

History of the Department of Statistics, now the Department of Statistics and Data Science, at Carnegie Mellon University, 1966-2011

October 8, 2021

The Department of Statistics at Carnegie Mellon University—then Carnegie Institute of Technology—was created in 1966. Prior to that time, Carnegie Tech’s Department of Mathematics served as training ground for an outstanding mix of young statistical talent, such as undergraduate and masters degree students Frederick Mosteller (1930s), David L. Wallace (1940s), and Miron Straf (1960s). Among the statistically-oriented instructors who began their distinguished careers in the Mathematics Department were Abraham Charnes, Carlton Lemke, Lester E. Dubins, Donald P. Gaver, Jr., and M.M. Rao.

In 1957, Morris H. DeGroot joined the Department of Mathematics as it was moving toward pure mathematics, a shift that did not bode well for the future hiring of statisticians. The only other statistician in the department was former Institute of Mathematical Statistics (IMS) President Edwin G. Olds who, despite being overshadowed in numbers by mathematicians for decades, was able to effectively make the case for statistics.

Olds’ influence on budding statisticians at Carnegie Tech can be traced back to the mid-1930s. In *The Pleasures of Statistics: The Autobiography of Frederick Mosteller*, Mosteller recalled having to compute the probability of casting a total of 9 and 10 using three ordinary dice in a physical-measurements course. When as a sophomore he asked how to solve the problem using a large number of dice, he was directed to Olds.

“He [Olds] began showing me slowly and carefully how to do the three-dice problem ... The method was that of generating functions, and it magically, and I do mean magically, counted how many ways there were to get each total with the dice. Although I had loved mathematics all along, this was the first time I ever felt that I’d been working with a peashooter when I could have had a cannon ...

“Soon I was a mathematics rather than a physics major.”

Hooked on probability and statistics, Mosteller would go on to earn a Ph.D. in mathematics from Princeton University. In 1957, en route to becoming one of the premier statisticians of the 20th century, Mosteller founded the Department of Statistics at Harvard University almost a decade before Carnegie Tech had one; (as fate would have it, one of his early Ph.D. students at Harvard was Stephen Fienberg, who later became head of Carnegie Mellon’s Department of Statistics and dean of its College of Humanities and Social Sciences (H&SS)). In 1964, Mosteller and fellow Carnegie Tech and Princeton alumnus Wallace authored their classic book, *Inference and Disputed Authorship: The Federalist*. When it was featured on the cover of *Time* magazine for its use of statistical methodology to solve the disputed authorship of the Federalist Papers, Carnegie Tech’s Mathematics Department basked in the reflected limelight.

By the mid-1960s, DeGroot’s interests had been diverging from his colleagues for years, leading him to yearn for a group of colleagues with whom to share his passion for research in decision making and other statistical interests. His friend and lifelong collaborator, Richard M. Cyert, then Dean of the Graduate School of Industrial Administration (GSIA; now the Tepper School of Business) agreed. Together they developed a plan for a Department of Statistics with DeGroot as its head and Gaver as the other senior faculty member. A year later, Carnegie Tech became Carnegie-Mellon University. (In 1987, the hyphen was removed and the school assumed its present name of Carnegie Mellon University.)

In its early years the department had only a few, albeit outstanding, faculty members, such as Norman Starr, Stanley Sclove and Michael Woodrooffe. The department, while not part of any college, reported to Cyert, first as Dean and then as Carnegie Mellon president. Notable graduate students during this time included Prem Goel (former chairman of the

Dept. of Statistics at Ohio State University), the late David Hildebrand (former chairman of the Dept. of Statistics at the University of Pennsylvania), and Benjamin Kedem (Professor of Mathematics at the University of Maryland).

In 1969, Jay Kadane joined the department. Three years later, DeGroot stepped down as head when his wife became ill and Kadane, who was six years out of graduate school, succeeded him. In addition to DeGroot and Kadane, the faculty at the time consisted of John Lehoczky, William W. Davis, Paul Shaman, Al Kinderman, and Gerry Ramage. There were a handful of graduate students, no undergraduates, and two staff members.

