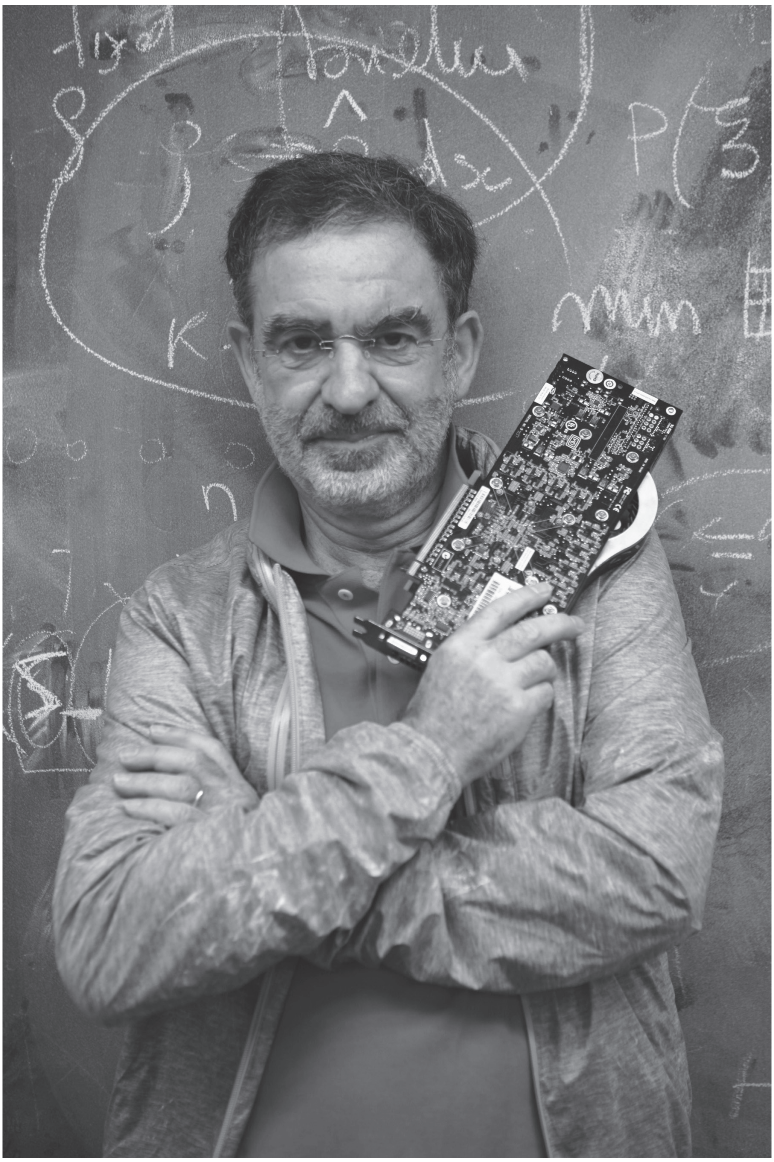
The History of Neuroscience in Autobiography Volume 8

Edited by Larry R. Squire
Published by Society for Neuroscience
ISBN: 978-0-615-94079-3

Tomaso A. Poggio

pp. 362–414

https://www.doi.org/10.1523/hon.008009



Tomaso A. Poggio

Born:

Genova, Italy
September 11, 1947

Education:

University of Genoa, PhD in Physics, Summa cum laude (1971)

Appointments:

Wissenschaftlicher Assistant, Max Planck Institut für Biologische Kybernetik, Tubingen, Germany (1978)
Associate Professor (with tenure), Department of Psychology and Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1981)
Uncas and Helen Whitaker Chair, Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology (1988)
Eugene McDermott Professor, Department of Brain and Cognitive Sciences, Computer Science and Artificial Intelligence Laboratory and McGovern Institute for Brain Research, Massachusetts Institute of Technology (2002)

Honors And Awards (Selected):

Otto-Hahn-Medaille of the Max Planck Society (1979)
Member, Neurosciences Research Program (1979)
Columbus Prize of the Istituto Internazionale delle Comunicazioni Genoa, Italy (1982)
Corporate Fellow, Thinking Machines Corporation (1984)
Founding Fellow, American Association of Artificial Intelligence (1990)
Fellow, American Academy of Arts and Sciences (1997)
Foreign Member, Istituto Lombardo dell'Accademia di Scienze e Lettere (1998)
Laurea Honoris Causa in Ingegneria Informatica, Bicentenario dell'Invezione
della Pila, Pavia, Italia, March (2000)
Gabor Award, International Neural Network Society (2003)
Okawa Prize (2009)
Fellow, American Association for the Advancement of Science (2009)

Tomaso Poggio began his career in collaboration with Werner Reichardt quantitatively characterizing the visuomotor control system in the fly. With David Marr, he introduced the seminal idea of levels of analysis in computational neuroscience. He introduced regularization as a mathematical framework to approach the ill-posed problems of vision and—more importantly—the key problem of learning from data. He has contributed to the early development of the theory of learning—in particular introducing the mathematics of radial basis functions (RBF)—and has supervised learning in reproducing kernel Hilbert spaces (RKHSs) and stability. In the last decade, he has developed an influential quantitative model of visual recognition in the visual cortex, recently extended in a theory of sensory perception. He is one of the most cited computational scientists with contributions ranging from the biophysical and behavioral studies of the visual system to the computational analyses of vision and learning in humans and machines.

Tomaso A. Poggio

Brains, Minds, and Machines

Most neuroscientists these days study the brain to understand its disorders. Of course, this is important—and is funded by the National Institutes of Health (NIH)—but from this point of view, the Year of the Brain and the present Brain Initiative are not very different from the Year of the Heart and the Colon Initiative. Instead, I am interested in the brain because it generates what philosophers call the mind and what we call intelligence. Thus, this is the autobiography of a hybrid neuroscientist and computer scientist working to understand mind and intelligence in brains and computers.

Early Years in Europe

Childhood in Genova

I was born in Genova (Genoa), a harbor city in Liguria with a long history stretching back to Roman times. Genoa was one of the four *Repubbliche Marinare* that dominated the Mediterranean Sea between 1100 and 1400, and the city was at war with Venice for several centuries. Marco Polo's *Il Milione* was written while he was in a prison in Genoa after being captured in one of the many naval battles between Genoa and Venice. "La Via Nuova" in Genoa is the best fully preserved Renaissance street in Europe. (Unfortunately it was later renamed Via Garibaldi with the typical bad taste of the Risorgimento.) It was a new street in 1500 at the core of what was a real estate development—probably one of the first in history—by Andrea Doria. La Via Nuova is just outside the old medieval center, which is probably the largest in Europe and which grew around the original Roman *castrum*. Genoa was the place where bonds were invented and used by Genoa bankers to finance various kings of Europe and their wars. Genoa is also the place where marine insurance has what are probably the world's oldest records. At the Banco of San Giorgio, where many of the old documents are kept, you will find that, in 1300, insuring the transportation of merchandise from Genova to Barcelona was equivalent to what it is today, despite all the pirates who then infested the Mediterranean. During my childhood in the 1950s, Genova was recovering from the destruction of the war; it was an industrial city with a busy harbor and nearby oil refineries. Minor oil spills made their way to the rocks on the beaches in the Riviera and were a regular annoyance of summer swims.

My father owned a textile industry in the outskirts of Genova and a marine supply magazine in the *vicoli* of the old city, next to San Matteo, the family church of the famous Genoa family—Doria. My maternal grandfather was an entrepreneur who started multiple industries in Italy, Argentina, and Brazil. His company in Genoa, Olio Moro, produced and distributed olive oil and tuna cans. We have no ancestors in academia on either side of the family, though my grandfather, Tommaso Moro, improbably claimed Thomas Moore as a distant relative. At the time when I began reading books about mathematics and physics, my grand father advised me to not waste too much time in study. Eventually, I built model airplanes and somewhat dangerous rockets in the big garden of his old country house in Quarto. Despite his opinion of my studies, I have surely inherited my curiosity from him. He had traveled a great deal in Europe and especially in South America.

Absent scholarly traditions were replaced by various encyclopedias, among them the *Enciclopedia Treccani*, an Italian version of the *Encyclopedia Britannica*, which became my main source of information. Through these volumes I developed a fascination with physics, especially the theory of relativity. Einstein was my main hero. The question of what made him such a genius became the focus of my curiosity. This early curiosity developed into a burning interest in intelligence—how we could improve it, how the brain works, and how to make intelligent machines that could solve the scientific problems that excited me, such as the possibility of time travel.

