Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Public Sector Operations Research: A Personal Journey

Richard Charles Larson,

PUBLIC SECTOR OPERATIONS RESEARCH: A PERSONAL JOURNEY

RICHARD CHARLES LARSON

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, rclarson@mit.edu

Not all OR careers are carefully planned from the age of ten.

LARCENY ON THE CHARLES

It all started at an MIT fraternity mixer party in September 1965. Another brother and I, noting the need for additional female attendees—to balance the gender ratio—fulfilled our brotherly duties by bringing over from Boston's Back Bay an additional six partygoers. The problem was that two of the recruits—whom we had not known previously—turned out to be professional thieves. Before we knew it, they had left the party, but not before ransacking the ladies' powder room, taking from purses cash, credit cards, and IDs. That fact did not become apparent until 1:00 A.M., when all the trusting young women went to pick up their purses that they had innocently left in the powder room. When I heard of this, I felt terrible, personally responsible. *Grand larceny*. Little did I know that this one event would change my life and shape my future career.

During the following weeks in an attempt to catch the thieves, I played the role of private detective. I followed up every lead they had given us. One said she went to a secretarial college in the Back Bay of Boston. I talked my way into the president's office, claiming that one of their students had lost a valuable gold watch at that party; as a result I was able to see the photographs of each of the students registered at that college—but no luck. Similarly, I followed all other leads to no avail. Soon in my "Detective Larson" zeal I found myself carrying a half a dozen books in criminology and forensic sciences checked out from the MIT libraries. I became fascinated, for instance, with the potential for computers to automatically recognize latent fingerprints. The following semester I registered in one of MIT's first courses in artificial intelligence to undertake the fingerprint project. One day MIT's Professor Alvin W. Drake noticed me carrying seemingly random library books and asked me what they were. I related to him the story, and he urged me to register with him under a special research project focusing on operations research problems in urban policing. Stunned that I could do such a thing at MIT, I leapt at the opportunity. Professor Drake soon arranged a high-level meeting involving executives at Arthur D. Little, commanders of the Boston Police Department, and yours truly. Thus began my many-year odyssey of riding in the back seats of police cars in Boston, St. Louis, Wilmington (Delaware), Worcester (Massachusetts), and elsewhere. Soon I realized that the police emergency response system was a priority-oriented tandem queueing system. At that time, in the mid-1960s, no one had studied urban policing from an operations research point of view. Wow, there was my master's thesis topic!

THE PRESIDENT'S CRIME COMMISSION

In the spring of 1966, the Director of the MIT Operations Research Center, Professor Philip M. Morse, received a phone call from the Institute for Defense Analyses (IDA) in Arlington, Virginia. The caller asked if there were any faculty or perhaps even students at MIT who had any experience with police departments and/or the criminal justice system. Shortly thereafter, I was invited for a job interview with Dr. Alfred Blumstein at IDA, who had just been appointed Director of the Science Technology Task Force of the President's Commission on Law Enforcement and Administration of Justice. In the summer of 1966, I had the opportunity to serve—as its youngest member—on the task force, working not only with Al Blumstein but also Saul Gass, who was another full-time member of the task force. I guess a career in operations research or any other area is a combination of good preparation, knowing good people, and good luck. Yours truly in this case happened to be in the right place at the right time.

Working with Drs. Blumstein, Gass, and others at the Science and Technology Task Force, I focused on two major areas: the police emergency response system and the total criminal justice system. This led in 1967 to my master's thesis, supervised by Professor Drake, "Operational Study and Police Response System" (Larson 1967). In the same year, Al Blumstein and I co-authored a paper, "A Systems Approach to the Study of Crime and Criminal Justice" (Blumstein and Larson 1967), an edited version of which was published in 1969 in *Operations Research*, "Models

Subject classifications: Judicial legal/crime: police and prosecution models. Professional: comments on. Hypercube queueing model, queue inference engine. Operations research applied to New York City.
Area of review: ANNIVERSARY ISSUE (SPECIAL).

of Total Criminal Justice System" (Blumstein and Larson 1969). "Thrilled" does not describe how I felt about my first paper in *Operations Research*!

NEW YORK CITY-RAND INSTITUTE

In 1967, Mayor John V. Lindsay of New York City negotiated with the RAND Corporation of Santa Monica, California to establish the New York City-RAND Institute (NYCRI). This Institute was to bring the think tank mentality that had been so successfully utilized at RAND for the Air Force and other U.S. Armed Forces to operational problems of the city, in this case New York City (NYC). Dr. Jan Chaiken (MIT Ph.D., mathematics and physics) of the RAND Corporation invited me to Santa Monica for a job interview. As a result of that I had the privilege and honor of serving as a primary consultant with the NYCRI for the entire seven years of its life. This in turn led to many more hours in the rear seats of police patrol cars, in the 911 emergency call center in Manhattan, and in other parts of NYC's criminal justice system. And here, too, I had the opportunity of working with distinguished colleagues including—in addition to Jan Chaiken—Ed Blum, Peter Kolesar, Warren Walker, Art Swersey, and Ed Ignall. That team presented—in baseball vernacular—a New York Yankee level powerhouse of analytical talent focused on NYC's many operational problems. Our work focused on problems of the police, fire, and ambulance services, whereas others focused on housing, welfare, and other domains of NYC's public services. And this work led to implemented results and numerous published papers, some as RAND technical reports and others as journal articles and books (Kolesar 1982, Kolesar and Swersey 1985, Larson 1971a and 1971b, Larson and Stevenson 1972, Larson and Chaiken 1972, Walker et al. 1979). Our analysis of New York's then new 911 system led to significant changes in staffing and scheduling of 911 call operators, implemented totally within four weeks of our study—still a record response time in my experience (Larson 1972a).

The overall work contributed greatly to my MIT Ph.D. research. The doctoral thesis, completed in 1969 and again supervised by Al Drake, was extended and modified significantly to be published as my first book, *Urban Police Patrol Analysis* (Larson 1972b). The following year, in 1973, I was honored to learn that the book was awarded the 1972 Lanchester Prize of the Operations Research Society of America (ORSA) for the best English language publication in operations research.

