JOHN WILDER TUKEY

16 June 1915 — 26 July 2000

Biogr. Mems Fell. R. Soc. Lond. **49**, 000–000 (2003)

John W Tukey

JOHN WILDER TUKEY

16 June 1915 — 26 July 2000

Elected ForMemRS 1991

BY PETER MCCULLAGH FRS

University of Chicago, 5734 University Avenue, Chicago, IL 60637, USA

John Wilder Tukey was a scientific generalist, a chemist by undergraduate training, a topologist by graduate training, an environmentalist by his work on Federal Government panels, a consultant to US corporations, a data analyst who revolutionized signal processing in the 1960s, and a statistician who initiated grand programmes whose effects on statistical practice are as much cultural as they are specific. He had a prodigious knowledge of the physical sciences, legendary calculating skills, an unusually sharp and creative mind, and enormous energy. He invented neologisms at every opportunity, among which the best known are 'bit' for binary digit, and 'software' by contrast with hardware, both products of his long association with Bell Telephone Labs. Among his legacies are the fast Fourier transformation, one degree of freedom for non-additivity, statistical allowances for multiple comparisons, various contributions to exploratory data analysis and graphical presentation of data, and the jackknife as a general method for variance estimation. He popularized spectrum analysis as a way of studying stationary time series, he promoted exploratory data analysis at a time when the subject was not academically respectable, and he initiated a crusade for robust or outlier-resistant methods in statistical computation. He was for many years a scientific adviser to Presidents Eisenhower, Kennedy, Johnson and Nixon. A 1965 report he wrote was the impetus leading to Congressional action that created the Environmental Protection Agency, and a later 1976 report (4)* confirmed the threat of halocarbons to stratospheric ozone. His work for the State Department on the Nuclear Weapons Test Ban Treaty led him to develop data-analytic tools to distinguish explosions from earthquakes. Evidence of his influence can be seen in a wide areas of science and technology, from oceanography to seismology, from topology to sampling and statistical graphics. Among many honours, he was awarded the S.S. Wilks award of the American Statistical Association in 1965, the Medal of Honour of the Institute of Electronic and Electrical Engineers in 1982, and the US National Medal of Science in 1973.

* Numbers in this form refer to the bibliography at the end of the text.

FAMILY AND EDUCATION

John Wilder Tukey was born in New Bedford, Massachusetts, the only child of Ralph H. and Adah M. Tukey, both high-school teachers trained in the classics. Recognizing John as a prodigy, they educated him largely at home until he was admitted to Brown University in 1933. This was possible partly because John's mother was, as a married woman, not allowed to hold a position as a full-time teacher in Massachusetts. Their educational method was to respond to John's queries by providing clues and asking further questions rather than giving a direct answer, a characteristic that John inherited and used throughout his career. By his own account, New Bedford had a wonderful public library, and to a great extent John's early education took place there. Not only did the library have the *Journal of the American Chemical Society*, but it also had the *Transactions of the American Mathematical Society*. In a 1995 interview on the occasion of his 80th birthday, John remarked that the reason he went to Brown as a chemist was that he could read the *JACS* but he could not read the *Transactions*.

John took his bachelor's degree in chemistry at Brown in 1936, and he stayed on for a further year for his master's degree. In 1937, he moved to Princeton, enrolling initially as a graduate student in chemistry. He was discouraged at Princeton to find that he could not serve as a laboratory assistant for undergraduate physical chemistry, as he had done at Brown, until he had his PhD. During his first year he attended courses in mathematics as well as chemistry. At the end of his first year he took the PhD qualifying examination, but in mathematics rather than chemistry. After a further year, his doctoral thesis, 'On denumerability in topology', was completed under the supervision of S. Lefschetz (ForMemRS 1961). It was published in the series *Annals of mathematical studies* under the title *Convergence and uniformity in topology* by Princeton University Press, and John was appointed Instructor in Mathematics in 1939.

His friends and fellow students at Princeton were an illustrious group, including L. Spitzer (ForMemRS 1990), F. Smithies, R.P. Boas, R.P. Feynman (ForMemRS 1965), O. Morgenstern and N. Steenrod. In his autobiography, Feynman (1988, p. 58) recounts an incident as a student at the Graduate College, involving the perception of the passage of time. Feynman discovered that he could keep track of time while reading, but he could not do likewise while typing or talking. The next day, he gave a demonstration to his fellow students. By Feynman's account, Tukey was amazed by this, but he could not understand why Feynman could not keep track of time while speaking. A quick demonstration showed that Tukey had no difficulty with this, but he could not do the same while reading. It turned out that while Feynman kept track of time by counting, Tukey kept track of time by visualizing the passage of a numbered tape, and that he was 'looking' at this tape as he spoke. This incident may be the origin of Tukey's oft-repeated remark that 'people are different'. Many others have subsequently remarked on John's ability to perform multiple tasks simultaneously, such as compiling the *Current index to statistics* and listening to a seminar at the same time. He occasionally referred to the second task as 'his knitting' implying that it occupied minimal CPU brain cycles and did not intrude on his other work.

Fred Mosteller recounts that when he first met John in 1939, all of his work had been in abstract mathematics and he had done no statistics. In 1941, in *American men of science* Tukey lists his interests as point set topology and analysis. The war changed all that. In 1941, Tukey went to work for Merrill Flood in Fire Control Research, meaning gun and artillery control, range finding and so on. There he met Charlie Winsor, an engineer turned statistician. Their friendship was deep, and Tukey gives Winsor the credit for his conversion from pure mathematician to statistician and scientist. Several of Tukey's papers, and the book *Exploratory data analysis*, are dedicated to Charles Winsor. Others credit Tukey's conversion in large part to George W. Brown, a colleague at Fire Control Research.

In 1950 John married Elizabeth Louise Rapp. Before their marriage Elizabeth was Personnel Director of the Educational Testing Service in Princeton. Afterwards, she developed a career as an antiques dealer and appraiser. Elizabeth reports in an interview (Fernholz & Morgenthaler 2000) that her family was shocked when they first met John, because he was so unconventional or eccentric. His informal attire left a lasting impression with her family, but she reasoned that his departures from social norms were not much above average for a New Englander. Elizabeth frequently accompanied John on his foreign travels, using the opportunity to further her antiques business. She died on 6 January 1998 after a long illness. They had no children. Elizabeth's sister Phyllis married Frank Anscombe, then a junior professor at Princeton, later Professor of Statistics at Yale. Anscombe and Tukey collaborated on several papers, the best known being one on residuals in 1963. The four Anscombe children were frequent visitors to the Tukey house.