At about the same time, I read the short story "Flowers for Algernon," which had just come out. It is a story about the eponymous Algernon, a laboratory mouse that had undergone brain surgery intended to increase his intelligence. The story is told in a series of progress reports written by Charlie Gordon, the first human test subject for the surgery and a moron. The mouse and the moron become progressively smarter and then their intelligence degenerated. I thought that this was the ultimate human tragedy: achieving genius-level intelligence and understanding exactly what was happening while slowly becoming stupid again.

School—Jesuit Arecco

Later, I had the opportunity to study with exceptional teachers, not only for Greek and Latin, but also for mathematics and physics. The Arecco was a Jesuit school in Genova, which I attended from age six to eighteen—elementary school up to la *maturità classica*. Most school hours were devoted to Latin, Greek, and Italian literature every day except Wednesday and Saturday. The time spent on math, physics, and biology was only two hours a week for each, but the quality of the science and math teachers (Padre Vergnano and Padre Spessa, both Jesuits who were also teaching at the University of Genoa) made up for the scarcity of time.

I read Plato in his original words, enjoyed the easy translations from Caesar's *De Bello Gallico* and compared them to the Asterix cartoons, wrote essays on the *Divina Commedia*, and absolutely loved Kant. Sports were also very important. There was some soccer at school but much more tennis. Later, sailing at the yacht club absorbed my time and my dreams. The yacht club site was the oldest in Italy and, of course, of English origin and tradition. It was tucked between the big ships at the entrance of the commercial harbor, and I managed to capsize in front of big oil tankers more than once. The Genoa that I remember seeing from the windows of my house was really the port, with the big ships waiting beyond the dam. I remember scrutinizing the weather to see what it would be like for that day's regatta. Sailing became my passion during those years. My father, who had driven racecars in his youth and had held some world's record for speed, introduced us to sailing to prevent us from being stricken with his racecar passion, which he deemed to be too dangerous. I went boating with my yacht club friends and with my brother. We were the crew of our Flying Junior and later of a 470, an Olympic class. We often won regattas, though I remember how badly we lost an Italian championship in the Bay of Naples despite being the pre-race favorites. Since my teenage years I have never believed more than 50 percent of what I read in the newspapers. The critical experience was reading reports of sailing races, in which I competed, in the sports section of the main newspaper in Genova, *Il Secolo XIX*: I knew exactly what had happened in critical moments of the race while the reporter was on the beach or on a far-away jury boat!

Physics at the University of Genoa

I passed the *maturità*, the Italian baccalaureate, with pretty good grades, especially in philosophy, and managed to enthrall my examiner with some far-fetched discussions of the relationship between Kant and Popper. I am still curious about what that idea was because I certainly cannot reproduce it. After a summer traveling to Spain and England, I decided to enroll in the physics program at the University of Genoa. I did not want to be an entrepreneur or a politician or a bureaucrat. I was interested in information (which at the time meant electrical engineering), and I was interested in the brain (which meant biology). But engineering looked boring and the biology department was just ancient zoology. Molecular biology had not yet arrived in Genoa. In retrospect, I think physics was the right choice. Physics in Italy had good teachers and a long tradition of excellence: one of my teachers had been a student of Enrico Fermi. Physics was and still is a broad introduction to applied mathematics and many deep and useful scientific tools.

After one year of physics, I became fascinated by mathematics. Professor Pucci and his assistant, Talenti, almost convinced me to switch from physics to mathematics. Toward the end of my studies, four years later, I had to make up my mind. I decided on Antonio Borsellino as an advisor, who was a particle physicist converted to biophysics. During my final years, Borsellino invited me to attend a biophysics conference that he had organized in Erice, Sicily, a magic village on a mountain near Trapani, where—according to mythology—Venus was born. Artificial membranes, lipid bylayers in particular, were the main topic of the conference. This was too close to biochemistry for me, so for my doctoral thesis I decided to work on coherent optics and the potential to store and process information this way. Dennis Gabor of Imperial College had just won the Nobel Prize in Physics for his invention of holography. Antonio took me to see Gabor, whom he knew, in his retirement house on the coast near Rome. Despite such august contacts, my thesis did not amount to much, partly because I was still in the regime of the old Italian *laurea*. Very little time was available for thesis work. My laurea in physics was recognized officially as a European-wide PhD in Germany, but it was certainly less arduous than a PhD today. My thesis, however, included one chapter that established a formal correspondence between the mathematics of holography and the mathematics of correlation. The correlation model had been proposed a decade earlier by another friend of Antonio Borsellino, Werner Reichardt, founder and director of the Max Planck Institute für Biologische Kybernetik in Tübingen, Germany. Antonio told me to apply for a European Molecular Biology Organization (EMBO) fellowship and to visit Werner Reichardt and his new institute. The fellowship was awarded, and I went to visit Tübingen for two weeks to meet Werner.

Tübingen and the Max Planck Institute

Arriving in Tübingen and at the Max Planck Institute was discovering a paradise of science. There were roses around the institute and a gardener taking care of them. Werner's house had a beautiful view and a beautiful garden. Germany at the time was rich and socially fair. The gardener, who also was the jack-of-all-trades at the institute, had a Mercedes, produced in nearby Stuttgart, that was larger than Werner's, the institute director. The institute and its roses mesmerized me. And of course, I was impressed by Werner, the person to whom I most owe my scientific career.

Werner Reichardt was born in 1924 in Berlin. During his school years, he worked in the laboratory of Hans Heinrich Hollmann, one of the pioneers in the electronics of the day.

At the beginning of World War II, Werner was assigned to work on radio projects in the air force. He was 19 years old when both of his parents were killed in an air attack on Berlin. He became a member of a resistance group and was arrested by the Gestapo; condemned to be executed, he was able to escape from jail. In postwar Berlin, Werner managed to set up a radio repair shop, and with Hollmann's help, continue his studies. He obtained his master's degree and his PhD at the Technische Universitat in Berlin.

From 1952 to 1955, Werner was a research assistant at the Fritz-Haber-Institut of the Max-Planck-Gesellschaft in Berlin, where he had among his teachers and advisors Max von Laue, who had received a Nobel Prize for Physics in 1914, and Emst Ruska, who was later (1986) awarded the Nobel Prize for the development of the electron microscope.

In the 1950s, Werner began working with Bernhardt Hassenstein on the experiments and the theory of motion perception in the beetle. They had met during the war when Werner was on duty in a radio station monitoring the state of the ionosphere. Werner the physicist had met Berhard Hassenstein the future biologist, who had a similar radio station assignment. Those two kids (they were 20 years old or less) decided then that they would someday start a novel institute in physics and biology. This was clearly a critical meeting for their mutual futures.

Because of their 1952 paper on the beetle, Max Delbrück offered Werner a postdoctoral position at the California Institute of Technology (Caltech) and convinced him to work as a physicist in biology. At about the same time (in 1954), Jim Watson and Niels Jerne were also postdocs with Delbrück. Werner made an adventurous coast-to-coast automobile trip with Jerne at the end of his postdoctoral period that was punctuated by car breakdowns and impromptu lectures at local colleges to help pay for gas and repairs. They offered a package of two lectures: one on horses, on which Niels Jerne had done his experiments on the clonal theory of immunology (for which he received a Nobel Prize), and one on beetles, the experimental subjects of Werner's theory of the optomotor response. Shortly afterward, in 1958, Werner became head, together with Bernhardt Hassenstein and Hans Wenking, of the Forschungsgruppe Kybernetik within the Max-Planck-Institut für Biologie in Tübingen. In the meantime, offers by Caltech, Massachusetts Institute of Technology (MIT), and Bell Labs triggered a counteroffer by the Max-Planck-Gesellschaft in 1960 that Werner accepted; in 1963, he started his own department in the Max-Planck-Institut für Biologie. In 1968, this department became, with the nomination of three other directors: V. Braitenberg, K. Goetz, and K. Kirschfeld, the Max-Planck-Institut für Biologische Kybernetik. For the new institute, Werner chose a system for studying visual information processing that was neither too simple nor too complex. The fly's brain, with its 10^6 neurons, is halfway on a logarithmic scale between unicellular organisms and man.

My EMBO-sponsored two-week visit at the institute took place a couple of years after the institute was founded (1968), when the building was still very new and partly unfinished. At the end of the two weeks, Werner offered me a one-year position as Wissenschaftliche Assistant. I accepted immediately. Arriving in Tübingen in September 1971, I stayed in the GastHaus for the first three months and then rented an apartment with a fellow researcher from Wales (Brian Rosser) in Walhauser Ost, a 10-minute walk to the institute, though I had a red Mini Morris station wagon of my mother's. I started working with Werner and Leo Heimburger, who was Werner's technician in charge of the quantitative torquemeter-based behavioral experiments with flies.