The NYCRI work received other honors. The 1974 ORSA Lanchester Prize was awarded to the paper, "An Algorithm for the Dynamic Relocation of Fire Companies," written by Peter Kolesar and Warren E. Walker (March–April 1974 issue of *Operations Research* **22**, pp. 249–274). The award citation concludes, "... the Kolesar and Walker paper provides an OR approach that can help save lives and reduce property damage from fires by making it possible for fire companies to respond more rapidly to serious emergencies." For their work with the New York City Fire Department, the team of Homer Bishop, Edward H. Blum, Grace Carter, Edward J. Ignall, Peter Kolesar, Arthur J. Swersey, and Warren E. Walker were awarded honorable mention in the 1974 competition for the Franz Edelman Award of The Institute of Management Sciences (TIMS).

The work of the NYCRI continued through 1975, when because of severe financial problems of the city, the new mayor chose to shut it down. In my mind, this was indeed unfortunate. If the NYCRI had been supported to this day, it most likely would have become the globally premiere R&D laboratory focusing on urban operational problems—building on the techniques of operations research and management science.

The many talented professionals in the NYCRI have gone off to heralded careers, both in operations research and in higher government offices. Jan Chaiken became the director of the Bureau of Justice Statistics. Peter Kolesar became a full professor at Columbia University, and Art Swersey a full professor at Yale University. Warren Walker has had a distinguished career in the public sector, much of it in western Europe. And one of its most talented members, Ed Ignall, died a tragic and premature death in 1979. There is now the Edward J. Ignall Research Fellowship, awarded annually to encourage and help support the research activities of a graduate student in Columbia's Department of Industrial Engineering and Operations Research.

RESEARCH APPLIED TO NATIONAL NEEDS

In the early 1970s, the National Science Foundation, desiring to target some of its research to national needs, created the Research Applied to National Needs (RANN). At MIT, we assembled a multidisciplinary team and drafted a proposal to NSF RANN, which resulted in a 3.5-year research grant called the Innovative Resource Planning Systems for Urban Public Safety Systems (IRP). That $850,000 grant led to a sequence of four published books (Colton 1978, Larson 1978a and 1978b, Willemain and Larson 1978), numerous articles, and many masters and doctoral theses. These include theses of Rick Jarvis (now at Clemson University), Mark McKnew (also at Clemson University), Shan Cretin (USC), Kenneth Chelst (Wayne State University), and many others. At one time, we had 28 people on the payroll of the IRP project.

In addition to the published research results, the project supported Professor Amedeo Odoni and me in writing a textbook on the subject, *Urban Operations Research* (Larson and Odoni 1981). That book is the basis for applied OR courses taught around the United States and in several foreign countries, including a course we have taught at MIT for over 25 years (Larson 1972c)! Arnie Barnett joined us early on during the course as a fellow instructor, and the course as it has evolved over time is as much his as it is Amedeo's and mine. While the book is technically now "out of print," its 560 pages live on—as MIT OpenCourseWare, free to the learner—at URL ⟨http://web.mit.edu/urban_or_book/www/book/⟩.

In the early days of our teaching the Urban OR course, prior to the textbook, we helped students undertake applied OR projects with local governmental agencies. There were lots of good stories surrounding these student projects. My favorite involved three students helping the Town of Framingham, Massachusetts, optimize its school bus system, i.e., minimizing the number of leased school buses and optimizing their routes. In following our suggestion to "get into the system," the students—driving their cars—followed school buses on their routes to learn the realities of permissible U-turns, stopping times, etc. One day they were stopped and interrogated by local police. Reason: There is a Massachusetts law that prohibits the deliberate following of school buses! There again, we got into trouble by diving into the system! The end result of this project was that the students reduced the town's annual outlay for school buses by 20%, but not before upsetting high school football fans (see Larson and Odoni 1981, §8.1.4, p. 537–539)!

HYPERCUBE QUEUEING MODEL

An urban police force responding to calls for service is a complex, spatially distributed multiserver queueing system. Much of our IRP research was directed at this observation, leading to mathematical models that helped planning processes. In designing a "police response system," there are many issues related to queue discipline, implemented at the police emergency communications center, and with spatial deployment of resources, implemented in the planning office and in the field. Queue discipline deals with prioritization and perhaps even deliberately delayed response to low priority callers to leave sufficient resources available for possible high priority incidents that may or may not occur in the near future. In-the-field deployment relates to optimal design of police patrol sectors ("beats"), determining appropriate numbers of police resources to have on duty by time of day and determining response sequences that would minimize travel times. A response sequence is needed because the police unit assigned to patrol the sector containing the call for service may be busy when a call is reported from that sector. A call from sector 7, for instance, may find the "sector 7 car" already busy on a previous call. Then the dispatcher needs an ordered list of cars, say 3, 1, 13; he or she goes down the list and dispatches the first available car that arises next on the list.

When examining the OR literature in the early 1970s for guidance on how to design these kinds of systems, we found references on location theory that did not include the probabilistic aspects of system operation. The fact that the server assigned in a police sector may already be busy on an earlier call was ignored in the literature. From a queueing point of view, the system was quite complex because the workloads of servers differed, due to local geography and variability of 911 work over the service region. We decided to focus initially on the spatial aspects of the police emergency response system, hoping that there may be some relevance in ambulance and also fire response.

The key result of this work was the hypercube queueing model (Larson 1974a), which depicted the state of the multiserver system as a vertex of an N-dimensional unit volume hypercube, where N is the number of police cars deployed in the district. Coordinates of a vertex are given by an N-digit binary number, the ones (zeros) corresponding to police patrol cars that are busy (idle) in that state. Our attention focused on computing in an automated way the upward one-step transition rates λ_{ij}, defined to be the instantaneous Poisson rate of transitions from state i to state j, where i and j are adjacent states on the hypercube. The values of these parameters depend on the particular dispatching strategy that is employed by the police department. We investigated five specific dispatching strategies (Larson 1974b, Larson and Rich 1987), one involving automatic police car location (Larson and Franck 1978), a relatively new concept in the mid-1970s and now a standard feature using GPS technology.