Tukey was a workaholic, but he was also an avid reader of science fiction and mystery novels of the trashy sort. He could read and process information extremely rapidly, and he claimed to be able to read one novel in an hour. At the time of his death, an entire room of the house in Princeton was filled with thousands of these paperbacks.

First impressions

My first recollection of Tukey comes from a 1977 visit to London, when Tukey was invited to give a seminar at Imperial College. I was a graduate student at the time, and Tukey was well known to all the students, if only by his colossal reputation. A sense of excitement and curiosity was detectable among the staff and graduate students because this was not an ordinary seminar. An extra-large parade of local luminaries occupied the front row, and the graduate students as usual were safely ensconced in the rear. After his introduction, Tukey ambled to the podium, a great bear of a man dressed in baggy pants and a black knitted shirt. These might once have been a matching pair, but the vintage was such that it was hard to tell. An array of coloured pens bulged from his shirt pocket. These, I later learned, are essential tools for data analysis Tukey style.

Carefully and deliberately, a list of headings was chalked on the blackboard. The words came too, not many, like overweight parcels, delivered at a slow unfaltering pace. For the most part, the words were familiar individually, but as phrases they seemed strangely obscure. What did they mean? Was it English? I had no idea, but I was only a graduate student. Surely someone must know, someone in the front row. The list was not long, but seemed to take a long time to write. When it was complete, Tukey turned to face the audience and the podium, a long desk of the type used for demonstrating chemistry or physics experiments. 'Comments, queries, suggestions?' he asked the audience, each word seeming to take a full minute to deliver. As he waited for a response, he clambered onto the podium and manoeuvred until he was sitting cross-legged facing the audience. This activity must have taken a full minute, but there was still no response. As he sat there in a perfectly relaxed Buddha pose, it became apparent that Tukey was in no hurry to deliver his message, whatever it was. He seemed no more ill at ease than if he were at the beach or a baseball game. We in the audience sat like spectators at the zoo waiting for the great bear to move or say something. But the great bear appeared to be doing the same thing, and the feeling was not comfortable. How long could this go on, we wondered. After a long while, as if to confirm the position, he extracted from his pocket a bag of dried prunes and proceeded to eat them in silence, one by one. The war of nerves continued … four prunes, five prunes.… How many prunes would it take to end the silence? The situation demanded leadership, so we looked anxiously to the front row for relief. Still no response … eight prunes, nine prunes.… Several prunes later he had had enough, so he passed the remainder in silence to the would-be audience, now spectators. After what seemed like an eternity, someone from the front row asked a safe question: 'John, could you explain what you mean by such and such.' That was all that was required, and the seminar then continued relatively uneventfully with active participation from the front row.

As I later came to understand, anyone who thought that the speaker might cave in first did not know John Tukey. The prunes undoubtedly had a simple dietary explanation, but to some in the audience it seemed that he had instituted his own private classification of seminar audiences, and we were a 12-prune audience.

Whether Tukey's teaching style was simply oblique, as Brillinger suggests, or whether he was deliberately obscure is hard to tell. My impression is that he liked to play games his way to get people to figure out for themselves the things that he already knew. More than anything else, he liked the give and take of an argument, but he also expected his views to prevail, and they usually did. Whatever the explanation, he was much more successful with individual students than in seminars or formal lectures.

War service

Tukey spent a large part of the war at the Fire Control Research Office in Princeton. The principal problems there were mathematical, involving ballistics, gun and artillery control, range finding, calculating leads for moving targets and so on. Information on Tukey's wartime activities is scarce, but there are suggestions that he may have been involved in code-breaking. After the war, Bell Labs was involved in defence-related activities connected with the Nike missile defence system, and Tukey's work on ballistics continued there. He also continued to work with the military on various technical problems, including uranium enrichment, details of which are hard to find. It seems that he had a hand in the development of the U2 high-altitude spy aircraft, which monitored Soviet military activities from 1956 to 1960. His expertise was sufficient that he was appointed a US delegate to the 1959 Geneva conference on nuclear disarmament. Throughout his career, the Army Research Office and the Office of Naval Research supported Tukey's research.

Princeton

After completing his PhD, Tukey was appointed Instructor in Mathematics in 1939, then Professor of Mathematics in 1949, and he spent his entire career at Princeton until his retirement in 1985. Throughout his career, he taught courses in statistics and he advised a steady stream of research students on a wide range of statistical topics. The Statistical Techniques Research Group was founded in 1956 with Tukey as director. Other members affiliated with the group included George Box (FRS 1985) and Stuart Hunter.

The cultural differences between statistics and mathematics were such that statistics could not prosper within a department dominated by pure mathematics. Tukey argued that statistics was not a branch of mathematics or applied mathematics, but an integral part of science. Accordingly, a new Department of Statistics was created in 1965 consisting of Wilks, Tukey, Hartigan, and later Geof Watson and Peter Bloomfield. Tukey was chairman for the first four years, followed by Geof Watson. Despite its size, in the 1950s and 1960s Princeton was considered to be the top statistics department in the USA, as judged by the number and quality of its PhDs. Nonetheless, from an administrative standpoint, the department did not grow and prosper, and it was clear by the early 1980s that it was unlikely to survive beyond Tukey's retirement. In the end, the remnants of the department were folded into Civil Engineering in 1985. Watson remained in Mathematics.

Bell Labs

In 1945, Tukey began a long association with Bell Labs, Murray Hill, and for the rest of his career he spent half his time at the Labs and half at Princeton. In addition, for much of his career, it seemed as though he spent a 'third' half of his time in Washington on Government business. This dual-track career suited his style, the Labs proving to be a rich source of challenging problems in data analysis. His colleagues included Walter Shewhart, R.W. Hamming and Claude Shannon (ForMemRS 1991), so he must have found it a stimulating environment. At that time and for many years until the court-ordered break-up of AT&T in 1983, Bell Labs was the premier industrial research group in the USA. Tukey found himself in the middle of one of the most exciting periods of research and development, including the transistor in 1947, the electronic computer, and the laser in 1958. Tukey was one of the contributors to electronic computation, although much of what he did is not recorded. Brillinger (2002*a*,*b*) notes, for example, that John had a major hand in the design of one of the early binary adders.