In June 1972, I was married in Genova in the ancient San Siro di Struppa. The original church was built in the fourth century; the actual building is new, dating from the twelfth century. My wife, Barbara Venturini Guerrini, joined me in Tübingen in what became a great journey together. She helped me more than anything or anybody else not only in life but also in my scientific career. The initial two-week visit to the Max Planck Institute stretched first to two years and then to ten years. It was a very happy period of our life together.

The Grand United States Tour

In the fall of 1973, after almost two years of work in the institute, Werner invited me to join him on a scientific trip to the United States. It turned out to be a dream tour for this (then) 26-year-old. In retrospect, Werner made a point of introducing me to many of his scientific connections. In New York, we stayed at the guesthouse of Rockefeller University and met with Haldan Hartline and Floyd Ratliff, who established the concept of lateral inhibition in limulus. We visited Gerry Edelman, who was just beginning to think of switching to neuroscience. I was incredibly impressed by Gerry and his broad knowledge of art and music. By chance, we bumped into Marc Kac the mathematician ("Can one hear the shape of a drum?") in the Rockefeller faculty club. At Columbia, we visited Eric Kandel, whom I already knew because of a talk he gave in Tübingen. In Boston, at Harvard Medical School, we met with David Hubel (I forget whether Torsten was also there), and, at MIT, we met with Lucas Teuber, Walther Rosenblith, and Frank Schmitt, who had then stepped down from the chairmanship of the biology department. Frank was very good friends with Werner, and somehow I was then invited to a meeting organized by the Neuroscience Research Program (NRP) that was to take place in fifteen days. Two weeks later, I was again flying to the United States, this time alone. At the NRP workshop, I met a good part of the remaining neuroscience community of Boston including Emilio Bizzi, Peter Schiller, John Dowling, Whitman Richards, and Walle Nauta.

I met David Marr for the first time in 1973 when I went to chat with Marvin Minsky at the artificial intelligence (AI) lab. Boston was wet, foreign, and dark. David came out of his office in the "playroom" and we exchanged a few words. His name was known to me, of course, because of his cerebellum paper, which was highly praised by many VIPs in the biological sciences. David also attended the NRP meeting in Boston. We had both been invited at the last moment and were not scheduled to speak. David sat quietly the whole time, listening to what people were saying about psychophysics and physiology. John Dowling joked with David about his red Mustang. Back at the AI lab, a scientific conversation took place about the ideas on the retina David was then developing. I had been asked by Werner Reichardt and Antonio Borsellino to invite David to give a series of lectures at the spring course on biophysics in Erice, Sicily. I was happy that he accepted immediately, and so our next meeting was already arranged.

Erice, Sicily

Erice is a beautiful old village on top of a mountain overlooking the Mediterranean Sea. For two weeks in 1974, the participants of the spring course on biophysics gathered together mornings and afternoons. Among them were Mike Fuortes, David Hubel, Bela Julesz, John Szentagothai, Sir John Eccles, and Michael Arbib. At lunch and dinner, we divided into small groups to explore the five restaurants of Erice, all above average for scientists. We also made several expeditions to the various beaches down the hill. I was impressed and obviously pleased by the interest and the respect David had for my lectures and my comments. David's approach to science was by far the most unconventional and for me the most interesting. We discussed our ideas at length while dining and lying on the beach.

The Fly Visual System: A Visuo-Motor Control System—Fixation and Chasing Behavior

During the time I collaborated with Werner (1971–1981), the behavioral and physiological work on the fly's visual system led to experiments, theories, and models at three levels of integrative neuroscience: the phenomenological theory of flight behavior, the algorithms for detection of motion and relative motion, and the underlying neural circuitry. I will briefly describe each one.

Werner had discovered that flies fixate; in other words, they fly toward small dark objects. He had developed a sophisticated flight simulator in which a flying fly is held fixed while its torque, measured by a very sensitive device developed by his longtime friend and colleague Karl Goetz, controlled the visual environment, thereby simulating free flight. In this way, it was possible to experimentally study and quantify the fixation behavior. Reichardt also developed a model of the visual fixation behavior, which was extended in joint work in 1971–75 into a quantitative description capable of accounting for the main features of fixation, tracking, and chasing in flies. The equations, derived from the experiments in the stationary setup, could predict the free-flight trajectory of one fly chasing another! Gadi Geiger, a dear friend of ours whom I met the first day I was in Tübingen, showed that the theory could also predict correctly a Kindof Mueller-Lyer illusion in flies, judging from the fixation behavior induced by the Mueller-Lyer figures!

Motion Detection: Algorithm and Circuits

Early on, long before I arrived in Tübingen, Werner had worked out the properties of motion detection in the beetle with B. Hassenstein and

D. Varjú. This early work is his best-known scientific contribution. The optomotor response of the beetle Cholorphanus is the animal's tendency to follow the movement of the visual surround to compensate for it. The beetle was glued to a rod so it could not move its body, head, or eyes relative to the surround but could express its behavior by rotating a "Y-maze globe" under its feet. The rules of the optomotor behavior are summarized in a precise, quantitative way by the correlation model, by now known as the Hassenstein-Reichardt model or simply as the Reichardt detector. The strict mathematical treatment of this model led to many counterintuitive predictions, which were, one by one, verified experimentally. The model holds up in many species and in many types of neurons. The influence of the Hassenstein-Reichardt model can hardly be overestimated. It inspired work on motion vision in many animals, including humans. Thus, the model set the standard for how researchers thought about visual motion detection and how they designed experiments.

In a more general sense, the Hassenstein-Reichardt model introduced mathematical techniques and quantitative modeling to biology. Visual stimuli were presented to the beetle Chlorophanus, held by its back in a fixed position while climbing on the grass-like ribs of y-maze globe. The frequency of left/right choices revealed its intended turns. Thus, it was shown that the beetle's motion detector requires at least two input sensors looking into slightly different areas of the visual field. These sensors are excited in sequence by pattern edges moving by. If one input signal is delayed relative to the other, the two signals become synchronous for the "preferred direction" of pattern motion and strongly asynchronous in the opposite direction. Multiplication of the two signals and time averaging of the result yields a direction-specific motion signal regardless of the polarity of the change of brightness achieved by a moving object. This process corresponds formally to an autocorrelation of the input signal—hence the name "correlation-type motion detector." This work describes the algorithm for motion used by the beetle and other insects such as the fly and also by primates. I was recently at a Janelia farm meeting (in 2013) in which several talks suggested that new genetic and connectomics methods are finally about to identify which neurons are the Reichardt detectors, including the site of the multiplication and of the subtraction.

Relative Motion Detection: Algorithm and Circuits

In 1974, Reichardt discovered that flies use motion discontinuities in the visual surround to distinguish objects from the background. Flies turn toward a single, randomly textured stripe in front of an equally textured background but only if it moves relative to the ground. Part of our theoretical and experimental work between 1974 and 1978 was devoted to characterizing the properties of the algorithm used by the fly's visual system to detect relative motion in a way similar to what Werner had done 35 years earlier for motion detection. The interplay of experiments and theory led to a class of model that could be characterized as a form of nonlinear lateral inhibition between motion detectors. Together with Klaus and Werner, we derived a skeleton model of the necessary neural circuitry and refined it further through quantitative experiments. It is a lateral inhibition network that finds and enhances discontinuities in the motion field, such as the discontinuities generated by an object moving relative to the background. Starting with the pioneering work of K. Hausen and of the late R. Hengstenberg, it became possible to record from several of the neurons that correspond to boxes in our original model. Thus, the original model and experiments led to a series of remarkable physiological experiments, which were broadly consistent with the outline of the model, while revealing its precise anatomical and biophysical features. This work took the problem of motion discontinuities to another level, describing the neural circuitry and the biophysical mechanisms. The close interplay between behavioral analysis, theory, and neurophysiology guided this work very efficiently and led to a profound understanding of these perceptual processes, very much as Reichardt had suggested way back in 1965.

Cambridge, Massachusetts—Levels Framework

In 1976, I came to the AI lab at MIT for a period of three months to work with David Marr and to clear my head so I could decide what to do next. For the first week or so, I was left relatively alone, free to play with Macsyma and LISP. I spoke about the fly visual system at David's vision seminar. It went well. After the lecture, we had a beer together in Harvard Square. David was very happy about my lecture. His enthusiasm and his praise were contagious. I felt great and alive!