To make the model tractable, we had to assume (negative) exponential service time distributions. But we recognized that the exponential service times were not necessarily valid in practice. So investigation of the model sensitivity to that assumption was a focus of Rick Jarvis' doctoral research. He used the "method of phases," substituting Erlang order K service time distributions for the exponential, preserving the Markov finite state depiction of the process but with many more states. The resulting performance measures, as computed by the generalized model, were so close to those computed by the exponential model that double precision floating point arithmetic was required to spot any differences (Jarvis 1975). Armed with those results, we felt confident that the results of the hypercube queueing model with its exponential service time assumption would be fairly accurate for planning purposes, even in those situations where the actual service time distributions more resembled an Erlang distribution.

Jarvis also investigated optimal dispatch strategies, building from the seminal work of the Carter-Chaiken-Ignall paper, done for the NYCRI (Carter et al. 1972). Jarvis derived an algorithmic generalization of the Carter-Chaiken-Ignall results, extending their results from two servers to an arbitrary N server system. Doctoral student Mark McKnew and I extended the hypercube model further by identifying a third activity state for a police patrol car—"self-dispatched"—working on patrol-initiated activities (Larson and McKnew 1982).

We received calls from U.S. police departments asking about possible implementation of the hypercube queueing model in various cities. Computers of the mid-1970s were quite limited. The state space of the model grows as 2^N, where N is the number of police cars within one district operating as a multiserver cooperating queue. The city of New Haven, Connecticut, for instance, desperately wanted the model but said that they didn't have a computer that could solve an equation set that grew so quickly. In fact, the city of New Haven was structured to have only one district where theoretically all police cars in the city interacted with each other as a 42-car multiserver queueing system. 2^{42} is a large number, even for today's computers. So, as a result, we went to work trying to find an approximation for the hypercube queueing model. The approximation would need to be computationally reasonable, unbiased and sufficiently accurate so that police planners could use it confidently. As a result, we created the so-called hypercube approximation procedure (Larson 1975). This procedure led to a set of N nonlinear simultaneous equations, a huge reduction compared to the 2^N required for the exact model. We tested this with data from Boston and other cities over a period of two years. We found that in almost all cases the computed results from the approximation procedure were unbiased and tended to be within 1% or 2% of the results computed from the more exact model. In all subsequent releases of the hypercube model, it was the hypercube approximation procedure that has been utilized, not the exact model.

The essence of the hypercube approximation procedure is the assumption that each server is available with probability of $1-\rho_j$, where ρ_j is the steady-state fraction of time that server j is busy in a queueing sense. So one might think that the chance that we might have to go to the fourth preferred server for a particular customer would be the product $\rho_7\rho_3\rho_{13}(1-\rho_8)$, where the dispatch sequence is {7, 3, 13, 8, ... }. It turns out that this is "B minus probability reasoning": The simple coin flipping analogy tends to be biased because it assumes independent trials. For instance, even if we know the ρ_is (which we do not), once one samples the most preferred server and finds that server busy, the conditional state probabilities for the system tend to be biased toward more congested states. Therefore the product of ρ'$sx(1-\rho)$ is a biased formula. To correct for the bias, we pretend we were going into an $M/M/N$ queueing system and sampling servers blindfolded without replacement. From this "Gedanken experiment," we create "correction factors" that are placed in front of the product of servers' busy-ness and availability factors to correct for the inherent bias in the simple approach. The theory of sampling servers without replacement seemed to work in terms of eliminating all biases.

The hypercube queueing model, as represented by the hypercube approximation procedure, has lived on at both MIT and off-campus companies with which I have been affiliated. It has been implemented in such cities as Boston, Massachusetts (Brandeau and Larson 1986); Dallas, Texas; Wilmington, Delaware; Caracas, Venezuela; and Orlando, Florida for policing, and in NYC for ambulance deployment. Professor Stephen Sacks (University of Connecticut) has devoted significant time to making the model computer-accessible. As a result, today there exists a version of the hypercube queueing model on PCs linking the geography of a city or region through georelational databases. Professor Sacks and I plan to seek funding support to fully document this version of the hypercube and to make it "industrial strength," to then give it away to whatever public sector agency requests it.

Here is an example of an OR methodology as it has evolved over a nearly 30-year period. The limitation to rapid deployment seemed not to be in the OR mathematics, but rather in the cost performance characteristics of computers, the technology of graphical user interfaces, and the ubiquity of geodatabases. All those factors seem to be coming together now to produce a management tool that may have wide appeal.

SCHOOL BUSSING

In 1974, I received a phone call from the office of W. Arthur Garrity, Jr., a Federal judge who resided in well-to-do suburban Wellesley, Massachusetts. He had just ordered forced bussing of children attending Boston public schools to balance the racial percentages in the schools. He knew that logistically this would be a complex problem conducive to mathematical computer-based treatment. I was asked to be the quantitative expert who would use math and computers to design the bus-routing logistics system. Technically, this was an intriguing linear programming problem. Philosophically, I had considerable difficulty with the judge's order. After a weekend of deliberation, I declined the offer. Decades of tumultuous events and unintended consequences followed, the turmoil not due to poor OR but rather to a significant lack of societal consensus on how to alleviate de facto segregation in the schools.

"PROFESSOR LARSON, I HAVE SOME TERRIBLE NEWS FOR YOU"

The city of New York offered urban fellowships to promising undergraduates throughout the United States on a competitive basis in the late 1970s. Typically, the student would spend an extra year off campus between junior and senior years, working in NYC with a line agency and applying the methodologies they learned in their undergraduate career to operational problems of that agency. Victor Li became MIT's first New York City Urban Fellow. Victor, who had started undergraduate research with me, spent the year in New York working with the New York City Police Department attempting to implement the hypercube queueing model.

In October of that year, about six weeks after he had been in NYC, he called me and said, "*Professor Larson, I have some terrible news for you.*"

I thought, "Oh my God, there's been a tragedy in his family," or some similar incident. I said, "Victor, what is it?"

The conversation went something like this.

"Professor Larson, you know the Manhattan distance metric that you teach in class?"

"Yes, in high school it is also known as a taxi cab metric, and it is a distance metric called L_1, which relates to the fact that the police cars in Manhattan can only travel north/south and east/west, not diagonally as the crow flies. Victor, what is it with the Manhattan distance metric?"

"It doesn't work, not even in Manhattan."

"Victor, is that the extent of the bad news?"

"Yes, Professor Larson. Aren't you disappointed?"

"Victor, why doesn't the Manhattan distance method work in Manhattan?"

"Central Park! Have you ever heard of Central Park?"