The department's offices in the early years were on the eighth floor of the brick-walled, dungeon-like Science Hall, now Wean Hall, with odd aromas provided by the nearby chemistry laboratories. While Carnegie Mellon had some computers which used racks of IBM cards, Statistics had none. The department also still lacked a college to call home, as it was reporting both to the dean of GSIA and to the dean of engineering. Blazing its own path, the department decided to focus on Bayesian statistics, applications of statistics, and, later, statistical computing, when none of these was widely recognized as crucial to a growing department. In keeping with the latter focus, Statistics rented its first terminal: an interactive terminal hard-wired into the university computer system.

An early important proposal by Kadane that was adopted by departmental consent was to evaluate applied research on the basis of whether it was a contribution to the applied area and not necessarily a contribution to statistics. Another part of that strategy was to forego operating a consulting center for drop-in statistical advice. Instead, faculty were encouraged to participate in joint work with researchers in other fields. As a result, the department managed to develop collaborative relationships with many scholars across campus for decades to come.

George Duncan joined the department in 1974, followed by William Eddy and Diane Lambert two years later. As almost none of the faculty had spouses or children at the time, there was a lot of after-hours socializing, such as Wednesday evening cocktails and hors d'oeuvres in the Faculty Club in Skibo. Everyone also gathered at noon each day in their

shared hallway in Science Hall to walk for lunch to Skibo (the campus student union named after Andrew Carnegie's ancestral home in Scotland).

In what would become a hallmark of the department to the present day—journal and magazine editorships—DeGroot followed up his 1971-1975 editorial work as book review editor of the *Journal of the American Statistical Association (JASA)* by serving as the publication's theory and methods editor from 1976 to 1978. Kadane was his deputy editor for the latter. Duncan took over that *JASA* editorship from DeGroot in 1979.

In 1978, Wen-chen Chen came aboard; about the same time, the department offices moved from the eighth to the seventh floor of Science Hall. Mark Schervish arrived in 1979 – an important year for collaborations. During a graduate seminar taught by Kadane, Schervish met University of Pittsburgh philosophy faculty member Teddy Seidenfeld (who later became a Carnegie Mellon joint philosophy/statistics faculty member). Their interactions would eventually evolve into weekly research meetings and numerous papers by the trio, both of which continue today.

Meanwhile, across the Atlantic at the First Valencia International Meeting on Bayesian Statistics in Valencia, Spain, DeGroot and Stephen Fienberg were beginning their collaboration on comparing probability forecasters. At this meeting, DeGroot was able to finally convince Fienberg, whom he and Kadane were aggressively courting, to join the department a year later in 1980. The final negotiating item involved finding opportunities in Pittsburgh for Fienberg and his son to play hockey.

The dawn of the 1980s was a time of change and development, with faculty numbering about 10 and graduate students roughly 15. Among them was Luke Tierney, in his first faculty post, and Ph.D. student Kathryn Chaloner. The couple met at Carnegie Mellon and later married. Today, both are professors and department heads at the University of Iowa—she in the Department of Biostatistics, and he in the Department of Statistics and Actuarial Science.

In 1980, Statistics finally landed a home when it joined H&SS, where it remains a core department today.

In 1981, Robert Kass was hired, and Fienberg became head just as he was starting his

six-year stint as chair of the Committee on National Statistics (CNSTAT) at the National Academy of Sciences. As department head, he oversaw submission of a proposal to the first Scientific Computing Research Equipment in the Mathematical Sciences (SCREMS) program at the National Science Foundation, for which the department received funding to purchase its first mini-computer for \$250,000. This was the start of the department's leadership in statistical computing beyond Carnegie Mellon, a position it retains to the present day.

Collaboration continued to flourish, as DeGroot, Fienberg, and Kadane began work on the use of statistics in legal settings, which would become the edited volume, *Statistics and the Law*. At a downtown Pittsburgh hotel, the department hosted the international Computer Science and Statistics 13th Symposium on the Interface, chaired by Eddy. This was to be the first of many conferences and workshops hosted by the department and its faculty.

Amidst a swirl of activity, tragedy struck. On July 3, 1981, while visiting family in Taiwan with his wife and year-old son, 31-year-old Wen-Chen Chen was found dead. Chen had undergone extensive questioning by Taiwan authorities regarding his alleged political activities in the U.S. With DeGroot as the driving force in bringing attention to what the Taiwanese government was dismissing, he and renowned U.S. pathologist and Pittsburgh resident Cyril Wecht traveled to Taiwan. While Wecht concluded through autopsy findings that the promising young faculty member was murdered, there was no further action by the Taiwanese government.