While at lunch together at the MIT cafeteria, David and I were still trying to define the nature of our approaches to the problem of the brain. Our views were already very near and converged rapidly. I told David the philosophy of the approach I had been following in my work with Werner. We discussed it at length and finally I suggested that we write a paper together that the Neuroscience Research Program (NRP—the influential program founded at MIT by Frank Schmitt) had asked me to write. The resulting paper eventually became an important part of Marr's book entitled *Vision* (Marr, 2010, reissued). The "levels of understanding" manifesto has been mentioned as one of the most enduring constructs of 20th-century cognitive science and computational neuroscience (Wilems, 2011). The argument in our paper and in Marr's book is that complex systems, such as a computer and the brain, should be understood at several levels. Here, let me list just three levels: the hardware, algorithms, and computation. Our paper and David's *Vision* book emphasize that explanations at different levels are largely independent of each other: a software engineer does not need to know the hardware in any great detail. The message was important at the time, 30 years ago. The study of the problems to be solved, and of the associated computations, is relevant in its own right and is needed for a full understanding of the brain. As I have argued more recently, I think that it is now time to emphasize the connections between these levels, and to extend the range of levels, if we want to make progress in computational neuroscience.

To elaborate, let me recount the background of the argument. Our 1977 paper was the original "manifesto" on our computational approach to the brain. We started from an argument described in a paper by Reichardt and Poggio (1976) on the visual system of the fly, where we distinguished three levels: single cells and circuits, algorithms, and behavior (of the organism). David insisted, correctly, on replacing the behavior level with the level of computation and computational analysis. This was important for defining the approach of computational neuroscience. But one key aspect of the original argument in Reichardt and Poggio (1976) almost disappeared in the process. We had stressed that one ought to study the brain at different levels of organization, from the behavior of a whole animal to the signal flow (i.e. the algorithms and the circuits and single cells). In particular, we expressed our belief, which Werner had written about even earlier, that (a) insights gained on higher levels help to ask the right questions and to do experiments in the right way on lower levels, and (b) it is necessary to study the nervous systems at all levels simultaneously. From this perspective, the importance of coupling experimental and theoretical work in the neurosciences is clear; without close interaction with experiments, theory is very likely to be sterile.

I believe that David would agree that it is time to look again at the levels of understanding framework, now emphasizing the connections between levels and their synergies. In particular, I believe that neuroscience can help computational theory and even computer science as suggested by recent models of the visual cortex, which are leading to interesting approaches in computer vision. In 1979, when David wrote *Vision*, our belief was that computational theories might be of use to neuroscientists. The rise of computational neuroscience during the last several years confirms this notion. Moreover, the table is now turning. In the near future, neuroscience may well provide new ideas and approaches to artificial intelligence.

Cambridge, Massachusetts—Stereopsis

As time went on, I developed a deeper understanding of David's work on vision. In retrospect, it took a lot of time. Really new ways of thinking cannot be appreciated easily. A thousand different facets must be communicated with the magic of a language and a fascinating style. David's early papers on vision have these rare properties. A good way to understand something new is to criticize it, playing the role of devil's advocate. In doing this with David, I found myself at some point defending recurrent networks. David formulated the challenge of solving stereopsis in his way. He explained it in terms of his analysis of the computational problem of stereopsis—which cells had to be excited and which ones had to be uninhibited. On a paper napkin at the MIT cafeteria, I wrote down the obvious equations for the recurrent network. I claimed that it would be very easy to prove its convergence; Liapunov-like functions constructed from conditional expectations were what I had in mind. Back at the lab the same evening, after a dinner at the Greek restaurant in Central Square, David programmed the recurrent algorithm, which seemed to work well in one dimension. The day after, the two-dimensional version of the algorithm showed encouraging signs of liking the stereograms constructed by Béla Julesz. One week later, when I had finally understood David's computational analysis, I also realized that an analysis of the convergence of the algorithm was going to be very difficult. All general standard methods failed. When I told David that I had to turn to a last resort, a probabilistic approach, the teasing began. The teasing became even more intense when I had to write a program to compute the result of the probabilistic analysis.

In the meantime, our creative collaboration was exciting, despite the headaches from the convergence problem. We began working closely together. It was a fantastic experience. David was very sharp; he had clear ideas about almost everything, and they were usually right. I slowly discovered other facets of him. He was passionate about music, Italian opera for instance. I heard him improvising a few times on the piano. He played with ease and emotion. I was impressed. I had to wait another year, however, before hearing him play his true instrument, the clarinet.

During those three months in Boston, I often went sailing on the Charles River. But the really new experience was flying. David was a pilot, and my presence in Boston triggered anew his passion. I took a few flying lessons. Several times we flew together from Hanscom Field in a rented Cessna. On those occasions, I would stay overnight at David's house. Early in the morning, we listened to the weather forecasts and then drove to Hanscom Field where we would get a plane for the day from Patriot Aviation and share expenses. One of the most beautiful flights brought us to the Lakes Region in New Hampshire. The sky was as clear and deep blue as the water beneath us. We landed on a grass strip on a little island. It was very quiet. We walked to the water a few hundred yards away. There we sat for a few hours. We reviewed what we had done on stereopsis and decided to write a short paper for *Science* about it. David already had the opening sentence and the overall formulation was clear. We just had to sit down the next day and write. There was one white sail on the lake. We were encompassed by green, blue, and silence. David was happy and relaxed. So many more ideas and flights and forests and lakes awaited us! On our way back to Hanscom, the weather changed and rain started suddenly. On the final approach, David was very tense, his mind totally concentrated on the plane, the control tower, and the instruments. Many people were afraid of this concentration, which they mistook as a sign of unfriendliness or aloofness. I knew well the alertness of David's mind when he was discussing science, lecturing, playing music, or flying, and I could physically feel the presence of his thoughts, his total concentration. There was an incredible intensity to his thinking. His reactions and his answers were incredibly quick and at the same time crystal clear, sure, and sharp. Our landing on the wet field at Hanscom was perfect. Five minutes later, the airport closed down.

On another flying expedition to Nantucket and Martha's Vineyard, we got sunburned on the beach. Two days later, red like lobsters, we gave a lecture on stereovision at Harvard Medical School. We were making grand plans to fly across the United States for a month or so. Life was going to be a lot of fun! A few weeks before my return to Germany, we were right in the middle of the U.S. bicentennial. The tall ships were coming into Newport on their way to New York. On the weekend of July 2, the weather was beautiful, and David decided we should fly down to Newport. We came above the bay with our Cessna 170 to find that the blue sky was filled with flying objects: balloons, choppers, a Goodyear "blimp," and many other planes. Tower instructions were to circle at a specified distance above the tall ships. The surface of the sea was covered by little white traces, glittering under the sun. Hundreds of boats of all sizes came to meet the tall ships. The scenery was superb. It was simply great, circling above the ships, together with so many other planes and boats. Hundreds of planes were scattered all over the field at Newport airport, many of them old-timers, happy and colorful. But in the afternoon, the weather deteriorated quite suddenly. There was a storm with ghastly winds. Back at the airport, we thought for a while about leaving the plane and going back to Boston some other way. David phoned several times to inquire about the weather at Hanscom. It was clear, so he decided to start out. Airborne again, drops of rain slashed across the windscreen until we came up from the low clouds out into the sun. It was the eternally beautiful weather to which poets have accustomed us. But the feeling in a small plane without instruments is quite different. David, however, was relaxed. There was nothing to do but fly straight and wait for the clouds to dissolve. Near Boston, the ground started to appear in short intervals through foggy holes in the white carpet below us. When we landed at Hanscom, the sun was setting against a clear sky.

It is not by chance that my deep friendship with David was associated with flying together. Flying and friendship, joy and beauty, freedom and living, are things that are made of the same substance. I did not fly anymore with a light plane after David became ill. I do not know whether I am ever going to do it again.

Tübingen—Biophysics of Computation

Sometime after I returned to Tübingen from MIT, I met Vincent Torre, who was a student of Borsellino in Genoa, as I had been. He was younger, so we never met during our studies. We met for the first time in the house of Antonio Borsellino in Genoa during one of my visits back home. Shortly afterward, we started working together in Tübingen on the biophysics of neurons and synapses. Later, while working with Christof Koch, I dubbed that direction of research "Biophysics of Computation" following the title "Physics of Computation," a chapter in a well-known book by Carver Mead (C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, Reading, Mass., 1980). Vincent, who went to work with Alan Hodgkin (whose theory of the action potential is a major achievement in biophysics) on photoreceptors in Cambridge, taught me about biophysics. I learned more about biophysics from a textbook that I used to teach cable theory at the University of Tübingen. I decided to teach because the topic was important but so boring that I would not have been able to finish the book by myself. At the end of the year, one student was left in my class!

Vincent and I started from a relatively well-known operation, the threshold operation of spike generation. We derived a mathematical formulation that could be used more easily to incorporate the key properties of single neurons in the quantitative description of a system consisting of many neurons and synapses. We ended up using a functional power series expansion called Volterra series (also called Volterra-Wiener series) that I had learned from a seminal paper of Bedrosian and Rice, and which is a natural extension of linear system theory. The fun part of the project came when we had to write the power series expansion of the inverse of a power series expansion. It turned out that Lagrange had solved the problem in his 1770 paper, and this was still the best place to find the result. In those years before the Internet, that meant visiting the rare book section of the Universitaet Bibliothek and consulting Lagrange's article directly. The old French was not a problem, but the mathematical notation required quite a bit of decrypting. What I remember most is that there were quite a few mistakes in the derivation that somehow canceled each other out at the end!

