BIOGRAPHICAL MEMOIRS

Robert Reuven Sokal

January 13, 1926–April 9, 2012
Elected to the NAS, 1987

A Biographical Memoir by F. James Rohlf and Douglas Futuyma

Robert (Bob) R. Sokal was a pioneering biologist and biostatistician whose innovative work in numerical taxonomy and biostatistics has left a lasting impact on the scientific community. Born in Vienna, Austria, Sokal's career spanned several decades during which he made significant contributions to the fields of taxonomy, ecology, ecological genetics, human variation, and statistical biology.

Early Life and Education

Bob was born in Vienna, Austria, on January 13, 1926. He remarked in a family memoir that he was not very distinguished academically as a *gymnasium* student and that he expected to become a doctor. The course of his life was changed when he was almost thirteen years old. It was on Kristallnacht in November 1939 when the intensity of Austrian anti-Semitism burst forth and changed the security of a middle-class boy. Regardless of any academic achievements, he and all other Jewish students were required to attend a *hauptschule* (secondary school) rather than a *gymnasium*, because they were not allowed an education that would permit them entry into an institution of higher learning. Bob's father, Siegfried Sokal, who owned paint stores in Vienna and also manufactured paint, was imprisoned in Dachau and later in Buchenwald when the Nazis took over. After innumerable visits to the Gestapo to intercede on his behalf, his mother, Klara Rathner, was able to obtain a decision that his father would be released if she could produce a valid visa and ticket for him to travel to another country. The only place they were able to travel to was Shanghai, China, as it did not require a visa and also because most other countries were now involved in the war. On April 9, 1939, the family took a train with the few possessions they were able to hide to Trieste, Italy, where they boarded a ship that would take them to Shanghai. Their story was published in German in *Letzte Zuflucht Schanghai: die Liebesgeschichte von Robert Reuven Sokal und Julie Chenchu Yang [Final Refuge Shanghai]*.[1]

Figure 1 Robert Reuven Sokal.

In Shanghai, they lived in a refugee area, but Bob was able to attend a British public school and learned not only Chinese but also French and English. He wrote that he rose to the top of his class because he feared that otherwise he would become a manual laborer, a profession for which he felt he was ill-adapted. As a penniless refugee, there was no

possibility of paying tuition, so he had to win scholarships. After graduation in 1943 he attended St. John's University, an American university in Shanghai, and earned a bachelor of science in biology in 1947. While a student, he founded the St. John's Biological Society to bring more life into the department. It was at the society's first meeting that he met Julie Chenchu Yang, who would become his wife. His thesis, upon which his first publication would be based, was on the anatomy of the head of a dragonfly species.

During his final year (1947) in Shanghai, his father was able to return to Vienna and reopen his paint store and find a place to live. It was then possible for his mother to also return, but Bob stayed in Shanghai to finish his degree. To support himself, he became a tutor to high school students and then was hired as a biology teacher at the Shanghai American School.

After completing his undergraduate degree, Bob applied to and was accepted by the University of Chicago's graduate program in zoology. The faculty were impressed that as a refugee he was able to carry out and publish independent research. His doctoral advisor was Alfred Emerson, an ecologist and termite taxonomist. Bob was also inspired by Sewall Wright. Emerson suggested that he study variation in the cottonwood aphid *(Pemphigus populitransversus*). His research was also inspired by Alfred Kinsey's studies of gall wasps. Wright supervised the completion of his thesis, "because he was the only one who could understand it."[2] His doctoral work, completed in 1952, laid the foundation for his future research in statistical biology. He was strongly influenced by Wright and by Thomas Park's meticulous lectures and experimental study of population dynamics in cultures of *Tribolium* (flour beetles).