By 1982, the department continued to grow and, with the university-wide tentacles of a sprouting bureaucracy compounding regulations and paperwork, Statistics hired its first business manager. A year later Joel Greenhouse arrived to begin a post-doctoral position supported by a National Institute of Mental Health (NIMH) grant for training in psychiatric statistics with Lehoczky and DeGroot as principal investigators.

In keeping with the the tradition of departmental editorships of journals and magazines, DeGroot was spearheading the creation of *Statistical Science*, dedicated to communication among statisticians regardless of specialty, while Kadane was in the middle of a three-year stint as applications and coordinating editor of *JASA*. On the honors front, Fienberg was

the 1982 recipient of the Committee of Presidents of Statistical Societies Presidents' Award (COPSS) as the outstanding statistician age 40 or younger—the first of three Carnegie Mellon Statistics faculty to receive this honor.

By the early 1980s, the department enjoyed a world-wide reputation in Bayesian statistics and decision making, and was developing strength in statistical computation. At the same time, Carnegie Mellon was stressing these same areas, as well as cross-disciplinary research, in its strategic planning. The university's recognition of the importance of statistics in such research would result, a decade later, in the department's playing a key role in several major cross-disciplinary efforts on campus, such as the Center for the Neural Basis of Cognition (CNBC), and the Center for Automated Learning and Discovery (CALD), now the Machine Learning (ML) Department.

In 1984, the department moved from Wean Hall to two newly renovated wings of Baker Hall, almost doubling its space in faculty, staff, and student offices, and acquiring two conference rooms, a lounge, and more. To ensure frequent interaction, graduate student offices were interspersed among faculty offices. A spring evening was the setting for faculty, staff, and university officials to unite in celebration as DeGroot was recognized for his decades of contributions and achievements with Carnegie Mellon's highest honor: a University Professorship. Later in the year, Fienberg stepped down as head, and Lehockzy assumed the post.

In 1985, Fienberg, now Maurice Falk Professor of Statistics and Social Science, and Eddy began developing the idea for a quarterly magazine for anyone with an interest in the analysis of data—*CHANCE, New Directions for Statistics and Computing*. It would reach fruition three years later, and eventually become a publication of the American Statistical Association for the next 20 years and counting. By mid-decade, the department housed about 15 faculty and 30 graduate students.

Academic statistics departments were afforded a unique opportunity to acquire their own equipment as a result of the awarding of the first SCREMS funds in 1982, followed by Dept.

of Defense funds under its University Research Instrumentation Program (DURIP). The Statistics Department moved from general availability of terminals in 1981, to a VAX 11/750 in 1983, to several workstations in 1985 – all maintained by Eddy. By 1986, the department’s equipment included one VAX 11/750 superminicomputer with 912 MB of disk storage, one VAXstation 500 color workstation, one CSPI Mini-Map array processor, five GIGI color microcomputers, and more. All machines were interconnected by a 10Mb Ethernet using DECnet and TCP/IP software.

The mid-1980s marked the graduation of the department’s first undergraduate statistics majors. Meanwhile, the labors of late-1970s through mid-1980s Ph.D. graduates were bearing fruit in government, industry, and academia. At the National Institutes of Health, Todd Sahlroot and Dean Follmann engaged in research on clinical trial and statistical modeling of biomedical data. Kathy Blackmond Laskey, at the Decision Sciences Consortium, analyzed data from the national pesticide survey of well water for the Environmental Protection Agency, while Sheryl Kelsey, as co-director of the Epidemiology Data Center at the University of Pittsburgh, oversaw research studies sponsored by the National Institutes of Health and other agencies. Diane Sapphire and Elizabeth Stasny were beginning their academic careers at, respectively, Trinity University and Ohio State University.

In 1987, Fienberg was appointed dean of H&SS and, a year later, Larry Wasserman was hired. Eddy was developing a journal to extend the use of computational and graphical methods in statistical and data analysis that would become the *Journal of Computational and Graphical Statistics*, a joint publication of the ASA, IMS, and the Interface Foundation of North America.