Tübingen—Multi-Input System Theory

During this same period, I became very interested in extending linear, one-input system theory to nonlinear multi-inputs systems. The motivation was the visual system in the fly and in people. I had read papers about the behavior of Volterra series under white noise inputs. In the process, I started reading papers by Norbert Wiener and some of his MIT students such as Lee and Schetzen, and later, Amar Bose. I also interacted with classical mathematicians such as Bernard Coleman, who could not understand the mathematically crazy distributions such as delta functions that engineers were starting to use. In the attempt to make my mathematics more sound, I began a collaboration with Guenther Palm, who had just started to work with Valentin Braitenberg. Guenther was a mathematician with a fresh PhD in functional analysis. A lot of what I know today about function spaces, distributions, and different types of integrals I learned from Guenther. Guenther was a gentle bear, often appearing at our home for a beer around midnight.

Tübingen—More Stereo

David came to visit in Tübingen in the beginning of 1977. He stayed in the guest room at the institute and walked over to our home every morning for breakfast. We worked on the probabilistic analysis of stereopsis, discovering more difficulties every day. David wanted to think about a theory of human stereopsis. Eye movements were important. At that time, I had just heard from Jack Cowan of his work with Hugh Wilson on spatial frequency channels. David brought out Mayhew and Frisby's *Nature* paper on rivalrous stereograms. Our starting point began with two ingredients, the apparent falsification by psychophysical experiments of our first algorithm and the need for eye movements. We read everything on stereovision from Barlow to Julesz. At some point, we were suffocating in my office under piles of bound volumes of the *Journal of Physiology* and *Vision Research*. We even did some informal experiments. At the end of the three weeks, we had written three-fourths of the analysis of the cooperative algorithm paper (I had to write the final quarter with Gunther Palm) and had some rough ideas about a new model of stereopsis.

David had brought his clarinet. I introduced him to Eric Buchner, a good cello player. With another friend, a very good pianist, they played together several times. We were all deeply impressed by David's music. During his visit in Tübingen, the members of the Scientific Curatorium of the Institute—including its chairman, Sir Bernhard Katz—came one day to meet with the members of our institute. In the evening after dinner, David and other friends played Mozart's "Clarinet Concerto in A Major." I had never been so deeply struck by music as I was that evening listening to David and his clarinet. It was so beautiful and perfect, as full of emotion as to be almost unbearable. It was quite clear afterward that the audience had a similar experience.

At that time, David was quite alone in his work. He did not have anybody back at the MIT lab with which to work in the same way we had. I suggested he try to work with Shimon Ullman and share responsibilities of the group and the students. At that time, I knew Shimon only superficially but what David thought about him and his work left no doubt. David promised he would do it. That was an easier promise to fulfill than his promise to finally get out of his "craziness" and his "women problem." He did not manage that until he met Lucia, one year later. Those three weeks in Tübingen were a lot of fun; life was full, warm, and happy.

In June 1977, David returned to Tübingen. He stayed a full month in "his" room in the institute. The first week of his visit was difficult because a month earlier I had started an anatomical project, the first and only wet experiment I had done all by myself. I was very interested in the circuitry of motion computation in the retina because of my work with Vincent Torre. I was going to explore the (small) possibility that cobalt injected in the ganglion cells could stain the presynaptic processes via retrograde transport and uptake by the presynaptic terminal synapsing onto the ganglion cells. The project failed and also managed to make David quite nervous about competing with a few frogs. After that, we worked hard together, developing our stereovision ideas and writing them down. The days were productive. The theory took form. Through all my work with David it was often impossible to say who came up with a specific idea; almost everything came from discussions and thinking together and reciprocal criticisms. But David had the power of veto: if I was unable to convince him, that was it. He also had the ability to keep us on course.

We finished our manuscript right on schedule with time left to take Polaroid pictures of the two authors sitting with the title in one hand and stereo glasses in the other. (In the original draft of the manuscript there were a few lines warning the secretary that at that particular point we had just had too much Courvoisier and, therefore, the following sentences were going to be particularly immortal.)

The whole month was continuously concentrated, happy, and playful. As so often with David, science was fun and freedom! I often ask myself why David's presence had this incredible power. I still find it very difficult to give a full answer. But I know that part of it was the clarity and especially the force of his mind, of his thoughts. To think with David was for me an inebriating experience, a special feeling of playing and creating. Skiing downhill on a sunny day in the Alps gives me some smattering of this intellectual fun. In the spring of 1977, Werner organized a neurobiology meeting to celebrate the 500-year anniversary of Tübingen University. I helped organize the lectures. Many friends came: David Marr, Vincent Torre, David Hubel, Denis Baylor, Emilio Bizzi, Gunther Stent, Jack and Max Cowan, Bela Julesz, and others. David's lecture was a beautiful jewel of intellectual brilliance and improvisation.

In the middle of October that same year, I flew to Toronto for the annual meeting of the Optical Society of America; I was invited by Whitman Richards, who organized a special session. The whole MIT vision group came. It was fun, although short and chaotic. A couple of days later, I flew to Boston to work with David for three weeks. It was a fight with LISP and probability (again!). Ellen Hildreth, a student of David's and a good friend, wrote that "David was a very stimulating person; the energy level in the lab would suddenly double when David walked in (it would quadruple if Tommy [Poggio] was there too . . .)."

I remember how part of the zero-crossing idea had originated. Coming out of the cafeteria in the AI building (which also housed the CIA, Polaroid, and IBM), I expressed my uneasiness about taking zero-crossing and peaks of the filtered images because filtering the images was roughly equivalent to making their second derivative zero-crossing correspond to extrema of the first derivative. This made sense. But peaks were something strange, at least at this level and from this point of view. For simplicity, and because of the relations between derivatives, difference of gaussians, and bandpass channels, I wanted to flush peaks and retain zero-crossing only. David thought a while and then decided that, for reasons I had not thought of, the idea was actually not too bad. It is still unclear whether he was right. At the end of my stay, we drove together in a rented car through the colorful fall foliage in New Jersey down to Bell Labs. I gave a lecture for Bela Julesz and his small group on a topic that was completely uninteresting to them, synapses. When we mentioned our probabilistic analysis of zero-crossings, Bela named some mathematicians at Bell Labs who had worked on somewhat similar topics. Among them was a name that we did not know, Ben Logan. We asked for the paper, and Bela sent his secretary to get reprints. Glancing through it I saw that his theorem was very suggestive of our notion of independent bandpass spatial frequency channels. In the hotel and later, in the car, I tried to convince David, who remained quite skeptical. The zero-crossing idea and its connection with Logan's theorem is the kind I immediately like. Unfortunately, such ideas are often too nice to be biologically correct, and David was very probably correct in his skepticism.

La Jolla—Crick and Marr

I met Francis Crick at about the time when Francis and Odile moved from Cambridge, England, to the Salk Institute in La Jolla, California (1976). At Salk, he became a theoretical neuroscientist, following his second passion—after the mystery of life, the mystery of the mind.

I had visited Francis and Odile during the summer at their house on Portugal Place in Cambridge, England, with its golden helix above the front door. In La Jolla, I went with them and with the Orgels on trips to the desert. I saw him at the F. O. Schmitt's Neuroscience Research Program meetings. And I saw him debating about consciousness with various guests at our home, and answering question from my son Martino, and from his Texas Instrument's Speak and Math.

In 1979, Francis invited David and me to spend a month at the Salk Institute trying to understand the connection between the architecture of visual cortex and several intriguing aspects of visual perception. Francis always had the time, the interest, and an infectious enthusiasm for discussing scientific problems. He regarded us skeptically but seriously. A quite faithful account of those unforgettable conversations (mainly between Francis and David with me and sometimes with Leslie Orgel playing the role of the fan club) is in the third part of David's posthumous book, *Vision*.

During this visit, I discovered the clarity of Francis's mind and his incredibly intense focus. After hours and hours of discussion on a problem with the solution still escaping us, David and I were often tired, confused, and ready to give up for the day. Not Francis; he was relentless, forceful, critical, and enthusiastic. He was not a mathematician, but he knew how to use mathematics and how to visualize it. I still have his letters full of diagrams and equations dating back to the time when we worked on an extension of the sampling theorem, trying to understand aspects of motion perception.