"Yes, I've even been there several times, having been born in New York City."

He then related that a police car on the west side of Central Park that is dispatched to the east side of Central Park would have to travel a detour distance around Central Park or through one of its one or two thoroughfares that gets one through the park. Central Park does not have the dense city grid system that characterizes the rest of Manhattan, so Central Park was a barrier or obstacle to travel that increased the travel distance. I related to Victor that this was not a tragedy of massive proportions. What he should do is create a travel time or travel distance matrix that reflected the reality of Central Park and use that in a hard-coded way for his junior/senior year abroad in NYC. When he returned to MIT, perhaps we could focus on the generalization of the problem for his master's thesis. This is exactly what Victor did. He returned to MIT, and in 1979 we co-authored the paper, "Finding Minimal Rectilinear Distance Paths in the Presence of Barriers" (Larson and Li 1981). That was Victor's first authored or co-authored refereed journal paper; he now has more than 200. The polynomial algorithm that we discovered generalizes to a set of barriers in any two-dimensional region, the barriers having arbitrary complexity. They need not be polygonal or convex. Apparently, this work has found its way into the robotics literature.

This "polypath" algorithm is the method of choice now when one creates the travel time or travel distance matrix for the hypercube queueing model. And it can be used for any other purpose in computing travel distances and travel times in a city, for instance, for taxi cabs. My doctoral student, Ghazala Sadiq, then generalized that work as she focused her doctoral thesis work on facility locations utilizing the Manhattan metric in the presence of barriers to travel (Larson and Sadiq 1982). These are additional examples of operations research contributions to the literature motivated by real problems experienced by real agencies operating in the real world. Operations research is indeed research on operations (see Larson 1992).

SPEEDING SERVICE BY INTRODUCING DELAY

Christian Schaack became a doctoral student with me in the early1980s at the MIT Operations Research Center. He and I noticed that we had some unfinished business with regard to accurate OR modeling of the dispatch queue for urban emergency services. Too often in practice, a police or ambulance dispatcher would dispatch the Nth out of N total servers in order to respond quickly to the immediate calls for service, thereby leaving an entire district devoid of available servers. Should an urgent call suddenly arise, that call would have to be delayed in queue, even though one or more of the busy servers might be handling low-priority calls, the proverbial "cat up a tree." There is little or no preemption in urban emergency services. For any reasonable cost function, one finds that an optimal policy is to delay deliberately lower priority calls in queue, even in the presence of available servers, in order to leave the system in a state that can best anticipate near-term urgent calls that may or may not arise. This result is similar to that found in airline yield management: Don't sell those last few airplane seats to discount customers who are on the phone asking for them in anticipation of likely full-fare business customers who have yet to appear. In policing, to speed service to near-term urgent calls, it may be necessary to delay deliberately lower priority calls. Politically, this policy can be difficult to implement because there is always the possibility of incorrectly categorizing a call, that is, a call that one designates as low priority could turn out to be urgent. In that case, it would be a tragedy if death or serious injury resulted from the dispatch delay due to the erroneous categorization. But Christian Schaack and I felt that in the preponderance of cases the advantage of studying such a policy would far outweigh the risks, so then we embarked on a research project studying N-server cut-off priority queues (Schaack and Larson 1986, 1989). Steven Sacks joined us in showing how these results could be applied in practice (Sacks et al. 1993). While we are aware of implementations of the hypercube queueing model, we are not aware of formal implementation of the cut-off priority of the queue models in urban emergency services. But I personally believe that its utility from a potential life-saving point of view far exceeds that of the hypercube queueing model.

PUBLIC SYSTEMS EVALUATION AND ITS DERIVATIVES

The IRP project, whose work was disseminated with a monthly newsletter, led to national publicity and recognition. As a result, we received phone calls from places like St. Louis, Missouri; Worcester, Massachusetts; and Wilmington, Delaware, asking us to play leadership roles in the design and evaluation of experiments and innovations in the police departments in those cities. But our MIT resources were maxxed out. So Richard Lister, then senior member of the National Institute of Justice, suggested that I start an off-campus company that could respond to these types of requests. I had never thought of starting a company, but it seemed to be an intriguing idea. So on the Ides of March, 1974, we started Public Systems Evaluation Inc. (PSE), a not-for-profit applied research firm. The founding vice president was Kent W. Colton, then a doctoral student at MIT, and the treasurer was Gilbert C. Larson, my father. I cannot tell you of the joy I had in reuniting with my father after so many years of separation because of academic work at MIT, IDA, and NYCRI. He and I had the opportunity to work closely together for approximately seven years until his retirement from PSE. We co-authored a refereed journal article that was published in an IEEE journal (Larson, Colton, and Larson 1977), a very special event for a father-son team, both of whom had degrees in electrical engineering.

Shortly after we opened the doors to PSE in March 1974, we had a backlog of $390,000, involving professional commitments in the three cities that were mentioned above. Over the years, PSE morphed into ENFORTH Corporation, which morphed into Queues Enforth Development (QED), which recently has morphed into Structured Decisions Corporation. Shortly after starting PSE, Dr. James M. Tien joined us as executive vice president, and Jim remains (part time) as a Principal to this day. He is also Professor and Department Chair at Rensselaer Polytechnic Institute (RPI). In our first joint project together, Jim and I worked on designating a different civilian type of responder to low priority police calls for service (Tien and Larson 1978); that idea is still quite applicable today. Michael Cahn, another colleague over the years, also joined in the late 1970s and serves as president of the company.

Several well-known operations researchers have spent from six months to two or three years at one of the versions of PSE before going on to the rest of their careers. These include my former doctoral students Ed Kaplan, now a professor at Yale University; Sam Chiu, a professor at Stanford University; Mark McKnew, a professor at Clemson University; Susan Hall, who teaches on occasion at Babson College; Alan Minkoff, a senior researcher at IBM; and Oded Berman, a professor at the University of Toronto. They also include former SM students, Margaret Brandeau, on the faculty of Stanford University, and the aforementioned Victor Li, a professor at the University of Hong Kong. Other PSE-related alumni include Stephen Sacks, a professor of economics at the University of Connecticut, and Octavio Richetta, a professor at University of Massachusetts in Boston.