By 1989, the department’s computing equipment included nine terminals, five printers, 11 personal computers, and 32 workstations. To maintain the growing inventory, the department hired its first full-time systems manager. That same year, faculty member Michael Meyer developed StatLib, a system for distributing statistical software, datasets, and information by electronic mail, FTP, and WWW, and which is hosted by the Statistics Dept. to this day.

Sadness gripped the statistical community with the death of Morrie DeGroot on Nov.

2, 1989. As former student and collaborator Prem Goel remarked at the memorial service, mirroring the department's sentiment: "... [that] Morrie has not gone anywhere, that he has merely changed his body which was worn out, that memories of all the good times with Morrie are still with me, and that his ideas and advice will keep on showing me the right path."

Two years later the department formally honored its founder through its on-going, bi-annual Morris H. DeGroot Lecture held in conjunction with the first of a long series of Case Studies in Bayesian Statistics workshops co-organized by Kass on applications of Bayesian statistics to problems in science and technology.

The department also transformed a conference room into the Morris H. DeGroot Memorial Library containing the bequeathed books and periodicals of its namesake. Over the years, the collection has expanded to include hundreds of faculty contributions and some new purchases.

By the start of the next decade the major thrusts of the department had evolved into Bayesian statistics, computational statistics, industrial statistics and stochastic modelling, and biostatistics and psychiatric statistics. In keeping with the latter, Brian Junker joined the department as a postdoctoral fellow in the Program in Psychiatric Statistics, jointly operated by the department and Western Psychiatric Institute and Clinic. Fienberg returned in 1993 following a two-year stint as academic vice-president at York University in Toronto. With the arrival of Kathryn Roeder and Christopher Genovese a year later, personnel rose to 14 tenure-track, one research scientist, and three full-time-equivalent visiting faculty.

By 1995, StatLib had grown to a collection of about 150 Mbytes, and the StatLib server was servicing about 60,000 monthly transactions. In the classroom, curricular and pedagogical innovations introduced in new H&SS core courses, "Statistical Reasoning" and "Introduction to Statistical Methods," were partly credited with the rise in the number statistics majors and minors, and with the overall interest in the discipline: by spring, 1996, there were over 900 undergraduates enrolled in department courses.

The development of the graduate program continued with roughly 36 students. The

number of special graduate education programs by now included the joint Ph.D. with the H. John Heinz III School of Public Policy and Management (now the H. John Heinz III College) and the joint Ph.D. in statistics and machine learning. The program expanded further with the Masters of Science in Computational Finance (MSCF), a joint venture of four departments, including Statistics, for future investment professionals. Besides local classroom instruction, MSCF students attend class via a remote television connection in New York City.

In 1995, Kass was named the new head, replacing Lehoczky after 11 years. On the research front, the brain provided fodder for ever-widening statistical collaboration. A new tool to measure the change in blood flow related to neural activity in the brain—functional Magnetic Resonance Imaging (fMRI)—offered statistical challenges for Eddy, Genovese, and new hire Nicole Lazar. Kass' interest in statistical modeling of neural activity culminated in a faculty appointment – and executive committee membership – in Carnegie Mellon's new CNBC, established to foster interdisciplinary research on the neural mechanisms of brain function. (A joint CNBC-Statistics Ph.D. program was approved in 2011.)

To keep pace with the increasing number of statistics undergraduate and graduate degree recipients, outreach initiatives were instituted, like departmental newsletters and the annual alumni breakfasts at the Joint Statistical Meetings. Meanwhile, the influx of young faculty not only provided a balance with senior faculty, but heralded a change in family profiles. The after-hours socializing of the prior decade gave way to afternoon baby showers and child care schmoozing. New faculty also meant new honors. Roeder was the 1997 recipient of the COPSS Award, with Wasserman accepting the award two years later. Genovese received an NSF CAREER Award for young researchers. For senior faculty, Lehoczky's being named the Thomas Lord Professor of Statistics raised the number of chairs to four: Kadane was the Leonard J. Savage Professor of Statistics and Social Sciences; Fienberg, recently elected president of the IMS and the International society for Bayesian Analysis, was named a University Professor; and Seidenfeld became the first recipient of the Herbert A. Simon Chair in Philosophy.