During discussions with Francis, we began to worry about the function of the large number of neurons in layer 4c beta of V1 (prompted I think by a visit by David Hubel at Salk). We speculated about their possible role in image interpolation and hyperacuity. We had a phone conversation with Gerald Westheimer at Berkeley and ended up writing a paper with Francis. Once back in Tübingen, I was in charge of the appendix, which dealt with a little extension of Shannon's sampling theorem to moving patterns. The writing of the appendix was a lot of fun because it involved a heavy exchange of handwritten letters full of equations and drawings with Francis. As a result of that side project, I began to work with Manfred Fahle in Tübingen. He was a bright new PhD–MD student, and this work led to interesting psychophysical work on spatiotemporal acuity.

Leaving the Max Planck Institute for MIT

Werner offered me the best position he could, just below directorship, with the promise that I might replace him once he retired. Being a director in the Max Planck Gesellschaft is unique: it includes tenure for the director and the entire lab and freedom from writing grants along with optional teaching. Werner was a great friend and my scientific father; we loved Tübingen and the Max Planck Institute—a kind of ivory tower of science. But it was too soon for paradise. My wife and I needed to experience the real world with competition and the fight for survival. So in 1980, I decided to move to the MIT Artificial Intelligence Laboratory and join the faculty in the same department (psychology) where David was.

We left Germany with plenty of insurance: both of us were given a two-year leave of absence and much encouragement to come back. I felt the need to buy a house in the United States to keep it from being too easy to go back! That was a tough act. In the summer that I was looking for a house, Paul Volker was fighting inflation and Barbara was on the beach in Levanto. Because of Paul, mortgage rates were rising every week by one percentage point, a rather unusual situation. I ended up with a 17 percent mortgage rate and a second mortgage generously contributed by MIT at the great rate of 18 percent! The person who was most instrumental in convincing us to move and who provided us with the most help, funding, and support was Whitman Richards. Whitman had created the computational vision group at the AI Lab, managing to pull together David Marr, Shimon Ullman, and me. Whitman had a long series of great students and was an inspiration for generations of MIT researchers.

When I arrived at the MIT, there were seven or eight faculty in the Artificial Intelligence Laboratory. Only Patrick Winston, who was the director, was born in the United States, the rest were foreign born. Patrick started his job as director of the lab when he was still a graduate student! I shared my LISP machine ($100,000 for a desk computer) with one of my most remarkable friends, Kobi Richter, then on leave from the Israeli Air Force. In return, Kobi helped me buy not one but two used cars by playing the tough guy and the expert.

My salary at MIT was considerably lower than in Tübingen, and my wife could not work because her neuropsychiatric specialization was not recognized in the United States. (She was the real breadwinner at home.) The mortgage was high. I figured out that most MIT faculty had consulting jobs one day a week to make ends meet. Fortunately, a consulting opportunity materialized soon in the form of Danny Hillis and "Thinking Machines."

Thinking Machines Corporation was a supercomputer company started in 1982 by Danny Hillis and Sheryl Handler. It was an effort to turn Danny's PhD thesis (with Marvin Minsky on massively parallel computing architectures) into a commercial product called the Connection Machine. When I arrived at MIT, Danny was finishing his thesis. As a graduate student in the AI lab, he had a rather large research group, composed mostly of undergraduates. The company's ambitions were more scientific than commercial. Its motto says it all: "We're building a machine that will be proud of us." Thinking Machines was one of the first high-tech companies with its own kitchen and chef. It was more like a club of smart people than a company. Among the people I met there, in addition to Danny and Marvin, were Jack Schwartz and Steve Wolfram (before Mathematica) and most notably, Richard Feynman, with whom I had long conversation about vision in the fly and color vision in general.

The first years at MIT were confusing. I was dealing with the loss of David Marr, who died of leukemia in 1980, learning a new way of life, and learning a different way of doing research; I was much more alone. I also had to get grants and even teach, though not very much. Emilio Bizzi helped me during this period. I met many people, often without really understanding who they were. For instance, one day Sidney Brenner, whom I knew through Francis, dropped by my office at the AI Lab with David Baltimore, mentioning that David was about to start a new biological institute, which turned out to be the Whitehead Institute. Another day, Steve Jobs came to find out about AI and computer vision. He was shown our new stereo head system that mimicked eye and head motions by Keith Nishihara. Jobs was young and stiff and quite out of place, dressed in a jacket and tie.

In retrospect, normal life at the AI lab in those days meant that, on any given day, Steve Jobs might appear, as would Bill Gates. Once a small group of Italians arrived and asked Patrick Winston, the lab director, if they could take pictures of him and his AI colleagues. We ended up as models in a special issue of *Vogue* (*Uomo Vogue*) dedicated to fashion modeled by scientists. I discovered only recently that the photographer was Oliviero Toscano. Oliviero is very well known for his famous campaigns (e.g., Colors of Benetton). He is a great guy with a deep understanding of the human condition.

I was also trying to cope with two offices, one in the AI lab at 545 Technology Square, and one in the old psychology department, as well as with two research directions, biophysics and computer vision/ psychophysics.

Biophysics

Christof Koch had been my first and only PhD student in Germany. Because I was not a professor at the University of Tübingen (where apparently Kant was denied tenure), Valentin Braitenberg had agreed to be Christof's official advisor. Christof started working as a technical assistant on a C program to digitize trajectories of flying flies in 3-D. We then decided on a PhD thesis on the biophysics of computation by studying the computational properties of synapses and dendrites. Part of the thesis was based on anatomical data obtained from retinal ganglion cells from the nearby lab of Heinz Waessle. Christof finished his thesis a few months after I moved to the United States and then joined me after Werner Reichardt helped to get him a postdoctoral fellowship from the Franz Thyssen foundation.

With Christof and Vincent, we worked on the idea that shunting inhibition could mediate nontrivial information processing operations because the relationship between synaptic inputs, which are conductance changes and voltage effects, is nonlinear. In particular, shunting inhibition could be used to implement almost logical gates distributed over the dendritic tree, making a neuron more a very large scale integrated (VLSI) chip than a single transistor. This research direction had begun a few years earlier in 1977 in Tübingen with the proposal that a specific biophysical mechanism could explain directional selectivity in the vertebrate retina. It was Vincent who suggested it, based on work he had done in the Consiglio Nazionale delle Ricerche (CNR) lab in Pisa with Marchiafava.

Stereovision

I also continued to work on the problem of human stereopsis. At the time, we had reasonable success in terms of the computer performance of our stereo algorithms, but it was still unclear how the visual cortex solved the stereopsis problem. These days, artificial stereo systems are pretty good in absolute terms, and disparity maps obtained by them are used in a number of applications including a vision system for cars (by Mercedes). We still do not know in full how the human visual system computes stereo, but the mystery is not about performance anymore. There is an irony here, and perhaps some interesting work to be done. Algorithms like ours (Marr and Poggio, 1976; see also Parvati Dev and Michael Arbib) showed that Julesz's random dot stereograms could be "solved" pretty well by simple excitatory and inhibitory interactions (our stereo algorithm was a Hopfield network before Hopfield networks were invented) that operate on low-level representations of the image. This demonstrated that Helmholtz was wrong, and recognition of objects and objects' parts is not needed for matching the images from the two eyes. In fact, all modern stereo algorithms work on low-level descriptions of the images. The irony is that, in our everyday stereovision, the Helmholtz approach may be as important as low-level matching. Here is a conjecture (good for a PhD thesis): Low-level and high-level matching should be used by a stereo algorithm to obtain the best stereo results, and both are likely to be used in human vision.

One specific memory connects research on stereovision with some peculiar American realities. One day, I was visiting with Eric Grimson, another MIT professor, at a military lab in Washington. We were taken at some point to a room with some of the most recent photogrammetry tools, including one of the first Cray 1 supercomputers. Suddenly, while our host was describing some of the work on stereo, big red lights started flashing and loud alarm bells went off. The place exploded in a frenzy of activity. Eric and I were whisked out of the room by military police. Later, we were told that non-U.S. citizens were not supposed to be there, and they had just discovered that I was Italian and Eric, Canadian!

The work on stereovision also involved writing a review about computation and physiology with the other Poggio, Gian Poggio. Max Cowan was a great friend of the other Poggio and the editor of the *Annual Review of Neuroscience*, and he invited us to write it. We had the same last name; we were born in the same city, grew up a few blocks away in a similar circle of families, but were divided by 15 years or so. Ironically, we met for the first time in Belmont at a dinner in the home of our mutual friend Emilio Bizzi.

Zero-Crossings and Texture

Cambridge and MIT brought many visitors and new friendships. I met Edwin Land in a few visits with David Hubel. I met Don Glaser, the physicist who invented the bubble chamber, in 1981 in Klosters in one of Eigen's skiing workshops. I met Don again later in Cambridge during his mini-sabbatical in 1982 in the Rowland Institute, a private lab, now part of Harvard, where

Edwin Land, the founder of Polaroid, was studying human color perception. Don came to a vision conference I organized in 1984 in Erice. In 1992, we organized a Dahlem conference in Berlin entitled "Exploring Brain Functions: Models in Neuroscience."

