among the ideas and techniques that Tukey proposed are: Tukey depth, Tukey median, projection pursuit, dyadic ANOVA, and the bagplot.

Other statistical topics. Among further areas that Tukey contributed to are: expert systems, fiducial probability, medical statistics, nonparametrics, path coefficients, propagation of error, randomization, and smoothing.

Information sciences

Starting in the mid-sixties John Tukey sought to bring order to the literature of statistics and probability by constructing indices of the papers of those fields. He had done extensive work previously for *Mathematical Reviews* and prepared bibliographies, e.g. for time series. In particular he constructed a citation index, the first outside of the legal profession. The various *Indices* he constructed are listed below.

Engineering

As World War II was ending Tukey designed the electronic adding circuit used in the von Neumann computer at the Institute for Advanced Studies, Princeton. His name is also on a number of patents. His work on the U2 has already been mentioned. He was an early worker in the field of acceptance sampling.

Tukey's scientific contributions continued right up until his death.

RECOGNITION

Tukey received many important honors. He was awarded seven honorary doctorates. The universities involved were: Case Institute of Technology, Brown University, Yale University, University of Chicago, Temple University, Princeton University, University of Waterloo.

Further he was: Guggenheim Fellow, Wald Lecturer (IMS), Member (National Academy of Sciences), Member (American Philosophical Society), Member (American Academy of Arts and Sciences), S. S. Wilks Medalist (ASA), Fisher Lecturer (COPSS), National Medal of Science Recipient, Hitchcock Lecturer (University of California, Berkeley), Scott Lecturer (Cambridge University, England), Shewhart

Medalist (American Society for Quality Control), Medal of Honor recipient (IEEE), Deming Medalist (American Society for Quality Control), James Madison Medalist (Princeton University), John von Neumann Lecturer (SIAM), Foreign Member (Royal Society of London), Monie A. Ferst Award Winner (Sigma Xi).

Throughout his career Tukey served in various capacities as an officer in professional societies: AMS Council, Vice-President ASA, Vice-President SIAM, President Institute of Mathematical Statistics.

CONCLUDING REMARKS

Tukey had scientific impact both by writings and in person, the latter perhaps being the most important. He had many students and seemed to enjoy teaching a great deal. The doctoral students are listed in [3]. An important part of his students' education was the Applied Statistics Seminar. The speakers were told that they did not need to have a solution to the problem. The participants were encouraged to speak up even when unsure of the correctness of what they had to say. He liked to argue saying that that was the best way to get the issues onto the table quickly.

Tukey was famous for his neologisms. A few are: ANOVA, bit, boxplot, cepstrum, hamming, jackknife, linear programming, stem-and-leaf.

There are interviews: [2], [1], [5], [6]. There are memorial articles in the December 2002 number of *The Annals of Statistics*. That issue also contains a list of Tukey's publications. He had many, many co-authors perhaps, in part, because he liked bouncing ideas off others. Tukey's books have gone to Brown University and his papers to the American Philosophical Society.

Acknowledgements. The author's work was carried out with the support of NSF grant DMS-02-03921.

David R. Brillinger

Statistics Department
University of California
Berkeley, CA 94720-3860, USA

brill@stat.berkeley.edu

References

[1] Anscombe, F. J. (1988). Frederick Mosteller and John W. Tukey: a conversation. *Statistical Science* 3, 136-144.

[2] Aspray, A. and Tucker, A. (1985). Interview of John Tukey. *The Princeton Mathematics Community in the 1930s*. Transcript Number 41. (PMC41). Princeton University.

[3] Brillinger, D. R. (2002). John Wilder Tukey (1915-2000). *Notices of the American Mathematical Society* February, 193-201.

[4] Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. Computation* 19, 297-301.

[5] Fernholz, L. T. and Morgenthaler, S. (1997). A conversation with John W. Tukey. Pp. 26-45 in *The Practice of Data Analysis* (Eds. D. R. Brillinger, L. T. Fernholz and S. Morgenthaler). Princeton University Press, Princeton.

[6] Fernholz, L. T. and Morgenthaler, S. (2000). A conversation with John W. Tukey and Elizabeth Tukey. *Statistical Science* 15, 79-94.

[7] Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Statist.* 33, 1-67.

Further Reading

The Collected Works

Tukey, J. W. (1984, 1985, 1986, 1988, 1990, 1992, 1994). *The Collected Works* Vol. I-II *Time Series*, III-IV *Philosophy and Methods of Data Analysis*, V *Graphics*, VI *More Mathematical*, VII *Factorial and ANOVA*, VIII *Multiple Comparisons.* Chapman & Hall, New York.

The Indices to Probability and Statistics

Dolby, J. L., Ross, I.C. and Tukey, J. W. (1973, 1973, 1975, 1973). *Index to Statistics and Probability*, Vol. 1 *The Statistics Cumindex*, 2 *Citation Index*, 3-4 *Permuted Title*, 5 *Locations and Authors.* R and D Press, Los Altos.

Other Books

Andrews, D. F., Bickel, P., Hampel, F. R., Huber, P., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location.* Princeton U. Press, Princeton.

Basford, K. E. and Tukey, J. W. (1998). *Graphical Analysis of Multiresponse Data.* Chapman & Hall, London.

Blackman, R. B. and Tukey, J. W. (1959). *The Measurement of Power Spectra from the Point of View of Communications Engineering.* Dover, New York.

Brillinger, D. R., Jones, L. V. and Tukey, J. W. (1978). *The Management of Weather Resources II: The Role of Statistics.* GPO, Washington, DC.

Cochran, W. G., Mosteller, F., and Tukey, J. W. (1954). *Statistical Problems of the Kinsey Report on Sexual Behavior in the Human Male.* ASA, Washington, DC.

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

Tukey, J. W. (1940). *Convergence and Uniformity in Topology.* Annals of Mathematics Studies, Number 2. Princeton University Press, Princeton, N. J.

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

Related Entries: Computer intensive statistical methods (Update 1); Configural polysampling (Update 1); Election predictions; Environmental statistics; Fisher's k-statistics; Folded transformations; Freeman-Tukey test; G- and h- distributions; Gapping; Graphical representation; Graphical representation of data; Hanning; Interactive data analysis; Jackknife methods; Kinematic displays; Median estimation, inverse; Median estimates and sign tests; Military statistics; Monte Carlo methods; Monte Carlo swindle; Multiple comparisons; Multiple time series; Multivariate graphics; Nonparametric tolerance limits; Notched box and whisker plot; One-wild distribution; Peeling data; Polykays; Problem solving in statistics; Quality control, statistical; Randomized blocks; Regression coefficients; Regressograms; Residuals; Resistant techniques; Robust estimation; Robust regression; Robust-resistant line; Robust smoothing; Robustification and robust substitutes; Siegel-Tukey test; Slash distribution; Spectral analysis; Spectral density; Spectral distribution; Spectral estimation; Spectral function; Stem-and-Leaf display; Studentized range; Student's t-tests; Studentized range; Tolerance regions, statistical; Trend test; Trimmed and Winsorized means, tests for; Trimming and Winsorization; Tukey-Kramer intervals; Tukey's confidence interval for location; Tukey's g- and h-distributions; Tukey's hanging rootogram; Tukey's inequality for optimal weights; Tukey's line; Tukey's median polish; Tukey's quick test; Tukey's test for nonadditivity; Tukey's test for ordered alternatives; Twicing; Weather modification.