GARBAGE IN, GARBAGE OUT

In 1981, the officers of Public Systems Evaluation, recognizing a conservative U.S. president, projected that the federal funds available for applied research in U.S. cities would most likely diminish over the decade of the 1980s, but that private sector work in that area perhaps would grow. As a result, the four of us who were Principals at PSE terminated PSE and formed Enforth Corporation. The four cofounders were James M. Tien, Kent W. Colton, Michael F. Cahn, and yours truly. Why *Enforth*? Because each of the cofounders' last names ends with the letter N—there were four Ns, implying N^4, therefore, Enforth.

We soon heard of a request for proposal (RFP) that dealt with the world's largest public landfill, the Fresh Kills Landfill (FKL) in Staten Island, New York. The FKL was a 500-foot-high mountain of garbage on an area that had been pristine sea-level Marine Park in the year 1900. It was about to be featured in *National Geographic* magazine as the largest "mountain" in the mid-Atlantic states that bordered on salt water. Because of environmental regulations and other considerations, all in-city landfills other than Fresh Kills were about to be closed. The city projected that the 10,000 tons of garbage and refuse that had been hauled by barges to FKL on a daily basis would, over the decade of the 1980s, be doubled to 20,000 tons per day. The fleet of 60 barges then being towed by tugboats had to be upgraded and increased in size to accommodate the doubling of daily tonnage. When city management asked operations in the Department of Sanitation what the new fleet size should be, the answer came back in a way that might not surprise many of us. The answer was, "Well, if we must double the daily tonnage and we are currently operating with 60 boats, then we obviously need 120 barges to accommodate the double tonnage." This linear thinking implied that a doubling of demand required a doubling of resources to handle that demand. Apparently the notion of economies of scale had not occurred to those in line operations.

The New York City Department of Sanitation (NYCDOS), in its good wisdom, decided to delay the implementation of a barge fleet growth plan and commissioned the problem to a study by consultants. Seven organizations responded to the RFP, and Enforth Corporation was selected—its first major contract. On the first day of the contract, initiating what was to be a trend in our applied operations research work, three of us flew to NYC and spent the day at FKL. Operations research is research on operations. But little did I know the consequences on one's shoes of spending a day in a major landfill. That night friends had purchased tickets for a Broadway play and I had not brought a change of shoes. And the more you try to wash your shoes to get rid of the landfill aroma, the worse it gets. Most likely, folks within a diameter of 10 people around me in that Broadway play knew where I had spent that day, to my continuing embarrassment. But this behavior marked our tradition from then to the present: The first day of each of our operations research assignments was spent on the scene experiencing the problem first hand, "living in the system."

The result of that work was a computer-implemented planning model, *Barge Operations System Simulator.* (BOSS). BOSS and its descendants lived on in NYC for more than 15 years. The initial implementation of it for the problem at hand, namely determining the optimal fleet size of the barge system, saved the City of New York $20 million (Larson et al. 1988, 1991). It showed that the number of barges that one needed was not 120, but rather 80–90. The NYCDOS acted cautiously here, and they decided to order 100 new barges rather than the 120 suggested by line operations. Because each barge cost one million dollars at the time, this was a direct savings to New York City of $20 million. The project to Enforth Corporation cost the city of New York $100,000. For every dollar invested in that study, New York City saved $200.

A recent analysis involving waste recycling and "resource recovery" demonstrated that the same OR BOSS methods, appropriately modified, helped shape the NYC refuse marine transport system of the 1990s (Richetta and Larson 1997).

FROM GARBAGE TO SLUDGE

Operating as Enforth Corporation, we learned to read the *New York Times* every day. One day we discovered that NYC was confronted with a problem of designing a marine-based sludge removal system. Sludge is the end by-product of primary and secondary wastewater treatment, the wastewater coming from the city's sewage system. In the mid 1980s, New York City had 13 primary and secondary wastewater treatment plants, each located at shore side. Small, dedicated tanker ships picked up the sludge, and the tradition had been to dump it in the New York City bight ⟨http://everest.hunter.cuny.edu/bight/index.html⟩—within eyesight of the Statue of Liberty. This was to be no more, as ordered by the federal EPA. Rather, the EPA had designated for sludge dumping at the "106-mile mark," a rectangular region of 200 square miles located 106 miles south-southeast of the harbor and due east of Atlantic City. I made a cold call to the Wards Island Headquarters of NYC's Department of Environmental Protection (DEP), reported our services in garbage, and said we would next like to focus on sludge. We were entered as a bidder in an RFP process and were ultimately selected as the provider of those services.

On day one of this work, I took a leisurely stroll around the acres of the Ward's Island plant. Later on it was reported that I was observed trespassing in what had been a hard-hat area where visitors must be accompanied by trained staff. So I guess I violated rules and regulations on day one. At least my shoes offered no clue as to where I had been. Our clients in New York felt that one class of tanker vessel should be used to pick up the sludge at the 13 plants and haul it directly out to the 106-mile mark. We suggested that an alternative approach would be to use small in-harbor vessels to pick up the sludge and bring it to one or more transshipment points, and then from the transshipment point(s) larger ocean-going vessels would transport the sludge to the 106-mile mark. Our clients said the transshipment should not be an option, that no analysis would support it. We suggested that we would "toss in" transshipment as an extra (free) option, include it in our analysis, and they could then decide whether to include transshipment or not as a result of our efforts. Our analysis eventually indicated the superiority of the transshipment option. The implemented $40 million capital investment decision was to utilize Ward's Island as the transshipment point in a system using two different kinds of vessels, one for the inner harbor and one for ocean-going purposes.

The work resulted in a new class of strategic inventory routing planning models, since generalized by MIT Ph.D. students Allan Minkoff (Minkoff 1985) and Ian Webb (Webb and Larson 1995). Our NYC model was called SLUDGE, Strategic Logistical Unified Design GEnerator (Larson 1988). In the great majority of cases of major operations research work done at Enforth as well as at PSE, new OR opportunities were identified which were then brought back onto campus, providing rich research materials for masters and doctoral candidates. In this way, our engagement with the real world enriched the on-campus educational experiences of our students and I believe made us better teachers when we were up there at the blackboard reciting not only equations but how they were applied in practice.