As the world prepared for the new millenium, the department set its sights on NSF fund-

ing that was awarded jointly with Carnegie Mellon's Department of Mathematical Sciences through a Vertical Integration of Research and Education grant, or VIGRE. The goal of the VIGRE program, as articulated by NSF, is to "increase the number of well-prepared U.S. citizens, nationals, and permanent residents who pursue careers in the mathematical sciences..." A few years after the joint award, Statistics was singly awarded VIGRE funding. Through that assistance, the Statistics Department enhanced its undergraduate, graduate, and postdoctoral programs. With the graduate and postdoctoral programs, mentored cross-disciplinary work, communication, and teaching were emphasized.

Newly instituted teaching teas and research teas encouraged interaction among regular faculty and VIGRE trainees at all levels. (By 2006, VIGRE participants numbered three faculty principal investigators, 15 faculty mentors and project leads, six VIGRE-supported post docs/visiting assistant professors, 11 VIGRE-supported graduate students, and 28 VIGRE-supported undergraduates.) The grant also gave birth to the department's eight-year Summer Undergraduate Research Experience (SURE), directed by Greenhouse. Students – primarily from historically black college and universities – interested in public health careers designed supervised research projects such as "Factors Associated with the Perceptions of Obesity and Depression Among Adolescents in the U.S.," and "The Hidden Cost of Caregiving: Emotional Stress." Students presented their work at a poster session to the campus community at the conclusion of the eight-week SURE.

Maintaining the focus on undergraduate education, by 2000 the department offered a variety of programs, with the most popular being: (1) the B.S. and minor in statistics in H&SS; (2) the statistics and operations research concentrations of the B.S. in mathematical sciences in the Mellon College of Science; and (3) the statistics and data mining concentration of the B.S. in information systems in H&SS. The college itself was expanding, as a 20,000-square-foot addition provided new research areas, multimedia classrooms, conference rooms, meeting space, and more. When H&SS lost its dean to another university, Lehoczky was named interim dean. Faculty honors during this period included Kadane being named a University Professor, and Fienberg's election to the National Academy of Sciences followed by his election to the Royal Society of Canada five years later.

Cross-disciplinary work among faculty continued to span a variety of scientific, business, and policy domains, including archeology, astronomy, biostatistics, data mining, education, environment, finance, government, marketing, neuroimaging, neurophysiology, and social science/public policy. In 2001, faculty published 67 articles and commentaries, and six books. A year later, an advisory board report called the department “world class in the caliber and rigor of its theoretical and methodological research, in its contributions to advancing the discipline of statistics, and in its education programs. Moreover, it is preeminently so in interdisciplinary research.”

Statistics faculty have held joint appointments in the departments of machine learning, mathematics, philosophy, social and decision sciences, as well as in GSIA (now the Tepper School of Business), and the Heinz College (public policy and management), and in numerous interdisciplinary centers such as the Center for the Neural Basis of Cognition, Cylab, and the Center for Computational Biology.

Oded Meyer, appointed in 1999, was named the department’s first regular-faculty lecturer. At the same time, reformulated classes like Advanced Data Analysis II, in which students prepare semester-long projects to present at Carnegie Mellon’s competitive annual “Meeting of the Minds” undergraduate research symposium, continued to attract newcomers to the field.

The growth of the graduate and undergraduate programs, and the research activities, also added pressure on the small amount of space the department occupied, and still does. Appropriate office space for new faculty hires can be problematic. Some graduate student offices are in non-abutting sites, and important computing equipment is located in rented remote locations. While the reconfiguration of former faculty offices into multi-desk offices has proven a viable option for fostering a sense of community for students, contiguous space remains the department’s most pressing issue.

In that vein, Eddy opened, in 2002, a 2500-square-foot “branch office” three blocks from campus with a staff engaged in imaging research – optical video imaging, microscopic histological imaging, and laser fluorescence imaging. Faculty honors during this period included

Eddy becoming chair of CNSTAT, Junker's appointment as editor of *Psychometrika* and Kass's as editor of the new electronic *Bayesian Analysis*, and Seidenfeld's being named a University Professor. Eddy and Fienberg were designated lifetime National Associates by the National Academy of Sciences.