On the suggestion of Emilio Bizzi, then the chairman of the new Department of Brain and Cognitive Sciences, I started a Center for Biological Information Processing (CBIP) with Ellen Hildreth, which was one of the first hubs for computational neuroscience in the United States. In 1985, John Hopfield came to CBIP from Caltech for a yearlong sabbatical. I put him in an office with Christof Koch. A bit later, back at Caltech, John started the computation and neural system program; the first hire of the new program was Christof! A year later, in 1986, David Mumford, acclaimed for his work in algebraic geometry, came from Harvard to spend a year of sabbatical at the artificial intelligence laboratory next to my office.

Regularization in Vision: Edge Detection, Motion, Markov Random Fields

In the first three years or so at MIT, my research had been productive but somewhat directionless and opportunistic in terms of problem choices. In addition to biophysics, the overall theme was computational vision, but I was not following any new "big" idea. A new unifying theme emerged thanks to Vincent Torre during a trip to Genoa, where he introduced me to Mario Bertero and to inverse problems. It became clear to me that vision was an inverse and ill-posed problem and that regularization techniques "a la Tikhonov" would provide a general approach to such problems. Everything suddenly fit together: constraints advocated by David and used in our stereo paper were needed to regularize a problem and make the solution of the corresponding optimization problem unique and well-behaved. All problems in vision and more general perception were inverse problems, going back from the image to 3-D properties of objects and scenes. They were also, as typical for inverse problems, ill-posed. We used regularization techniques to "solve" specific vision problems such as edge detection and motion computation. In the process, we found that some of the existing algorithms for shape-from-shading, optical flow, and surface interpolation were a form of regularization. Our main contribution was to recognize ill-posedness as the main characteristic of vision problems and regularization as the set of techniques to be used for solving them. We wrote general papers on this (the *Nature* paper was first rejected at a neural network conference). We recruited Christof to explore an intuition I had—that analog electrical networks, and possibly simple properties of neurons and dendrites, could be used to implement regularization. As always, I was trying to connect different levels of analysis; in this case, I was trying to bridge the computation to the hardware.

BK-SFN-NEUROSCIENCE-131211-09_Poggio.indd 384 16/04/14 5:25 PM

Beyond Tikhonov Regularization: Markov Random Fields and Fusion of Information

It was quite obvious to me that Tikhonov regularization was a good beginning but not the full solution for regularizing ill-posed problems. Tikhonov regularization is a constraint on the norm of the solution. Usually, this norm constraint imposes smoothness of the solution, depending on the representation and its metric structure. In general, one would like to have a broader choice of regularization constraints than just smoothness. With Alessandro Verri, one of my great friends who came from Italy as a postdoctoral fellow, we tried to explore more general shape constraints in the framework of regularization—but the mathematics was hard.

At the time, I did not know about reproducing Kernel Hilbert spaces (RKHS) and focused on two possible directions, both probabilistic, for extending Tikhonov regularization. The first one was based on the observation that Tikhonov regularization could be interpreted in Bayesian terms. Minimization of the Tikhonov regularization functional is equivalent to a maximum apriori estimate (MAP) of the conditional probability distribution of the solution given the data under a particular Gaussian prior and the assumption of Gaussian additive noise in the data. The observation suggests that Tikhonov regularization may be a special case of a more general Bayesian approach. Estimates of conditional distributions may allow the expression of more complex and specific constraints than "smoothness." I did not follow this approach at the time, mainly because of computational and conceptual difficulties. The Bayesian approach is so general (a very big hammer), and one pays so much for it, that I wanted to use it only as a last resort; it is almost equivalent to giving up understanding and resorting to greedy optimization with no guarantees. The second direction was more specific. It was the use of Markov random fields (MRFs) with line processes to deal with regularization with discontinuities as in surface interpolation. The research came from a very good student of Sanjoy Mitter and mine, José Marroquin, and matched well with the use of the Connection Machine, the first parallel supercomputer produced by Thinking Machines Corporation (TMC). As it turned out, Markov random fields were in our hands at the time a rather brittle technique, sensitive to parameter choices. The best computer vision use of them was to integrate information from different sources such as color, intensity, motion, and depth to find surface discontinuities in the scene. My daughter's teddy bear demonstrated it on the cover of *Science*!

Color Constancy and Learning

The work on regularization had a number of specific applications in vision. We followed up on some of them, such as the computation of optical flow and color. I was, however, thinking of steering back toward my ultimate interest, the problem of learning. The opportunity came with a graduate student who was working on color perception. Anya Hurlbert was one of a small number of brilliant MD–PhD students in my career. It seemed to me that all of them were slightly masochistic; as if a PhD at MIT was not enough, they complemented it with an MD at Harvard. In any case, Anya, who was extraordinarily good in research, kept explaining to me the mysteries of color. Most natural objects are visible because they reflect light. The proportion reflected varies with wavelength, defining the surface reflectance function. The light reaching the eye is a product of the surface reflectance function and the spectral power distribution of the illuminant. If color is to provide a reliable cue to the surface, the visual system must compensate for changes in the illuminant. Object colors do tend to appear unchanged under different lighting, showing that color appearance is closely tied to surface reflectance. Many factors are thought to contribute to color constancy, which is a typical inverse problem. The illuminant changes during the time of the day and from outdoor to indoor, but our brain discounts the changes and computes a good approximation of the reflectance properties of objects, which is invariant to illuminant variability. Edwin Land had proposed an algorithm to achieve this goal. My question was whether one could do something similar by learning to associate to a variety of images with different illuminant the underlying "true" reflectances. We tried it out in a simple associative learning scheme in the spirit of the old work done in Tübingen, following up my interest in holographic memories. The subject could learn to extract the reflectance from the images reasonably well in toy problems. Moreover, the solution was a linear filter very similar to the algorithm proposed by Land!

Learning and Regularization

In 1988, I was still thinking about learning. The whole neural network fashion had started because of Hopfield networks (1982) and backpropagation (1986). I was skeptical. I claimed it was the usual recurrent epidemic of ideas about intelligence. Epidemics of the same flu virus tend to reoccur about every 20–25 years for obvious reasons. Scientific ideas have similar periods, about a generation long, as shown by William Goffman in a seminal *Nature* paper on the field of mathematical logic. I challenged Geoffrey Hinton in a meeting at Snowbird (that eventually became NIPS), asserting that there was nothing magical about neurons, that the term neural networks was just a sexy name for statistics, and that the field would in fact become part of statistics. In retrospect, I was mostly right. I missed completely, however, the importance of good metaphors for triggering good people to work in a field!

In 1988, I saw a paper about radial basis functions by Broomhead and Lowe on multivariate function approximation. The idea of looking at learning from the point of view of function approximation was appealing to me. It immediately connected the problem of learning with well-known mathematics. Furthermore, I saw a direct connection with regularization. I knew from work in computer vision, and in particular on edge detection, that Tikhonov-type regularization can be seen as a linear combination of a set of basis functions centered at the data point. I even knew how to get a Gaussian basis function from a particular regularizer. I had seen this in a paper by Alan Yuille and Norberto Grzywacz on a completely different problem—motion. Everything seemed to click: learning was an ill-posed problem as well, maybe the ultimate ill-posed problem of predicting from a subset of data, exactly what science tries to do. Fortunately, a great guy, Federico Girosi, had just joined my group, continuing a tradition that started with Vincent Torre (Alessandro Verri was a student of Vincent's; Federico Girosi was a student of Alessandro's). Federico was a better and more careful mathematician than I was; he was the ideal collaborator in that exciting time. Together, we wrote an AI memo entitled "A Theory of Networks for Approximation and Learning," which generated a number of papers in Institute of Electrical and Electronics Engineers (IEEE) and *Science* and remains the most highly cited of all of my publications.

Learning, the Brain, and Object Recognition

In 1990, I took my first six-month sabbatical at a new institute in Trento, Italy, called Instituto di Ricerca Scientifica e Technologica (IRST). Luigi Stringa asked me to form and lead a vision group. I was helping with the vision of the institute. Bruno Kessler and Luigi Stringa made quite an impression on Barbara and me at a dinner at our home. I was happy to have the opportunity to spend more time with my mother in Genoa. She also came to spend time in my apartment in Trento. I went to Trento with some of my students and close collaborators such as Federico Girosi. After our theoretical formulation of the learning problem in terms of regularization and regularization networks, the playground was open to many applications.