AN ARRESTING EXPERIENCE

In the late 1980s, Enforth was selected in an RFP process to undertake a major study called the arrest-to-arraignment (ATA) study in NYC. Numerous complaints of spouses and friends of individuals who had been arrested by the NYC police sparked this study. Their whereabouts could not be determined for up to 72 hours after the initial arrest. After arrest and booking in a neighborhood police precinct stationhouse, the defendant would be transported for overnight accommodations awaiting preliminary arraignment. The "data management system" revealing the nightly locations of these individuals turned out to be a set of cut-off cardboard orange juice quart containers, each container corresponding to a particular police precinct station house. Inside each container would be the names of the individuals locked up there for that particular evening. Some hotel management system!

Keeping with our tradition of living in the system, one of us actually volunteered to simulate the arrest process. NYC professionals videotaped the entire sequence of steps from initial arrest through the entire arrest-to-arraignment process. It was more frightening than *Scary Movie II*! This work led to a very detailed Monte Carlo simulation model implemented on a PC showing the intricate working relationships between the seven different agencies that maneuvered defendants through the ATA system. The recommendations of the project were implemented, as reported by the mayor of NYC in a televised press conference. The work became a finalist for the 1992 Edelman Prize competition of TIMS (Larson et al. 1993). Not winning the prize was surprisingly a very proud moment for me, because the 1992 winner was my former doctoral student Ed Kaplan (with Elaine O'Keefe), *Let the Needles Do the Talking! Evaluating the New Haven Needle Exchange*.

TWO QUEUES IN MANHATTAN

I recall one day when I went to NYC in a three-piece suit and spent the morning in a lower Manhattan office of welfare. There I was in this dingy room, which had 40 seats and windows that showed no sign of recent washing. Welfare mothers with their children were waiting in queue for two to three hours to be seen for services. I was there to help the Department of Welfare with their very visible and real queueing problem. At noon, I hailed a cab and went uptown to the headquarters of Citibank, to spend the afternoon on teller and ATM queueing problems for Citibank. The irony that crossed my mind in that cab ride was that the same multiserver queueing equations were applicable in the morning to the welfare department and in the afternoon at Citibank. I wore my three-piece suit for Citibank. I don't know what the appropriate clothing would have been if I had been visiting only the Welfare Department, but they certainly needed our services more than the fairly efficiently run Citibank did. They still do to this day.

During the decade of the 1980s, Enforth undertook about a dozen major studies with NYC, working additionally with the police department, the fire department, emergency medical services, the probation department, the corrections department, and the Office of Management and Budget. In a sense, I believe we replaced for NYC the role formerly played by the NYCRI.

Space does not allow us to go through all the different studies we did for NYC in the 1980s. But that decade remains a most precious one in my mind for work that we did, work that had substantial impact and resulted in publishable papers, both directly from Enforth professionals and from doctoral and master's students from MIT. The last major study we did during this phase was implementation of the hypercube queueing model to locate and dispatch ambulances throughout the five boroughs of NYC.

THE BIG APPLE GETS THE ORSA PRIZE

During the 1970s and 1980s, we became quite close to senior professionals in NYC's Office of Management and Budget (OMB). OMB began to see the benefits of operations research as applied to complex city operations. We encouraged them to build an in-house capability to do much of this themselves. It is with mixed sadness and pride that we report that they did indeed do this. They built an excellent in-house staff to do operations research analyses of the myriad operational issues and problems of NYC, an effort that culminated in the OMB OR staff being awarded the heralded ORSA Prize in 1993. The citation reads: "For their demonstrated effectiveness in applying Operations Research for the evaluation of costs of providing important city services, the Office of Management and Budget of the City of New York has made Operations Research an integral part of the budget planning process and of evaluating the City's public services."

QUEUES, LIMITED

While it was not initially a public sector application, in the late 1980s I was doing research at MIT under NSF support on multi-attribute aspects of queueing. A framework for this work first appeared in *Operations Research* in 1987 (Larson 1987), a paper that took on a life of its own, first shortened and republished in MIT's *Technology Review*, then appearing as a full-page cover story in the Sunday Business Section of the *New York Times*. All this gave me my eight minutes of national fame in an airing of ABC TV's program 20/20, where they dealt with the psychology of queueing after two segments on Khmer Rouge atrocities in Cambodia.

Boston-based BayBanks asked me, as part of the NSF funding research, to take a look at their problem of sizing ATM (automatic teller machine) sites. For some ATM sites, they felt they needed more machines because the monthly usage was too high. For other sites they felt as if they might have one or two machines too many and could save money by removing them or replacing them by limited function ATMs. Because operations research is research on operations, this was just the kind of assignment that we craved. So as dutiful operations researchers, we asked them for data. A month later they gave us 12 inches of full-page computer printouts containing detailed transactional information from several ATMs. At first, what I was going to do was aggregate the data to find λ_t, that is, a customer arrival rate at each ATM by time of day t, estimate a service time distribution, and apply traditional textbook steady-state queueing theory.

Looking at the data, I figured that we had a data set that I had not personally seen before. We had the time of service initiation and the time of service completion for each customer served. By examining this closely we could figure out which customers had been delayed in queue and which ones hadn't. So I asked the question if there might be more information in this data set than one customarily gets from aggregate estimates of λ_t. Using concepts of order statistics, we soon discovered something that we labeled the *Queue Inference Engine* (QIE). This was published in Larson (1990) and gave rise to a cottage industry in the study of queue inferencing, involving such other authors as Dimitris Bertsimas, D. J. Daley, Lee Jones (Jones and Larson 1995), (S. H. Patel, A. Mandelbaum, Les Servi, and S. Zeltyn.

What the queue inference engine allows one to do from a data set containing the start time and end time of service of every customer served is to deduce in a backward way the transient queueing behavior of the queue as experienced by the customers. One can get the state probabilities of the system over time, and from that the conditional distribution and moments of the waiting time of each customer served, by time of day. When we first discovered this I personally thought this was an error. I did not believe that much information could be deduced from such a data set. But intuition was wrong here, and not for the first time!

We transitioned the algorithm through the MIT Licensing Office and spun off a company called Queues, Limited. Queues, Limited embarked on a major contract with Citibank of New York. Soon I discovered it was quite difficult to be associated with two companies off campus, be a dedicated full professor at MIT, and also a dedicated father and husband. So we merged Queues, Limited with Enforth Corporation in 1991, creating QED (Queues Enforth Development, Inc.) With Phase I and Phase II Small Business Innovation Research (SBIR) support from the NSF, we created from the QIE a commercial product called *Swiftline*™.