Tukey's strength was not as a day-to-day organizer or administrator, but he could be effective when he set his mind to these tasks. In 1952 the Department of Statistics and Data Analysis was set up at the Labs under the guidance of Walter Shewhart, Claude Shannon and John Tukey. Milton Terry, Martin Wilk, Ram Gnanadesikan and Colin Mallows were early members who subsequently became head of one of the groups. Much of the early work was concerned with experimental designs for electronic experiments involving Mylar capacitors and transistor diodes.

By the time I arrived at the Labs in 1984, growth by meiosis had transformed the original statistics group into two, one title being a permutation of the other. The group was best known for its strength in computation, graphics and data analysis, all areas in which Tukey had an active interest. Throughout his career, John avoided direct contact with the computer, relying instead on his own formidable computational skills with pen and paper, and frequently resorting to multi-coloured pens. Nevertheless, he had a substantial influence on the field of statistical computing and statistical graphics, including dynamic graphics.

In the mid-1960s, and for many years subsequently, strategies and criteria for statistical computing were a constant theme of discussion at the Labs. Contributing to the discussion were numerous visitors, including Frank Anscombe, David Brillinger, David (later Sir David) Cox (FRS 1973), Michael Godfrey, Michael Healy and John Chambers. Much of the discussion centred on flexible systems, and the influence of that discussion can be seen in the way that statistical software, notably the S system, has subsequently developed.

Tukey acted as a consultant to a large number of US corporations throughout his career, even after his retirement in 1985. The range is quite astonishing: Educational Testing Service, Merck Sharp and Dohme, NBC (for whom he did election-night forecasting), Xerox and Bellcore. He had a hand in at least nine patents on information retrieval awarded to Xerox in the 1990s. During much of this time he also worked with the military, the Bureau of the Census, the National Bureau of Standards and the Scripps Oceanographic Research Institute.

Research style

Despite the abstract nature of his PhD research, John's mode of thinking on statistical matters is best described as eclectic, pragmatic, algorithmic and idiosyncratic. Although he was not averse to the use of formal statistical models by others, we seldom find explicit models in his own work where serious applications are involved. Right from the earliest papers on one degree of freedom for non-additivity, the emphasis is on algorithms and strategies for the most convenient way of arranging the computations and presenting the results. The emphasis on algorithms was not a passing phase, because we see it in his seminal work on spectrum analysis, in the fast Fourier transformation in the 1960s, his work on exploratory data analysis and graphics, and his work on robustness. In the 1950s and 1960s, the predominant style of published statistical research in the USA as exemplified by S. Wilks, A. Wald, J. Neyman (ForMemRS 1979), E. Lehmann, J. Kiefer and others was very mathematical, so John's style set him apart as a maverick. He was a man with a mission to change the profession, and he seldom seemed concerned about what others might think of his work.

The most visible progress in mathematics takes place through abstraction, axiomatization, precise definitions and deductive reasoning. Statistical theory uses a wide range of mathematical concepts, so it is common for statisticians with mathematical training to employ a similar mode of deductive reasoning in applied work. Tukey believed passionately that deductive reasoning with precisely defined concepts is not the path to progress in the sciences, an opinion that was not always well received by his mathematical colleagues. He very much resisted expressing his ideas or justifying his procedures in formal terms such as a statistical model, possibly because of the fear that people might come to believe in the model as though it were a scientific fact. On occasion, he might agree temporarily to proceed as though a certain model were true, but invariably he wanted to justify his procedures by being reasonably good over a wide range of scenarios. He was a genuinely original thinker, almost to the point of professional eccentricity, and he felt no obligation to conform to standard practice or standard terminology. On the contrary, he felt that his obligation as a scientist lay in overturning standard practices by reminding his fellow statisticians of true but unpleasant facts. In short, he was a constructive scientific anarchist.

In 1984, Bill Cleveland persuaded Wadsworth to publish the collected works of John Tukey while the author was still alive and active and in a position to put the papers in their historical perspective. The volumes are arranged by subject matter, each with its own editor. Eight volumes were published in this series (table 1), the last appearing in 1994. In addition to the published papers, these volumes also include numerous previously unpublished papers and book manuscripts. The list of these volumes gives a good idea of the breadth of Tukey's statistical work. Even so, entire areas such as robustness are not included.

Table 1. Collected works of John W. Tukey

title	year	editor
I. Time series	1984	D. Brillinger
II. Time series 1965–1984	1985	D. Brillinger
III. Philosophy and principles of data analysis	1984	L.V. Jones
IV. Philosophy and principles of data analysis	1986	L.V. Jones
V. Graphics	1988	W.S. Cleveland
VI. More mathematical	1990	C.L. Mallows
VII. Factorial and ANOVA	1992	D.R. Cox
VII. Multiple comparisons 1948–1963	1994	H.I. Braun

[AUTHOR: two titles with the same name?]

Published research

Introduction

Tukey's early papers, covering roughly the 20-year period up to 1960, cover a wide range of topics in statistical theory and applications. Even in his theoretical work we find an acute awareness of what is needed in applications. In his paper on partition models, for example, applications to polymer chemistry and bacterial mutation are discussed. Other topics covered in this period include a series of papers on non-parametric estimation, including confidence intervals and tolerance intervals for prediction, various papers on distribution theory, and a paper on transformations.

In view of the fundamental nature of his PhD work in topology, it is surprising that Tukey did not set out to establish a firm foundation for formal statistical inference, a theory designed to meet the needs of applied work. In the 1950s and early 1960s many applied statisticians had become convinced that the rigid Neyman–Pearson–Wald decision theory was unsuitable as a foundation for statistical inference. At about the same time, the Bayesian revival began, and one rigid theory began to supplant the other. It is clear that John did give a great deal of thought to the matter over a long period, but by 1984 he had come to regard 'belief in a unified structure for inference as a dangerous form of hubris'. Tukey's attitude in methodological matters was pragmatic and eclectic. He could tolerate formal statistical models in the work of others but he worked hard to avoid them in his own applied work. He made extensive use of confidence intervals, not because he regarded the associated formal theory as an adequate basis for a theory of inference, but because they seemed to suit his pragmatic style. He made use of formal probability models, although frequently expressed in his own gnomic terms, but he found little use for decision theory or formal Bayesian inference. Nonetheless, some of the techniques that he used in election forecasting, particularly the idea of 'borrowing strength' from similar data elsewhere, are most easily described and justified in Bayesian terms.

