Richard E. Bellman Heritage Award Acceptance Speech

I feel honored and grateful for this award. After spending so much time on dynamic programming and writing books about its various facets, receiving an award named after Richard Bellman has a special meaning for me.

It is common in award acceptance speeches to thank one's institutions, mentors, and collaborators, and I have many to thank. I was fortunate to be surrounded by first-class students and colleagues, at high-quality institutions, which gave me space and freedom to work in any direction that I wished to go. As Lucille Ball said, "Ability is of little account without opportunity."

Also common when receiving an award is to chart one's intellectual roots and journey, and I will not depart from this tradition. It is customary to advise scholarly Ph.D. students in our field to take the time to get a broad many-course education, with substantial mathematical content, and special depth in their research area. Then upon graduation, to use their Ph.D. research area as the basis and focus for further research, while gradually branching out into neighboring fields, and networking within the profession. This is good advice, which I often give, but this is not how it worked for me at all!

I came from Greece with an undergraduate degree in mechanical engineering, got my M.S. in control theory at George Washington University in three semesters, while holding a full-time job in an unrelated field, and finished, two years later, my Ph.D. thesis on control under set membership uncertainty at the Massachusetts Institute of Technology (MIT). I benefited from the stimulating intellectual atmosphere of the Electronic Systems Laboratory (later LIDS), nurtured by Mike Athans and Sanjoy Mitter, but because of my short stay there, I graduated with little knowledge beyond Kalman filtering and LQG control. Then I went to teach at Stanford in a department that combined mathematical engineering and operations research (in which my background was rather limited) with economics (in which I had no exposure at all). In my department, there was little interest in control theory and none at all in my thesis work. Never having completed a first course in analysis, my first assignment was to teach to unsuspecting students optimization by functional analytic methods from David Luenberger's wonderful book. The optimism and energy of youth carried me through, and I found inspiration in what I saw as an exquisite connection between elegant mathematics and interesting practical problems. Studying David Luenberger's other works (including his nonlinear programming book) and working next door to him had a lasting effect on me. Two more formative experiences at Stanford were studying Terry Rockafellar's *Convex Analysis* book (and teaching a seminar course from it), and most importantly teaching a new course on dynamic programming, for which I studied Bellman's books in great detail. My department valued rigorous mathematical analysis that could be broadly applied and provided a stimulating environment where both could thrive. Accordingly, my course aimed to combine Bellman's vision of wide practical applicability with the emerging mathematical theory of Markov decision processes. The course was an encouraging success at Stanford and set me on a good track. The course has survived to the present day at MIT, enriched by subsequent developments in theoretical and approximation methodologies.

After three years at Stanford, I taught for five years in the quiet and scholarly environment of the University of Illinois. There I finally had a chance to consolidate my mathematics and optimization background, through research to a great extent. In particular, it helped a lot that, with the spirit of youth, I took the plunge into the world of the measure-theoretic foundations of stochastic optimal control, aiming to expand the pioneering Borel space framework of David Blackwell, in the company of my then-Ph.D. student Steven Shreve.

I changed again direction by moving back to MIT, to work in the then-emerging field of data networks and the related field of distributed computation. There I had the good fortune to meet two colleagues with whom I interacted closely over many

years: Bob Gallager, who coauthored with me a book on data networks in the mid-1980s, and John Tsitsiklis, who worked with me first while a doctoral student and then as a colleague, and over time coauthored with me two research monographs on distributed algorithms and neurodynamic programming, and a probability textbook. Working with Bob and John and writing books with them was exciting and rewarding and made MIT a special place for me.

Nonetheless, at the same time I was getting distracted by many side activities, such as books in nonlinear programming and dynamic programming, getting involved in applications of queueing theory and power systems, and personally writing several network optimization codes. By that time, however, I realized that simultaneous engagement in multiple, diverse, and frequently changing intellectual activities (while not recommended broadly) was a natural and exciting mode of operation that worked well for me and also had some considerable benefits. It stimulated the cross-fertilization of ideas and allowed the creation of more broadly integrated courses and books. It also worked well for me to write textbooks and teach from them. It proved to be the best way to understand a subject in depth, and after a while it became a familiar and enjoyable activity.

In retrospect I was very fortunate to get into methodologies that eventually prospered. Dynamic programming developed perhaps beyond Bellman's own expectation. He correctly emphasized the curse of dimensionality as a formidable impediment in its use but probably could not have foreseen the transformational impact of the advances brought about by reinforcement learning, neurodynamic programming, and other approximation methodologies. When I got into convex analysis and optimization, it was an emerging theoretical subject, overshadowed by linear, nonlinear, and integer programming. Now, however, it has taken center stage thanks to the explosive growth of machine learning and large-scale computation, and it has become the linchpin that holds together most of the popular optimization methodologies. Data net-works and distributed computation were thought promising when I got involved, but it was hard to imagine the profound impact they had on engineering, as well as the world around us today. Even set membership description of uncertainty, my Ph.D. thesis subject, which was totally overlooked for nearly 15 years, eventually came to the mainstream and has connected with the popular areas of robust optimization, robust control, and model predictive control. Was it good judgment or fortunate accident that steered me toward these fields? I honestly cannot say. Albert Einstein wisely told us that "Luck is when opportunity meets preparation." In my case, I also think it helped that I resisted overly lengthy distractions in practical directions that were too specialized, as well as in mathematical directions that had little visible connection to the practical world.

An academic journey must have companions to learn from and share with, and for me these were my students and collaborators. In fact it is hard to draw a distinction, because I always viewed my Ph.D. students as my collaborators. On more than one occasion, collaboration around a Ph.D. thesis evolved into a book, as in the cases of Angelia Nedic and Asuman Ozdaglar, or into a long multiyear series of research papers after graduation, as in the cases of Paul Tseng and Janey Yu. I am very thankful to my collaborators for our stimulating interactions and for all that I learned from them. They are many and I cannot mention them all, but they were special to me and I was fortunate to have met them. I wish that I had met Richard Bellman; I only corresponded with him a couple of times (he was the editor of my first book on dynamic programming). I still keep several of his books close to me, including his scintillating and highly original book on matrix theory. I am also satisfied that I paid part of my debt to him in a small way. I have used systematically, for the first time I think in a textbook in 1987, the name "Bellman equation" for the central fixed point equation of infinite-horizon discrete-time dynamic programming. It is a name that is widely used now, and most deservedly so.

Dimitri Bertsekas