The Kansas Years

Bob wrote in his memoir that he did not have an affinity for any group of organisms, but that his career track was based on his interest in discovering patterns in nature and in inferring process from pattern. He joined the Department of Entomology at the University of Kansas (KU) on a research fellowship and joined the lab of Charles Michener, who was studying insecticide resistance (but transformed the project into a study of quantitative genetics). He became an instructor in 1951 and then was hired in a tenure-track position as an assistant professor in 1953; by 1961, he had been promoted to a full professor of statistical biology. During this time, his children David and Hannah were born, and he taught courses in population genetics, multivariate statistics (called "Correlation and Causation in Biology"), insect ecology, and biometry. His initial research was on the genetics of DDT resistance in houseflies, because grant funding was available for that topic.

As reported by Berry J. Brosi and Paul R. Ehrlich in 2016, his unexpected creation of the field of numerical taxonomy began at an informal lunchtime seminar in the Department of Entomology.[3] Bob rather rashly suggested that statistical methods could classify insect species more reliably than the traditional systematic approaches, in which an authority on a taxon would assert that certain specified features (such as form of the genitalia) were more suitable than others (such as wing length). The systematists in the Department of Entomology, understandably, strongly disagreed. Charles Michener, an eminent authority on bees, agreed to a challenge to compare Sokal's suggested new approach with the conclusions he had reached in an extensive analysis of a group of solitary bees. Bob then had to develop an actual method and perform the computations. With no computers at this time, the effort required a large amount of calculations by hand with tabulating machines and mechanical calculators. Michener and Sokal concluded that the results showed that Bob's new methods "can be used to remove some of the subjective bias from taxonomy."[4] His approach came to be called "numerical taxonomy."[5] It was very controversial, and there was tremendous resistance to this new method (and to some, even to just the idea that statistical methods and computers might be useful in systematics). Although some saw the possibilities of their use, many others resisted. The curator of the Snow Entomological Museum at KU said he would delay his retirement just to make sure Ehrlich's prediction about computers in museums would be wrong.[6] Even among those who accepted the usefulness of computers to carry out statistical calculations there was resistance to the idea that cameras could be connected to computers so that eventually they could analyze images directly. Sokal engaged in many debates, some with angry opponents, and proved to be a good debater who showed how one could deal with scientific controversy in a reasoned and professional way. In all, he published two books and 68 papers in numerical taxonomy.

The Stony Brook Years

In 1968, Bob and his former student F. James Rohlf (who also worked on numerical taxonomy) accepted positions at the State University of New York at Stony Brook (now Stony Brook University). They joined the newly formed Department of Ecology and Evolution, led by Lawrence Slobodkin. During his tenure at Stony Brook, Bob held several positions, including department chairman, graduate program director for the doctoral program in anthropology, vice provost for research and graduate studies, leading professor, and distinguished professor.

By then, numerical taxonomy had become established and was sometimes called a "phenetic" approach because most studies simply described how species could be objectively