The major themes from the early years are sampling theory, analysis of variance, time series, fiducial inference, and multiple comparisons.

Sampling theory

Tukey had a strong admiration for the work of R.A. (later Sir Ronald) Fisher FRS, so much so that in 1950 he arranged for a volume of Fisher's selected papers to be published by Wiley. In characteristic Tukey fashion, the selection was such that round numbers were assigned to the 'big' papers as Tukey saw them. Occupying the number 20 spot was Fisher's 1929 paper, which replaced the problem of estimating sample moments with the more natural problem of estimating sample cumulants. If ever there was an algorithmic paper, this was it, and its panache and style must have appealed to the young Tukey.

Fisher's k-statistic of degree r is a symmetric homogeneous polynomial k_{rn} of degree r in $n \geqslant r$ variables, also called the sample cumulant. Among symmetric functions, k_{rn} is the unique unbiased estimate of the population cumulant κ_r. The first two are the sample mean and the sample variance. Tukey made the key observation that if $(x_{i_1},\ldots,x_{i_n})$ is a simple random sample selected without replacement from a finite population $(x_1,\ldots,x_N)$,

$$\text{ave}\, k_{rn}(x_{i_1},\ldots,x_{i_n}) = (k_{rN}(x_1,\ldots,x_N),$$

where the average is taken uniformly over the $\binom{N}{n}$ distinct samples. Statistics having this property were said to be inherited on the average. Inheritance is thus a way of associating with each integer $r \geqslant 1$ a natural sequence of polynomial symmetric functions $\{k_{rn}\}$ in which k_{rn} is a function of $n \geqslant r$ variables. Inheritance is a fundamental idea because, without it, no logical relation exists between the symmetric functions k_{24} and k_{25}, for example. Having made this observation, it becomes apparent that there must be two such statistics of degree two, three of degree three, five of degree four, and so on. More correctly, the set of homogeneous polynomial symmetric functions of degree r having the inheritance property is a vector space of degree $p(r)$ where $p(\cdot)$ is the number of partitions of r. The most natural basis for this space is the set of functions termed polykays, these being generalizations of Fisher's k-statistics. In his 1956 paper, Tukey extends these ideas to two-way arrays in which the sub-sample operation acts independently on rows and columns.

Looking back at the development of the theory of sample moments over the past century, it is striking that after Fisher's paper, the entire outlook was changed, and the same is true once again of Tukey's work. To a certain extent, Tukey's paper can be said to have killed the field. As Speed (2002) has remarked, there have been virtually no subsequent papers on the topic in the past 50 years.

Analysis of variance

The term 'analysis of variance' refers to the Fourier-style decomposition of the total sum of squares into components. For a randomized blocks design or a two-factor design with equal replication, the components are the row sum of squares, the column sum of squares, the interaction sum of squares, and the replicate sum of squares. In certain circumstances, these may be further decomposed. The interpretation of the various sums of squares is very much dependent on context, but generally speaking the interpretation is likely to be simpler if the interaction mean square is comparable to the replicate mean square. When this occurs, we say that the data are additive. In many cases, however, no replicate observations are available, so there is no replicate sum of squares available for direct comparison. Nonetheless, Tukey showed how one degree of freedom could be isolated from the interaction sum of squares and used to detect a type of non-additivity removable by transformation.

Tukey's original description was presented in algorithmic fashion for two-way designs and later for Latin-square designs. Reading through these computational recipes, it is hard to believe that Tukey did not have a full grasp of the most general version of the test, which is to regress the residuals on the square of the fitted values under the additive model. A computer version of the scheme is as follows. (i) Fit the additive model $E(Y) = X\beta$ and compute the vector of fitted values $\hat{Y} = X\hat{\beta}$. (ii) Fit the extended model $E(Y) = X\beta + \gamma\hat{Y}^2$ with the squares of the fitted values as an additional covariate. The reduction in residual sum of squares is the one degree of freedom for non-additivity (1DOFNA), and the standard t-test using $\hat{\gamma}$ is the appropriate test statistic. More importantly, if non-additivity is found, the suggested remedy is to transform the data $Y \mapsto Y^\lambda$, where the data-suggested value $\lambda = 1 - 2\hat{\gamma}\bar{Y}$ serves as a rough indicator.

This 1DOFNA scheme has several hallmarks characteristic of Tukey's most successful work: it is based on a simple idea, it is algorithmic and easily computed, and it has had a substantial impact on applied work.

Fiducial inference

The papers on fiducial inference are Tukey's only contributions to what might be called formal theories of statistical inference. Tukey's attempt to understand Fisher's ideas on the subject led him to tackle the problem of the Nile, first posed by Fisher in his Harvard tercentenary lecture in 1936. Given a family of probability distributions $\{F_\theta\}$ on the same space S, the problem of the Nile is to find a set $A \subset S$ such that A has probability α under each distribution in the family. It was known from the work of Linnik that such regions do not exist in the Behrens–Fisher problem. Tukey showed that, if the family of distributions is finite and each distribution F_θ is non-atomic, the desired regions always exist, and the problem of the Nile is thus solvable. In statistical terminology, similar regions and similar test statistics exist whenever the parameter space is finite.

Fisher's theory of fiducial inference was a natural target for study because Tukey had a high regard for Fisher's work, and the theory of fiducial inference was not well received by the statistics community. Perhaps influenced by his pure mathematical training, Tukey tackled the uniqueness of Fisher's construction rather than its interpretation. In his 1956 paper, he describes a two-parameter twisted normal example that demonstrates that the fiducial construction is not unique for two-parameter models. In the published volume of Fisher's correspondence (Bennett 1990) the correspondence between Fisher and Tukey makes interesting reading.