On July 1, 2004, Schervish became head, replacing Kass who had served nine years in the post. In addition to the 17 regular faculty, there were seven visitors, two instructors, and two faculty with external appointments. On the academic front, the department launched the new B.S. in economics and statistics degree program, and began restructuring its graduate program so students could engage in research earlier in their careers with fewer required courses. Fienberg was elected to the Royal Society of Canada in 2004 and to the American academy of Arts and Sciences in 2007, joining Seidenfeld who was elected several years earlier. On the social front, departmental celebrations kicked into high gear over the next five years with events in honor of Kadane attaining emeritus status (first in the department's history), Fienberg's 65th birthday, Eddy being named the first John C. Warner Professor of Statistics, and Lehoczky's 40 years at Carnegie Mellon.

By 2008, the number of undergraduate majors totalled about 100; not coincidentally, that same year Meyer won the Elliott Dunlap Smith Award for excellence in teaching in the college. On the graduate side, the new Masters in Statistical Practice (MSP) program premiered a year later: a one-year, two-semester professional masters degree program emphasizing statistical practice, methods, data analysis, and workplace skills, and directed by Greenhouse.

The department concluded one decade and began another with a resounding mix of junior and senior faculty recognition: Jiashun Jin was winner of the 2009 Tweedie New Researcher Award from the IMS, and Rebecca Nugent received the 2009 Chikio Hayashi Award from the International Federation of Classification Societies to a promising young researcher. Kadane was elected to the American Academy of Arts and Sciences in 2010, while Eddy completed his second three-year term as chairman of CNSTAT, becoming the only person to have chaired both statistics committees at the National Academies, having previously chaired the Committee on Applied and Theoretical Statistics. In 2011, Lehoczky joined the ranks of

University Professors.

As of late 2010, 145 undergraduates were pursuing majors in the department, and seven were pursuing minors. There were 882 students enrolled in undergraduate statistics courses.

Fall, 2011, is record-setting on a number of fronts: There are 65 graduate students – our most ever! – which includes 18 MSP students – another high! Faculty, who are comprised of professors, special faculty, professors in residence, postdoctoral research associates, and more, total 33 – one of our biggest numbers – with the newest members being assistant professor Ryan Tibshirani, and Heinz College associate professor Amelia Haviland, a 2003 department alumnus with a joint Ph.D. with Heinz; she has a courtesy appointment in the Statistics Department.

Finally, but certainly not lastly, H&SS became the Marianna Brown Dietrich College of Humanities and Social Sciences in honor of the mother of William S. Dietrich II, who intends to provide a gift of a \$265 million fund to support Carnegie Mellon.

It will be the eighth largest gift ever by an individual to a private higher education institution in the U.S.

What Has Made Statistics at Carnegie Mellon Special?

Several features of the research and other activities of the department have made it distinctive at Carnegie Mellon, and a leader among its peers at other universities around the world.

First, from its founding, the department has been home to a focus on Bayesian inference, and this focus flourished as faculty developed computational approaches such as Laplace's method, and in the 1990s with the emergence of Markov chain Monte Carlo methods. This intellectual focus fit nicely with the department's emphasis on statistical computing and its involvement in the creation of the Machine Learning Department in the School of Computer Science.

Second, the department's emphasis on collaborative research established it within Carnegie Mellon as an essential locus of research and teaching linked to almost all units of the university. Statistics graduate students and faculty engaged in interdisciplinary research before

it became fashionable in most other departments around the United States, and faculty and students alike publish widely in scientific journals outside the field of statistics; they are often known as much for their substantive contributions in other areas as for their publications in statistics journals. Many of the department's students are enrolled in joint programs, further solidifying this interdisciplinary focus.

Finally, the department and the university have taken pride in the professional leadership roles played by its faculty, from editorships of journals and book series, to leadership in professional societies and at the National Academy of Sciences.

Ed Olds might not recognize the department were he alive today, but he would surely appreciate its commitment to teaching, to its undergraduate and graduate students, and to the professional leadership role which he initiated over a half century ago. The introduction in 2010 of an electronic application system for the Graduate Program (M.S. and Ph.D.) and the MSP Program garnered record applications to both. The department's many and varied successes, including its swelling numbers of undergraduate majors, remain cornerstones crucial to the department's strategic plan, which is the same as at the department's founding 45 years ago: training problem solvers to be in a position to choose amongst a variety of professional opportunities in the U.S. and beyond.