This product was initially directed at the private sector, which is not the topic of this paper. But the U.S. Postal Service (USPS), which is part of the public sector, took note of the QIE and contracted with QED to implement Swiftline in two post offices in North Carolina in 1994. At the same time on the MIT campus, Susan Hall was completing her Ph.D. work (Hall 1992). She added greatly to the mathematics of the QIE under circumstances in which one has limited instrumentation—providing real-time partial information about the queue (e.g., when the queue length exceeds 10 feet). Combining Susan Hall's work with Swiftline allowed the USPS to obtain much improved statistics on the waiting time performance of customers at post offices around the country. At that time in the mid-1990s the post office infrastructure in terms of data collection and aggregation and storage was not sufficiently advanced to support widespread implementation of Swiftline. As I write this in the year 2001, that infrastructure has been upgraded, the USPS has committed to a concept "*Service in Five*" (meaning they aim to serve every customer in five minutes or less) and a discussion of a widely implemented Swiftline is now very much in the picture.

QED also undertook an extensive amount of work for the USPS in the area of manpower planning and management. Among other things, this required yours truly to testify in public under oath before a five-person compulsory arbitration board to secure an increase in the number of USPS personnel who could be scheduled in a flexible rather than a rigid manner. It also gave rise to a sequence of published papers with my long-term colleague and friend Oded Berman (Berman and Larson 1993, 1994, and Berman et al. 1997).

STRUCTURED DECISIONS CORPORATION

By the late 1990s, QED had focused a large fraction of its work in the public sector, particularly the design, creation, implementation, and support of computer aided dispatch systems and records management systems for police, fire, and ambulance services. By the year 2000, QED had over 120 of these systems installed at over 65 sites in 12 states in the United States. These systems were supported 24/7, as their robust 24-hour operation was required literally to save lives. During the year 2000, QED's 25–30 professionals recognized that the company had really become two companies. One, the original one, which focused on intriguing and previously unsolved OR problems, and the second one, a production-oriented software company whose customers required 24/7 support. Eventually, we grew to recognize, perhaps reluctantly, that growth opportunities for each part would be best served by splitting the company in two. The first, QED, would stay with software products for police, fire, and ambulance services. The second, the birth of a new company, Structured Decisions Corporation, would be the OR firm that has existed since the Ides of March, 1974. So that decision was made and implemented on January 1, 2001.

As I write this, Structured Decision Corporation has a portfolio of clients from both the public and private sectors. We find that much of the public sector research, especially the OR methodologies involved, extend over into private sector operations. Current work at Structured Decisions Corporation in the public sector includes major projects with the U.S. Department of Justice and continuing work with the United States Postal Service, particularly with Swiftline. Also at the present time, as mentioned above, Steven Sacks and I will be seeking support to create a desktop graphical user interface (GUI) version of the hypercube queueing model that can be literally given away to police, fire, and ambulance services throughout the United States. So the public sector work moves on. But it also moves on through our students. Seeing their progress and their impact is the most rewarding part of our work.

SUGGESTED MODUS OPERANDI

In going through these personal reminiscences, I have tried to offer some ways to approach operations research, especially for students:

- Don't worry if you don't have a career plan. Be open and be willing to take moderate risks, then opportunities will come to you.
- Operations research is research on operations. It is not about writing the 972nd journal article about a problem that was first identified in 1947. Which is better: Writing that 972nd article on someone else's problem, or writing the first article on a new major problem that you have identified and solved?
- OR is not applied math. OR is all about broad-based problem solving, much of it admittedly quantitative, but also involving interpersonal skills, psychology, keen observation, and data analysis know-how.

Oh, the 1965 grand larceny—never solved. My partner fraternity brother was Eric Cosman, who went on to become a professor of physics at MIT and now as emeritus professor, a successful entrepreneur. If I could meet today those 1965 Ford-Fairlane-driving professional thieves, I would take them to dinner and thank them for starting a career richer than any I could ever have possibly imagined.

EPILOGUE

Over the years, I have witnessed first-hand the immense bravery and professionalism of the men and women of New York City's emergency services: police, fire, and emergency medical (now merged with fire). The paper you have just read was accepted for publication on September 4, 2001. On September 11, our nation's history was changed forever. A criminal band of suicidal zealots perpetrated outrageous crimes against humanity, with two of four hijacked airplanes being deliberately crashed into the twin towers of the World Trade Center in New York City. Approximately 3,000 innocent people were killed, from over 60 countries—representing all races, religions, and ethnicities.

Brave men and women of the New York City fire and police departments responded quickly to the first calls; the key goal: saving human lives. They saved many by their rapid and effective response. But when the twin towers suddenly fell, the NYPD lost 23 police men and women and the NYFD lost an astonishing 343 firefighters and paramedics. I join many millions of other Americans and hundreds of millions worldwide to extend my deep and heartfelt sympathy and condolences to the victims and their families.

REFERENCES

Berman, O., R. C. Larson. 1993. Optimal workforce configuration incorporating absenteeism and daily workload variability. *Socio-Econom. Planning Sci.* **27**(2) 91–96.

——, ——. 1994. Determining optimal pool size of a temporary call-in work force. *Eur. J. Oper. Res.* **73** 55–60.

——, ——, Edieal Pinker. 1997. Scheduling workforce and workflow in a high volume factory. *Management Sci.* **43**(2) 158–172.

Blumstein, A., R. C. Larson. 1967. A systems approach to the study of crime and criminal justice. P. M. Morse, ed. *Operations Research for Public Systems*. MIT Press, Cambridge, MA.

——, ——. 1969. Models of a total criminal justice system. *Oper. Res.* **17**(2) 199–232. Republished in *Analysis of Public Systems*, 1972. Republished in *Systems Approach and the City*, 1972, M. D. Mesarovic and A. Reisman, eds., American Elsevier, New York. Republished in *The Economics of Crime and Law Enforcement*, 1976, L. R. McPhetors and W. B. Stronge, eds. C. C. Thomas, Springfield, IL.