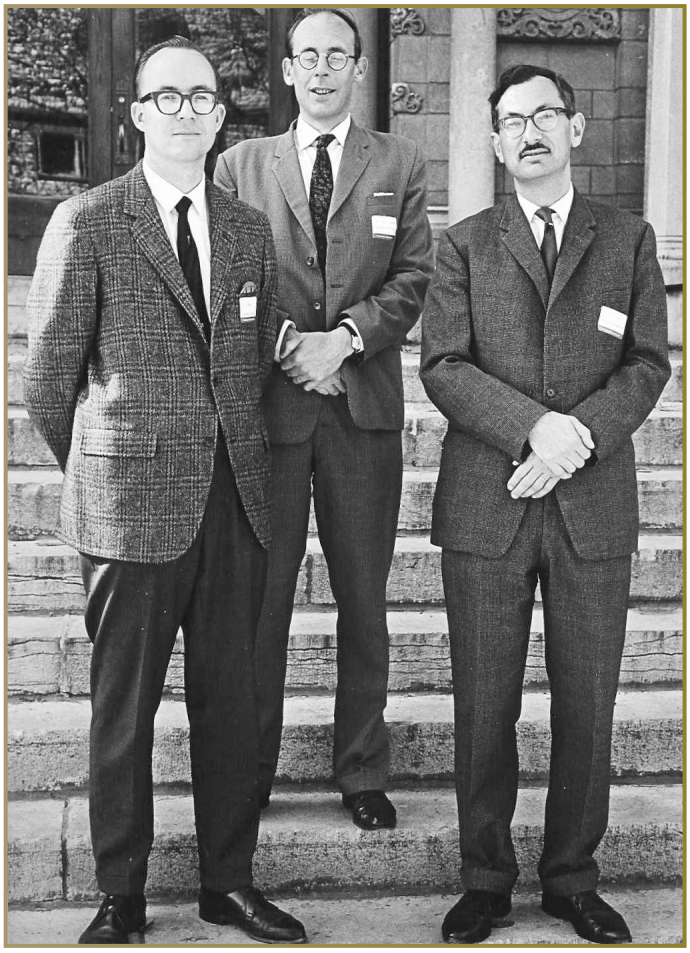

Figure 2 From left, F. James Rohlf, Peter Sneath, and Robert Sokal at a Numerical Taxonomy conference in Portugal in 1974.

sorted into groups within groups, without claiming that these necessarily corresponded to historical evolutionary units ("clades"). Thus, the purpose of most studies differed from the proposition, at first associated with the entomologist Willi Hennig, that taxa should represent monophyletic lineages (clades), each composed of all the descendants of a common ancestor. This was despite the fact that Sokal and his KU colleague Joseph Camin developed the first numerical method and software to estimate a phylogenetic tree based on a form of parsimony.[7] Systematics thus became divided into opposing camps with even more controversy. Bob then became involved in the "cladistic wars" of the 1970s requiring some painstaking work to check claims of the stability and accuracy of different methods that seemed to him to be unlikely to be true. The limitation of all methods at the time was that they were based mostly on morphological characteristics that in most cases were arbitrarily defined. This field would likely have developed very differently if extensive DNA data were as available then as they are today.

At Stony Brook, Bob established a large lab devoted to experimental studies of houseflies and especially *Tribolium* (following the examples of Thomas Park). The emphasis was on ecological genetics, especially frequency changes in mutant alleles as a function of population density. These studies continued for eighteen years and resulted in thirty-four publications. Some studies yielded evidence of inverse frequency-dependent fitness, in which a genotype's fitness is greater when it is rare than when it is common. During this period, he also revived his studies of morphological variation in the aphid *Pemphigus populitransversus* based on samples collected during extended summer journeys with his family, crisscrossing the United States in a camper van. (Students of his textbook *Biometry* became familiar with *Pemphigus*, as the source of data for many test questions.)

The last five of Bob's research years were focused almost exclusively on variation within and among human populations, owing in part to his European heritage and his command of languages in Europe and China, and also in part because of the abundant information on humans. As he noted later regarding a 1987 lecture at Stony Brook, "This work has turned me into a physical anthropologist," with forty-nine publications on gene frequency variation among populations, correlations between genetic and linguistic variation, and ethnohistory. Anthropologists agreed that he had become one of their own: in 2004, the American Association of Physical Anthropologists presented him with the Charles R. Darwin Award for "lifetime achievement in physical anthropology." (He was very pleased to receive an award named for one of his intellectual heroes.) And in 2003, he was chosen as the Raymond Pearl lecturer by the Human Biology Association. To give just one example of his studies of humans, he, Jiangtian Chan, and Merritt Ruhlen analyzed the correspondence between language differences and genetic differences among 130 populations, worldwide, using a genetic distance matrix that allowed for differences among populations in the size (degree of completeness) of the genomic samples.[8] Both for complete and partial genome samples, they found that genetic difference and language difference were correlated, even when geographic distances were held constant. They noted particularly strong differences between populations from eastern Asia, the Arctic, and Australia and those from Africa and Europe.