In 1955, Tukey sent Fisher a draft of his paper, exploring through examples the matter of non-uniqueness of fiducial probability. In the opening sentence of his reply, Fisher refers to the paper as 'your long screed about probability statements concerning parameters…'. And later in the letter, Fisher's fraying temper becomes even more apparent in the remark 'This of course throws no doubt on the verification of any probability statement by observed frequency, if the appropriate frequency is observable, as you would see if you could ever get your bull-headed mind to stop and think'. Fisher was notorious for his intemperate remarks, and no doubt Tukey was aware of this, but the correspondence continued, Tukey remaining polite but insistent. In 1956, Tukey visited Fisher at his office in Cambridge to discuss fiducial probability, hoping perhaps that a face-to-face discussion would be easier than written correspondence. The visit was a notable disaster, an unstoppable force meeting an immovable object. By one account, Fisher blew a fuse and threw Tukey out of the office.

This incident marked the end of their correspondence, and the end of Tukey's interest in fiducial inference. Given Tukey's admiration for Fisher's other work, it must have been a hurtful rejection.

Time series

Tukey's interest in time series began at Bell Labs with a problem posed by H.T. Budenson in radar tracking, which involved calculating the power spectrum for an observed series. A paper written jointly with R.W. Hamming entitled 'Measuring noise color' followed, and Tukey, simultaneously with M. Bartlett (FRS 1961), became convinced that spectrum methods were more useful and informative than covariance methods for stationary time series. The concept of aliasing was developed at about this time, as were refined smoothing methods. A bewildering variety of problems in all sorts of applications, from airplane dynamics to seismology and oceanography were tackled by using power spectra. Many of the papers were published as Bell Labs Technical Memoranda: some remained unpublished until the appearance of the collected works in 1985.

One of my favourite Tukey papers on time series, no. 12 in the collected works, was published in 1963, jointly with Bruce Bogert and Michael Healy. The title, 'The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking', is self-explanatory, at least to an expert in crossword puzzles. Echoes occur commonly in certain types of acoustical noise, but this paper was motivated by echoes in seismological series. The key technical idea could not be simpler, but the insight and consequences are remarkable. The effect on the power spectrum of a simple echo in the signal is to add to the log spectrum an approximate cosinusoidal ripple whose frequency is the echo delay. Now, the spectrum is a function of frequency, so the echo delay shows up as the frequency of the frequency, i.e. a time. Duality of this sort is a mathematician's delight, and Tukey's creative hand is evident in the subsequent wordplay. Because its purpose is to detect echoes, the spectrum of the log spectrum is dubbed the cepstrum. By the same reasoning, the cepstrum must have a quefrency and a gamnitude, as well as a saphe at the origin. More complicated echoes are called rahmonics. Saphe cracking, a characteristic Tukey coinage, refers to the analysis of the saphe (phase) of the cepstrum. The paper concludes with a complete table of twelve dual terms, familiar terms from spectrum analysis, with dual echoes for cepstrum alanysis. Thus low-pass filter becomes short-pass lifter. This blizzard of neologisms led Hamming to refer to the author as J.W. Cutey.

In some of his writings, Tukey used the term 'spectral analysis' rather than spectrum analysis, but he later regretted this usage, expressing the hope that 'spectral analysis', 'power spectral analysis' would become ghosts. His wish has not been granted. Of the 100 titles in the *Current index to statistics* during the 1990s where one or other term occurs, only 13 used Tukey's preferred term.

Tukey's interest in time series continued throughout his career. Undoubtedly his most influential contribution was the promotion of the spectrum as a viable, useful and informative description of the sources of variability in time series. His most celebrated contribution was his invention of the algorithm for the fast Fourier transformation, published jointly with James Cooley in 1965. The algorithm, which had the effect of reducing the computational time from $O(n^2)$ to $O(n \log n)$ for a series of length n, created a revolution in signal processing. With the computing power available in the mid-1960s, the time taken to compute the spectrum of a series of $n = 8192$ points was reduced from half an hour to eight seconds. The fact that all of the ideas behind the fast Fourier transformation had been around in various forms for several decades at least does not in any way diminish the impact that the Cooley–Tukey algorithm has had. Rabiner & Rader (1972) describe it as 'the major event of the decade in signal processing'. Up to that point, much of signal processing was done electronically by analogue devices; afterwards all was digital.

Multiple comparisons

The theory of confidence intervals can be regarded as a theory designed to control error rates. Thus a statement with a 95% confidence interval has a 5% error rate, at least when the assumptions are met. In certain types of experiment, however, there might be multiple statements of similar type, all addressing essentially the same hypothesis. Although the error rate is 5% for each statement individually, the overall error rate might be considerably larger. In the early 1950s, Tukey set out to devise a scheme to produce an honest 5% error rate for a set of pairwise comparisons of batch means.

Over a period of two to three years he drafted a 300-page manuscript with the intention of publishing it in book form. The manuscript was widely circulated and became very influential, but the book did not appear, largely because of Shewhart's advice that he should not publish until his ideas on the matter had crystallized. Tukey later remarked that this was the one occasion where Shewhart offered bad advice. In essence, Tukey invented the subject of multiple comparisons or simultaneous inference, and solved the major problems in this manuscript (5). Various bits and pieces of the manuscript were published in modified form over the subsequent 15 years. The subject has once again come to prominence under the title 'false discovery rate' in biological applications connected with micro-array experiments.

Data analysis

> For a long time I have thought that I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. And when I have pondered about why such techniques as the spectrum analysis of time series have proved so useful, it has become clear that their 'dealing with fluctuations' aspects are, in many circumstances, of lesser importance than the aspects that would already have been required to deal effectively with the simpler case of very extensive data, where fluctuations would no longer be a problem. All in all, I have come to feel that my central interest is in *data analysis*, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make the analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.

With this opening salvo, Tukey began his 1962 paper 'The future of data analysis', a challenge to the mainstream statistical community, declaring data analysis as a field of study in its own right. Tukey relished the grand gesture, and it is no accident that his paper reads in part like the 1776 Declaration of Independence. Optimization, and the misplaced confidence that flows from exact answers to precisely formulated mathematical decision problems, when taken seriously, were seen as great dangers to data analysis and to science. Certainly there was good reason to be critical of the direction in which mathematical statistics was headed in the 1950s and 1960s. The gap between theory and practice had become a chasm, and nothing short of a crusade would help to restore a reasonable balance. Although Tukey's paper was selected by Kotz and Johnson as one of the breakthroughs of the twentieth century, it is seldom read today. Nonetheless, I think it did have the desired effect of redressing the balance over the long term. A conference was held in Madison in 1967 on 'The Future of Statistics', and another in Edmonton in 1974 on 'Directions for Mathematical Statistics'. These conferences caused statisticians to sit back and look at the larger picture, and I think they did have the effect of keeping statisticians more closely in touch with data collection and with science.