Brandeau, M., R. C. Larson. 1986. Extending and applying the hypercube queueing model to deploy ambulances in Boston. A. Swersey and E. Ignall, eds. *Delivery of Urban Services*. North Holland, New York.

Carter, G. M., J. M. Chaiken, E. Ignall. 1972. Response areas for two emergency units. *Oper. Res.* **20** 571–594.

Colton, K. W., ed. 1978. *Police Computer Technology.* Lexington Books, Lexington, MA.

Hall, S. 1992. New directions in queue inference for management implementations. Ph.D. thesis, MIT, Cambridge, MA.

Jarvis, J. P. 1975. Optimization in stochastic service systems with distinguishable servers. TR-19-75, MIT Operations Research Center, Cambridge, MA.

Jones, Lee K., R. C. Larson. 1995. Efficient computation of probabilities of events described by order statistics and applications to queue inference. *ORSA J. Comput.* **7**(1) 89–100.

Kolesar, P. 1982. Ten years of research on the logistics of urban emergency services. J. P. Brans, ed. *Operational Research '81*. North Holland, New York.

——, A. Swersey. 1985. The deployment of urban emergency units: A survey. *TIMS Stud. Management Sci.* **22**.

Larson, R. C. 1967. Operational study of the police response system. MIT Operations Research Center Technical Report #26, Cambridge, MA.

——. 1971a. Measuring the response patterns of New York City police patrol cars. New York City Rand Institute, R-673-NYC/HUD.

——. 1971b. Responses of emergency units: The effects of barriers, discrete streets, and one-way streets. New York City Rand Institute, R-675-HUD.

——. 1972a. Improving the effectiveness of New York City's 911. A. W. Drake, R. L. Keeney, P. M. Morse, eds. *Analysis of Public Systems*. MIT Press, Cambridge, MA.

——. 1972b. *Urban Police Patrol Analysis*. MIT Press, Cambridge, MA.

——. 1972c. A new MIT graduate course: Analysis of urban service systems. *Urban Anal.* **1** 105–113.

——, J. Chaiken 1972. Methods for allocating urban emergency units: A survey. *Management Sci.* **19**(4) 110–132.

——, K. Stevenson. 1972. On insensitivities in urban redistricting and facility location. *Oper. Res.* **20** 595–612.

——. 1974a. A hypercube queueing modeling for facility location and redistricting in urban emergency services. *J. Comput. Oper. Res.* **1**(1) 67–95.

——. 1974b. Illustrative police sector redesign in district 4 in Boston. *Urban Anal.* **2**(1) 51–91.

——. 1975. Approximating the performance of urban emergency service systems. *Oper. Res.* **23**(5) 845–868.

——, K. W. Colton, G. C. Larson. 1977. Evaluating a police-implemented AVM system: The St. Louis experience (Phase 1). *IEEE Trans. Vehicular Tech.* **VT-26**(1) 60–70.

——, ed. 1978a. *Police Accountability: Performance Measures and Unionism*. Lexington Books, Lexington, MA.

——, ed. 1978b. *Police Deployment: New Tools for Planners*. Lexington Books, Lexington, MA.

——, E. Franck. 1978. Evaluating dispatching consequences of automatic vehicle location in emergency services. *J. Comput. Oper. Res.* **5** 11–30.

——, V. O. K. Li. 1981. Finding minimum rectilinear distance paths in the presence of barriers. *Networks* **11** 285–304.

——, A. R. Odoni. 1981. *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, NJ.

——, M. A. McKnew. 1982. Police patrol-initiated activities within a system queueing model. *Management Sci.* **28**(7) 759–774.

——, G. Sadiq. 1982. Facility locations with the Manhattan metric in the presence of barriers to travel. *Oper. Res.* **31**(4) 652–669.

——, 1987. Perspectives on queues: Social justice and the psychology of queueing. *Oper. Res.* **35**(6) 895–905.

——, T. Rich. 1987. Travel time analysis of New York City police patrol cars. *Interfaces* **17**(2) 15–20.

——. 1988. Transporting sludge to the 106-mile site: An inventory/routing algorithm for fleet sizing and logistics system design. *Transportation Sci.* **22**(3) 186–198.

——, P. Gregory, A. Minkoff. 1988. Fleet sizing and dispatching for the marine division of the New York City department of sanitation. An invited chapter, in B. L. Golden, A. A. Assad, eds. *Vehicle Routing: Methods and Studies* North Holland Press. (Updated and modified version published in *Waste Management Research* 1991. **9**, 267–276, A computer simulation model for fleet sizing for the marine division of the New York City department of sanitation.)

——. 1990. The queue inference engine: Deducing queue statistics from transactional data. *Management Sci.* **36**(5) 586–601.

——. 1992. Curriculum corrections: Teaching operations research as research on operations. *OR/MS Today* 36–40.

——, Michael F. Cahn, Martin C. Shell. 1993. Improving the New York City arrest-to-arraignment system. *Interfaces* **23**(2) 76–96.

Minkoff, A. 1985. Real-time dispatching of delivery vehicles. Ph.D. thesis, MIT, Cambridge, MA.

Richetta, O., R. C. Larson. 1997. Modeling the increased complexity of New York City's refuse marine transport system. *Transportation Sci.* **31** 272–293.

Sacks, Stephen R., Richard C. Larson, Christian Schaack. 1993. Minimizing the cost of dispatch delays by holding patrol cars in reserve. *J. Quant. Criminology* **9**(2) 203–224.

Schaack, C., R. C. Larson. 1986. An N server cutoff priority queue. *Oper. Res.* **34**(2) 257–266.

——, ——. 1989. An N server cutoff priority queue where arriving customers request a random number of servers. *Management Sci.* **35**(5) 614–634.

Tien, J., R. C. Larson. 1978. Police service aides: Paraprofessionals for police. *J. Criminal Justice* **6**(2) 117–131.

Walker, W., J. Chaiken, E. Ignall, eds. 1979. *Fire Department Deployment Analysis*. North Holland Press, New York.

Webb, I., R. C. Larson. 1995. Period and phase of customer replenishment: A new approach to the strategic inventory/routing problem. *Eur. J. Oper. Res.* **85**(1) 132–148.

Willemain, T. R., R. C. Larson, eds. 1977. *Emergency Medical Systems Analysis*. Lexington Books, Lexington, MA.