Key Contributions

Robert Sokal's pioneering work in numerical taxonomy emphasized the use of statistical methods to classify organisms based on their characteristics. This approach was detailed in his influential book, *Principles of Numerical Taxonomy,* which he co-authored with Peter Sneath.[9] This work laid the groundwork for modern systematic biology. But Sokal's

contributions extended beyond taxonomy. His research in spatial autocorrelation and its applications in biology provided new insights into the genetic and geographic distribution of species. His work with colleagues on the spread of agriculture in Europe and the origins of the Indo-European languages showcased the interdisciplinary nature of his research, combining genetics, geography, and linguistics.

The reason he worked on several different organisms was because, as he later wrote in unpublished notes, "I was never so much in love with *Pemphigus* or with *Tribolium* the way some people are with their particular organisms. To me it was the pattern, the regularity, the process that was the important thing, and I did not care what organism provided the insight."

Sokal also played a major role in graduate student training. His course in biometry (and the textbook that now has about 100,000 citations over the four editions) is recognized as an essential part of the education of many graduate students who intend to pursue biological research.

Legacy

Throughout his career, Sokal maintained his very high standards for intellectual honesty. His achievements in any one of these fields would have been a sufficient accomplishment for most academics. Perhaps his most significant and lasting contributions were in developing quantitative and statistical methods for studying complex systems. Numerical taxonomy was a revolutionary approach to systematics, simply by being quantitative and more objective and repeatable than traditional approaches. Quantitative, statistical methods in phylogeny and systematics have increased greatly in power and objectivity since his contributions, but he blazed the trail. Similarly, he approached the historical causes of geographic variation in human populations quantitatively, using and extending spatial autocorrelation and other statistical approaches. In doing so, he influenced the development of a subject far removed from his training in entomology and evolution. His rich personal life informed his professional one. Fluent in German, French, Portuguese, Chinese, and Hebrew, he was a polyglot who embraced diverse cultures and intellectual traditions.

Sokal was a member of numerous prestigious societies, including the National Academy of Sciences, the American Academy of Arts and Sciences, and the Linnaean Society of London. He served as editor of the *American Naturalist* from 1969–1974. His numerous awards and honors, including the Guggenheim Fellowship and the Charles R. Darwin Award for Lifetime Achievement, are testaments to his impact on the scientific community.

His contributions to science extended far beyond his innovative research. His dedication to teaching, his interdisciplinary approach, and his ability to inspire others have left an indelible mark on the scientific community. This memoir not only celebrates his scientific achievements but also honors the man whose work and life continue to inspire and guide future generations.

Figure 3 Robert and his wife Julie at his retirement celebration in 1995.

Personal Notes

I (FJR) have known Bob since June 1958, when I became his new graduate student to work on what became numerical taxonomy. I had little idea then that more than fifty years later we would still be close colleagues. My years as a graduate student were exciting as Bob was developing this new scientific field. He made a tremendous impression on me and the other students at KU. We learned by example that hard work was necessary to develop new findings and one also had to understand past work by others. We also learned how to deal with controversy. Over the years, I was continually amazed by what he accomplished in this new field while at the same time accomplishing so much on unrelated projects on the genetics of houseflies, the ecology of flour beetles, geographic variation in aphids, and human migration patterns. The preparation of this memoir was greatly assisted by his own preparation. He left behind autobiographical documents for his family and also published his scientific autobiography in the same year he died.[10] I can also attest to the fact that during his publishing years he had a very strict code of conduct concerning joint papers. He collaborated with many, and these were true collaborations. If his name was listed, he had done a significant part of the work. As he has written in one of the many documents he left behind, "There is hardly a paper on my publication list for which I have not personally done at least half of the work, conceptually, organizing it, or starting it. I may not have done the hands-on work, but I will have done the analysis and the writing in large part. Thus, I absolutely will not add my name to a student's or postdoc's

paper unless I myself have done a significant part of the work, or unless the paper is unlikely to see the light of publication without my collaboration." I remember many hours I spent working with him at his tape recorder dictating papers. He did most of the talking, and I passed notes and kibitzed from the side. He has also commented, "I am the archetypal collaborator," and much of his bibliography is coauthored.