Exploratory data analysis and graphical methods

Tukey's credo in promoting exploratory methods of data analysis is summarized by his remark that the best single device for suggesting, and at times answering, questions beyond those originally posed is the graphical display. Graphs and graphical presentation of data have always been an important component of science and statistics, but effective graphical display has always been more of an art than a science. Throughout his career, Tukey sought to develop a set of principles for effective graphical presentation, and in doing so he invented an enormous range of graphical techniques. Some of these are now accepted and standard, such as stem-and-leaf plots and box-and-whiskers plots. Experience with data from a wide variety of sources indicates that transformation—or re-expression, in J.W.T.'s language—statement is almost always beneficial, either for studying distributions or for studying relationships; logs for amounts, logits or flogs for proportions. Some additional principles of good graphical display include: (i) obvious effects should be eliminated before plotting the remainder; (ii) the null pattern should be a straight line, preferably horizontal; (iii) deviations from the null pattern should have simple interpretations;

Much of what goes on in science and in statistics can best be described as exploratory. The rules, if there are any, are hard to discern. By contrast, virtually all of formal statistical theory is confirmatory in style. A hypothesis is formulated, data are collected, and the theory is tested. Modifications are proposed if necessary. Throughout his career, but especially in the 1960s and 1970s, Tukey determined to do for exploratory data analysis what others had done for confirmatory analysis. The campaign was a daunting one, and the fact that any measure of success was achieved is a tribute to Tukey's bull-headed determination, to borrow Fisher's apt phrase.

Exploratory data analysis is an entirely different statistical world, consisting of transformations, graphs, tables, stem-and-leaf plots, box-and-whiskers plots, and other Tukey inventions. It is a world rich in examples, where rules are unclear, and the guiding principles are as likely to be found in psychology as in mathematics or probability. Over the years Tukey amassed a bewildering range of techniques and graphical displays, eventually paring these down to a book-length manuscript. The long-awaited book, *Exploratory data analysis*, culminating the campaign was published in a gaudy orange binding in 1977. This was followed in the same year by another, less idiosyncratic, book *Data analysis and regression*, published jointly with Fred Mosteller.

Robust statistics

As early as the late 1940s, Tukey and Winsor noted that when a Gaussian sample is contaminated by as little as 1% by another Gaussian of the same mean but three times the variance, the efficiency of the sample mean is severely degraded. This observation led Tukey to consider ways of analysing data to protect against outliers. Such methods were called robust against outliers.

The term 'robustness' was introduced into statistics by George Box in the early 1950s in connection with the sensitivity of certain statistics to minor distributional perturbations. Over the past 40 years, the term has come to be associated with the crusade begun by Tukey in about 1960 with the aim of making statisticians more aware of the sensitivity of classical techniques to outliers. Resistant or robust methods were advocated, and robustness was viewed as a property of procedures. The heyday of robust statistics came in the 1970s, beginning with the Princeton robustness study in 1970–71. During that year, a group of visiting statisticians at

Princeton began a systematic study of various compromise estimators of location, estimators designed to behave well over the range from Gaussian to heavy-tailed distributions.

The modest aim of the Princeton robustness study, the estimation of location from a single sample, meant that it could not expect to have a big direct effect on applied work where the structure of the design is seldom so simple. For those not directly involved in the project, it seems odd that so many statisticians should have agreed or been persuaded to devote a full year to this effort. Perhaps it was seen as the prototype for more useful methods to follow, such as robust techniques for regression or the study of dependence (Huber 1981).

As with many large-scale amorphous projects in which Tukey was involved, it is hard to pinpoint a simple, clear and lasting contribution. But he certainly retained a belief in the importance of robust statistical procedures throughout his life. During the year that I spent at Bell Labs in 1984–85, he urged me to 'robustify' generalized linear models. To the best of my knowledge this has not yet been done. Certainly, the robustness study (3) made statisticians more aware of the effect of outliers and the need to examine residuals and look for high-leverage points. To that extent, the level of general statistical practice has been improved.

Other activities

The Kinsey report

After the publication of the first Kinsey report on human sexuality in January 1948, the initial euphoric reception by the public was soon replaced by more critical academic reviews, many questioning the process by which, from samples of convenience, anything useful could be said about the US adult male population. To clarify the matter, in 1950 the National Research Council, which had sponsored Kinsey's research, asked the American Statistical Association (ASA) for advice on problems of research in human sexual behaviour. The members of the ASA committee were W.G. Cochran, F. Mosteller and J.W. Tukey. Their report, *Statistical problems of the Kinsey report on sexual behaviour in the human male*, published in 1954 (1), pinpointed clearly the strengths and weaknesses of the Kinsey study. The weaknesses were, principally, the reliance on volunteers, and the absence of a probability sampling scheme. On the latter point, the report notes that if a respectable approximation of a probability sample is used, the step from sampled population to target population is usually short and the inferences strong. Otherwise, the inference is often tortuous and weak, depending on subject-matter knowledge and intuition, and other barely tangible considerations. These matters deserve to be brought to the reader's attention, and to be discussed, and on this point the Kinsey report failed. The result is that it is impossible to set confident limits on the incidence of the activities under study.

In designing his study, Kinsey made a large number of choices: the choice of orgasm as the *summum bonum* of sexual behaviour, the reliance on volunteers and prisoners convicted of sexual offences, and the reliance on reported behaviour in face-to-face interviews. The ASA report discusses these choices sympathetically, noting that sexual orgasm is not a matter of general quantitative methodology and is thus beyond the scope of the report. The decision to interview rather than use a questionnaire is discussed at length. Kinsey's argument in favour of face-to-face interview, that it permits and encourages variation in the form of the questions to suit the subject, is the same argument used by others against it. The committee members even went so far as to agree to be interviewed, and to have their sex histories recorded, reporting that the interview impressed them as an extraordinarily skilful performance.

Despite the criticism, the final ASA report contained a substantial number of positive remarks, particularly on the matter of comparisons with other studies of a similar sort. Kinsey's study was found to be superior in almost every dimension.

The committee spent five days at the Institute for Sex Research at Indiana University, Bloomington, in October 1950. Public interest in the report was prurient and insatiable, and Kinsey's research support was up for renewal, so the stakes were unusually high for both sides. In certain respects, Kinsey and Tukey were similar characters. Both were stubborn, confident, dominant personalities; both liked nothing better than a good debate; both expected to win. Tensions were high, and the battle was begun, the committee going over Kinsey's report page by page, demanding evidence and proof where called for. As Jones (1997) tells it, Kinsey defended his position 'like a lion' but eventually had to concede on most statistical points. 'Each defeat seemed to deflate him, as though someone had let the air out of his ego' (Jones 1997, p. 643). In a typical exchange of fire, Tukey stated that he would prefer a random sample of three to a Kinsey sample of 300. Despite everything, the veneer of civility was such that the committee was invited to the Kinsey house for dinner followed by a musical programme. Kinsey's ability to charm and win people over to his side was legendary, but his authoritarian style that evening failed miserably. Even the passage of 20 years failed to dim Clara Kinsey's unpleasant memories of the event, for she reported in a 1971 interview, 'I never fed any group of men that I would have so much liked to have poisoned.… Tukey was the worst' (Jones 1997).

Environmental activities

The publication in 1962 of Rachel Carson's eloquent and passionately argued manifesto *Silent spring* (Carson 1962) set the stage for the rancorous environmental debate that followed. Seldom has a book been so violently attacked and its author so vilified by those few who felt their interests threatened, chiefly the US Department of Agriculture and the chemical and food processing industries. In response, the President's Science Advisory Council (PSAC), of which Tukey was a member, created an Environmental Pollution Panel with Tukey as chairman to look into the matter. Their 1965 report, *Restoring the quality of our environment* (2), is a monumental well-reasoned document, largely vindicating Carson's thesis without mentioning it. As Tukey reported in his 1995 interview, the thought that the PSAC report might mention the Carson book had the agriculture folks and other entrenched interests practically weeping in their beer. It is a very modern document, recommending the radical proposal that polluters be required to pay a tax proportional to the damage inflicted, at the same time insisting that no one has a right to pollute.

With such strongly held views on both sides, the emergence of such a balanced report is a tribute to Tukey's stubborn forceful personality and encyclopaedic knowledge of science. When he staked out a position, he invariably had strong supporting arguments, and he always expected to win an argument. The report notes that even without new regulations, much could be achieved by the enforcement of existing regulatory laws. However, if pollution abatement is to be considered, environmental quality standards must first be established. 'Such standards imply that the community is willing to bear the costs or to enforce these costs on others in order to maintain its surroundings at a given level of quality and utility.' The environmental battle of the 1960s was a major turning point, leading to the creation of the Environmental Protection Agency in 1970.

The environmental report looked into such matters as soil contamination from agricultural activities, from pesticides such as DDT, from farm effluent, from detergents, and from mining and other industrial activities. They looked also at municipal and industrial sewage, consumer goods waste, and air pollution from a variety of sources, particularly motor vehicles and factories. The policy of draining wetlands was specifically recognized as a special type of pollution. Climatic effects of pollution were also considered, particularly the increase in atmospheric carbon dioxide. One difficulty that the committee faced was the lack of reliable data. The body of the report, a mere 38 pages long, contains the principal recommendations, a list of principles, a list of recommended actions, studies to be carried out, specific measurements to be made, and recommendations on research programmes. Technical details are contained in 18 appendices.

Tukey chaired two further committees on environmental matters. The PSAC Panel on Chemicals and Health produced its report in 1973. The issue once again was how best to respond to the increasing number of chemicals released into the environment, many possibly benign but a few not. A major theme of the report is that where knowledge is inadequate, the proper response is to seek more knowledge, not either to take drastic action or to do nothing.

In 1975–79, Tukey served as chairman of the National Research Council (NRC) Committee on Impacts of Stratospheric Change. The Molina–Rowland paper (Molina & Rowland 1974), for which the authors won the Nobel Prize in 1995, drew attention to the fact that certain stable aliphatic hydrocarbons, chiefly CF_2Cl_2 and $CFCl_3$ used as aerosol propellants and refrigerants, have the potential to destroy stratospheric ozone by catalysis, so that a single chlorine molecule can destroy many ozone molecules. In response, the NRC committee was convened to report on the magnitude of the likely effects and the uncertainties associated with any prediction. Because of his background as a chemist and statistician, and his ability to make sensible recommendations in the face of great uncertainty and limited information, Tukey was the ideal choice to chair this committee. Direct evidence concerning depletion rates was tenuous at the time, but the task force concluded that fluorocarbon releases to the environment were a legitimate cause for concern. The major concern, then as now, was the depletion of stratospheric ozone and the resulting biological effects associated with the increase in ultraviolet radiation reaching the Earth's surface. Even then, the panel recommended restricting the uses of chlorofluoromethane to the replacement of fluids in existing refrigeration and air-conditioning equipment only. The subsequent finding that the reaction proceeds much more rapidly in polar regions was wholly unexpected but served to emphasize the wisdom of the NRC reports.

Bureau of the Census

In 1989–91 Tukey served on a Special Advisory Panel to advise the Bureau of the Census on the politically charged matter of adjusting the census to accommodate undercount. The statistical community was divided on the question of whether more a more accurate assessment of population numbers could be obtained by means of a post-census survey in selected areas with the intention of estimating the undercount or overcount. Among those who judged this to be technically feasible, some thought it politically inadvisable to broadcast the fact that census figures are subject to statistical adjustment.

Tukey took the view that adjustment was technically feasible and desirable, and he argued his case before Congress. He took the view that one could do better by borrowing strength from similar figures elsewhere, a view that is most easily explained by Bayesian arguments. Tukey, of course, had his own decidedly non-Bayesian justification for this procedure. In the end, political considerations won out, and the adjustment lobby failed to win the argument.