As a new Ph. D. who joined the faculty in 1970 (with the garb and hair style of the age), I (DJF) greatly respected but felt intimidated and judged by this renowned, dignified European in suit and tie. As years passed, we developed a strong relationship, and I have reason to think he acted to my benefit. Bob's research assistants remember him with respect and gratitude. One writes that he engendered a contagious enthusiasm for science by asking his lab group every day, "What's new and exciting?" She writes, "Bob was one of the most important people in my life [who] shaped my career in many ways." He was "a scientist of uncompromising integrity… kind, honest, and a great storyteller,… a mentor, a generous boss, and a dear friend." Another research assistant describes how she developed seizures that her neurologist said were not treatable. Bob connected her with another neurologist who used experimental brain surgery that revealed a malignant tumor. He successfully removed it, and she recovered completely. Bob, she writes, "saved my life … [T]he man behind the science really cared for people."

Note

Robert Sokal's extensive research notes, manuscripts, correspondence, lectures, and related documents from 1953 to 2004 are available to researchers in the Stony Brook University Faculty Collections. A description of the contents of the thirty-four boxes is available online at: https://www.stonybrook.edu/commcms/libspecial/archives/collections/faculty/sokal.php.

References

1 Schomann, S. 2008. *Letzte Zuflucht Schanghai: die Liebesgeschichte von Robert Reuven Sokal und Julie Chenchu Yang. [Final Refuge Shanghai].* München: Heyne.

2 Sokal, R. R. 2012. A sketch of my scientific autobiography. *Hum. Biol.* 84:489–505.

3 Brosi, B. J., and P. R. Ehrlich. 2016. Charles Duncan Michener, 1918–2015. *Proc. Nat. Acad. Sci. U.S.A.* 1131:1963–1964.

4 Michener, C. D., and R. R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution* 11:130–162.

5 Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38:1409–1438.

6 Ehrlich, P. R. 1961. Systematics in 1970: Some unpopular predictions. *Syst. Zool.* 10:157–158.

7 Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19:311–326.

8 Chan, J., R. R. Sokal, and M. Ruhlen. 2012. Worldwide analysis of genetic and linguistic relationships of human populations. *Hum. Biol.* 84:573–580.

9 Sokal, R. R., and P. H. A. Sneath. 1963. *Principles of numerical taxonomy.* San Francisco: W. H. Freeman.

10 Sokal, R. R. 2012.

Selected Bibliography

Numerical Taxonomy

1957 With C. D. Michener. A quantitative approach to a problem in classification. *Evolution* 11:130–162.

1958 With C. D. Michener. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38:1409–1438.

1963 With P. H. A. Sneath. *Principles of Numerical Taxonomy.* San Francisco: W. H. Freeman.

1965 With J. H. Camin. A method for deducing branching sequences in phylogeny. *Evolution* 19:311–326.

1973 With P. H. A. Sneath. *Numerical Taxonomy: The Principles and Practice of Numerical Classification.* San Francisco: W. H. Freeman.

1974 Classification: Purposes, Principles, Progress, Prospects. *Science* 185:1115 –1123.

Experimental Ecological Genetics

1970 Senescence and genetic load: Evidence from *Tribolium*. *Science* 167:1733–1734.

Analysis of Geographic Variation

1978 With N. L. Oden. Spatial autocorrelation in biology: 1. Methodology. *Biol. J. Linn. Soc.* 10:199–228.

1979 Testing statistical significance of geographic variation patterns. *Syst. Zool.* 28:227–232.

Human Population Movement in Europe

1982 With P. Menozzi. Spatial autocorrelation of HLA frequencies in Europe support demic diffusion of early farmers. *Am. Nat.* 119:1–17.

1988 Genetic, geographic, and linguistic distances in Europe. *Amer. J. Phys. Anthro.* 75:273–273.

1991 With G. M. Jacquez. Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis. *Evolution* 45:152–168.

With N. L. Oden and C. Wilson. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351:143–145.

1992 With N. L. Oden and B. A. Thomson. Origins of the Indo-Europeans: Genetic evidence. *Proc. Natl. Acad. Sci. U.S.A.* 89:7669–7673.