Concluding remarks

Tukey devoted a large part of his energies to five major research themes: spectrum analysis of time series, multiple comparisons, data analysis as a discipline in its own right, robustness, and exploratory data analysis. Of these, his work on spectrum analysis was the most successful, both in terms of the originality of his contributions and the impact that the work has had on applied fields. Although he probably did not invent the concept of simultaneous inference, Tukey's 1953 manuscript (5) established the field and solved most of the major problems. It is not so easy to judge the degree of success achieved in the three remaining campaigns. To do so, we must ask how these initiatives have altered matters for the applied statistician, either by providing computational tools or by greater awareness of pitfalls or by change of approach. Certainly the computational tools available today are much improved since the 1960s or 1970s, and some of that advance can be attributed indirectly to Tukey. In addition, the direction of evolution of the subject may have been altered by the force of Tukey's arguments for data analysis. Exploratory data analysis seemed like a good idea at the time, and still does, but few could claim that the book *Exploratory data analysis* has been a success either as a textbook or as a reference work.

Tukey's major statistical publications, the 1DOFNA paper, the cepstrum paper, the fast Fourier transformation and so on, do not begin to explain the extent of his influence in statistics. At the annual meetings, his dominant personality, sharp intellect and forthright, if sometimes gnomic, remarks made good theatre, and marked him as a formidable discussant who could quickly perceive the crux of the matter. Whether or not one agreed with his approach to applied statistics, his emphasis on exploratory work and robust computation, he was a force that could not easily be ignored. In short, he was a cultural phenomenon, revered by some, feared by others, understood by few.

Acknowledgements

Much of my information on the Tukey family comes from published interviews, particularly the interview on the occasion of John's 80th birthday (Brillinger *et al.* 1997), and a second interview (Fernholz & Morgenthaler 2000). Francis Anscombe, Tukey's nephew, provided additional family information. The first issue of the *Annals of Statistics* in 2002, dedicated to John W. Tukey, contains seven papers and a bibliography. (Benjamini & Braun 2002; Brillinger 2002*b*,*c*; Dempster 2002; Friedman & Steutzle 2002; Huber 2002; Speed 2002; Tukey 2002) discussing various aspects of Tukey's statistical work, plus a list of publications and other information both personal and professional.

A list of publications up to 1993, compiled by Mary Bittrich, his long-time secretary at Bell Labs, provided the starting point for the bibliography. David Brillinger's obituary (Brillinger 2002*a*) is a useful source of information, as is the web site http://cm.bell-labs.com/cm/ms/departments/sia/tukey. The eight volumes of the collected works have been a great help. I have also benefited from conversations with D. Andrews, F. Anscombe, G. Barnard, P. Bloomfield, H. Braun, J.M. Chambers, Sir David Cox FRS, D. Donoho, D.A.S. Fraser, M. Hansen, D. Hoaglin, P. Huber, K. Kafadar, C. Mallows, Lord May PRS, P. Meier, F. Mosteller, J.A. Nelder FRS, D. Pregibon, L. Shepp, S. Stigler and D.L. Wallace.

The frontispiece photograph shows Tukey teaching at Princeton University and was taken by Elizabeth Menzies.

References to other authors

Benjamini, Y. & Braun, H. 2002 John W. Tukey's contributions to multiple comparisons. *Ann. Statist.* **30**, 1576–1594.

Bennett, J.H. (ed.) 1990 *Statistical inference and analysis. Selected correspondence of R.A. Fisher.* Oxford University Press.

Brillinger, D.R. 2002*a* John Wilder Tukey (1915–2000). *Not. Am. Math. Soc.* **49**, 193–201.
Brillinger, D.R. 2002*b* John W. Tukey: his life and professional contributions. *Ann. Statist.* **30**, 1536–1575.
Brillinger, D.R. 2002*c* John W. Tukey's work on time series and spectrum analysis. *Ann. Statist.* **30**, 1595–1618.
Brillinger, D.R., Fernholz, L.T. & Morgenthaler, S. 1997 *The practice of data analysis.* Princeton University Press.
Carson, R. 1962 *Silent spring.* Boston, MA: Houghton-Mifflin.
Dempster, A.P. 2002 John W. Tukey as 'philosopher'. *Ann. Statist.* **30**, 1619–1628.
Fernholz, L. & Morgenthaler, S. 2000 A conversation with John W. Tukey and Elizabeth Tukey. *Statist. Sci.* **15**, 79–94.
Feynman, R.P. 1988 *What do you care what other people think?* Harmondsworth: Penguin.
Friedman, J.H. & Steutzle, W. 2002 John W. Tukey's contributions to robust statistics. *Ann. Statist.* **30**, 1629–1639.
Huber, P. 1981 *Robust statistics.* New York: Wiley.
Huber, P.J. 2002 John W. Tukey's contributions to robust statistics. *Ann. Statist.* **30**, 1640–1648.
Jones, J.H. 1997 *Alfred C. Kinsey*. New York: W.W. Norton.
Molina, M.J. & Rowland, F.S. 1974 Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone. *Nature* **249**, 810–812.
Rabiner, L.R. & Rader, C.M. 1972 *Digital signal processing*. New York: IEEE Press.
Speed, T.P. 2002 John W. Tukey's contributions to analysis of variance. *Ann. Statist.* **30**, 1649–1665.
Tukey, J.W. 2002 The publications and writings of John W. Tukey. *Ann. Statist.* **30**, 1666–1680.

BIBLIOGRAPHY

The following publications are those referred to directly in the text. A full bibliography appears on the accompanying microfiche, numbered as in the second column. A photocopy is available from The Royal Society's Library at cost.

(1) (73) 1954 (With W.G. Cochran & F. Mosteller) *Statistical problems of the Kinsey report*. Washington, DC: American Statistical Association.
(2) (155) 1965 (With 13 others) *Restoring the quality of our environment. Report of the Environmental Pollution Panel, President's Science Advisory Committee.* Washington, DC: Government Printing Office.
(3) (191*a*) 1972 (With D. Andrews, P.J. Bickel, F.R. Hampel, P.J. Huber & W.H. Rogers) *Robust estimation of location: survey and advances*. Princeton University Press.
(4) (213) 1976 *Halocarbons: environmental effects of chlorofluoromethane release*. Washington, DC: National Academy Press.
(5) (415) 1994 The problem of multiple comparisons. *The collected works of John W. Tukey*, vol. VIII (*Multiple comparisons, 1948–1983*), pp. 1–300. New York: Chapman & Hall.