I have always loved to collaborate across disciplines, and our development of a family of learning algorithms opened a big playground for me. I started working on applications of the new approach in a number of domains. Not surprisingly, the first problem I worked on was the problem of the brain in a manifesto entitled "A Theory of How the Brain Might Work," which was a more modest version of the original "A Theory of How the Brain Works." The paper claims that the ability to learn, generalize, and to predict is at the core of intelligence. Furthermore, regularization shows that learning is much more than memory—it can be regarded as an interpolating look-up table. The paper's main message is that intelligence is memory-based computation. I believe this is a useful point of view and possibly a powerful insight, consistent with the evolution of intelligence.

At the same time, I thought about a computational model based on this point of view for explaining pose-invariance of object recognition in humans and primates. The model was equivalent to a network with neurons tuned to views of an object from different viewpoints. It was a toy model but made the point that 3-D models, then the standard wisdom, were not needed. Shimon Edelman, a new postdoctoral fellow, did the implementation. We used 3-D clips—that Shimon and Heinrich Buelthoff were starting to utilize for psychophysical experiments—as our test objects. We wrote a short paper for *Nature*. I went back to hyperacuity to find an explanation for visual perception in terms of the new learning framework. I had first started to work on this topic with Francis Crick at the Salk Institute and later with Manfred Fahle in Tübingen. This time I teamed up with Shimon Edelman and Manfred, who was spending his sabbatical in my group. We showed that hyperacuity is quite easy to learn from a relatively small set of examples without any additional circuitry.

Entrepreneurship

Life continued with fun science and fine friends coming to visit. Francis Crick used to come to visit from time to time. One of his visits started a tradition in my group that went on for ten years or so: a "Crick dinner" was a dinner for all of my students in which a well-known scientist was present, such as Francis on that first occasion. Don Glaser, Marvin Misky, Steve Smale, David Mumford, Thomas Kuhn, Graeme Mitchison, Anya Hurlbert, Christof Koch, Gian Poggio, and Shimon Ullman were all at one time or another the excuse for a Crick dinner. Some friends and colleagues were leaving for better positions (e.g., Heinrich Buelthoff became a professor at Brown, Nikos Logothetis took a job in Houston, and Manfred Fahle took a job in Tübingen and then Bremen). Unfortunately, some were leaving and not coming back—ever. Werner Reichardt died on September 18, 1992, a few days after a very nice workshop in Tübingen with all of his main collaborators celebrating his retirement. From 1968 to 1992, Werner Reichardt made important contributions to the understanding of visual perception, using the fly as a model organism. Werner was a member of many academies in Germany; a member of the prestigious order, Pour Le Merite; a foreign member of the National Academy of Sciences; a recipient (with his friend B. Julesz) of the Heineken prize of the Royal Dutch Academy; and also the recipient of many other honors of which he was rightly proud.

It had not escaped my attention that the learning framework had commercial applications. Furthermore, I had been consulting for TMC for almost a decade, but that important line of intellectual activity and extra income was in trouble. After receiving a great offer from IBM, which was declined by Sheryl Handler, Thinking Machines—the world leader in parallel supercomputing—went through a rough patch, eventually filing for Chapter 11 in 1994. In the meantime, I had ideas about applying the learning framework to graphics. If vision was an inverse problem, graphics was a direct problem, but one could look at it as the inverse of vision and use learning techniques to synthesize new images under the control of a few parameters from a training set in which pairs of images and control parameters were available. The idea turned out to be equivalent to multidimensional morphing. We had a plausibility demonstration with Roberto Brunelli at IRST. After Roberto declined to pursue the approach, I happened to team up with Steve Librande, a great student at the media lab at MIT. At the time, he was developing software tools to help animators draw characters. His group had funding from the creator of "Garfield."

Steve combined his great artistic and software skills with radial basis functions and produced a video that is still captivating today on how to synthesize new drawings from a small set of original ones in the artists' style. A number of visitors at the MIT media lab, which was just becoming glamorous, were interested. We were naturally led to a start-up while fighting with Nicholas Negroponte about intellectual property (the original patent was held by IRST and MIT; the media lab had to share it, and Nicholas did not want to). The technology licensing office suggested an investor with whom I met at the old AI lab, Charles Harris.

I remember that I offered to drive Charles to the airport because I liked this slightly odd businessman who seemed to be really interested in research and was wearing a perfect business suit with slightly unusual sneakers. Charles was a real gentleman. I came to appreciate enormously his and his wife Susan's friendship, his intelligence, his wisdom, his curiosity, and his enthusiasm for life and for friends.

The MIT Technology Licensing Office (TLO) had proposed a possible chief executive officer (CEO) (Hoomin Toong); nFX started and soon moved to Silicon Valley in Mountain View, California. As it turned out, our technology was good but too early, ahead of the Web, and our CEO was the opposite of what the open atmosphere of the valley would have liked; nFX was not an easy ride for me and even less for people directly involved such as Steve Librande and Joel Voelz, the (very good) successor of Hoomin. Harris and Harris lost a couple of million despite the eventual sale of nFX to Adobe.

The other obvious application of learning techniques was to predict the stock market. In the late 1980s, I had been invited by Paul Glaser (via Patrick Winston), the technology chief at Citicorp under John Reed, to be part of an advisory group of faculty from some of the best computer science (CS) departments in the country, called the Strategic Technology Evaluation Program. We were about 14 hand-picked, world-class computer science experts chosen to advise Citicorp on innovative applications of technology. I remember Michael Stonebraker, Ron Rivest, Tomas Lozano-Perez, and Sal Stolfo. Through my work with Citi, I got in touch with Paul Ardern, head of a trading unit of Citicorp in London, who asked me to consult for them on their trading of Japan warrants. At the time I had a PhD student, Jim Hutchinson, who was getting ready to make a lot of money using machine learning in finance. I got him to help me consult with Citicorp. We set up a consulting partnership to which we quickly added a third partner, Xiru Zhang, who worked at Thinking Machines, and together with Jim had won an early competition by the Santa Fe Institute in predicting the future of a couple of time series. There was a potential conflict of interest because Jim was my student. I asked Patrick, the director of the AI lab, for advice. Patrick suggested that we get a thesis co-advisor, which we did. This is why I have a paper in a rather prestigious finance journal, the *Journal of Finance*. This is also how I started to collaborate with Andrew Lo, one of the smartest people I know at MIT, a collaboration that continues today. With our personal funds, PHZ started trading futures on the Nikkei from an office in Cambridge next to MIT. With some luck, we managed to have a good trading record for a few months. We then looked for outside investments. Harris and Harris and, a bit later, Commodity Corporation came in and became shareholders together with the three of us. A bit later, all four of us moved to Wayland to be closer to where Jim and Xiru lived. We grew to about nine people and to a capital of almost a billion dollars under management, which was then quite a bit of money. We did quite well until 2007, when the Lehman debacle surfaced. Unfortunately, Lehman was the broker for our European trading and quite a bit of our trading capital disappeared when Lehman went bankrupt. We recovered some of it years later, but, in the meantime, we had to close our doors.

Neuroscience of Object Recognition

Our model for viewpoint-independent object recognition was eventually supported by an elegant psychophysical experiment by Heinrich and Shimon, which showed, along with many other later papers, that Biederman's object-based claims are not correct for immediate perception. I was thus able to convince Nikos Logothetis to do related physiology experiments to check the prediction of view-tuned units in a monkey trained to recognize the same objects from different viewpoints. I did very little work for the paper itself. My main contribution was the initial idea of the experiment, but much more importantly, to have pestered Nikos for so long and so effectively that the experiments were actually done.

Object Recognition and Computer Vision

Don Glaser and I continued to stay in close touch. When my son was in his last year of high school, I took him on a tour of the colleges on the West Coast. We started in March 1995 with Berkeley, where we stayed with Don and Lynn Glaser, who organized a fantastic party for us at their beautiful home at the top of the Berkeley hills. This was the time of Berkeley's "golden parachute," an offer to coax retirement-age faculty to indeed retire.

The offer was so generous and flexible that everyone at the party had just retired including Don himself. It is there that I again met Steve Smale, who had also retired and accepted a new job in Hong Kong. The tour of the West Coast colleges ended with the University of California, San Diego, (UCSD) and a lunch with Francis Crick, who at the time was the president of the Salk Institute. Afterward, I took Martino to see the "gliderport" next to Salk on the cliff above Black's Beach. On the impulse of the moment, while Martino was looking at postcards, I signed up for a flight and took off with an instructor, jumping on the Pacific and then "crabbing" alongside the cliffs.

Don was a hero to me. He was my first example of someone who was successful in science and in business. There are of course others, such as Shimon Ullman, but at the time, Don was the only one I knew. On that occasion and so many others in the years to come, I loved discussions with Don and the observations he made. Don was a charming mix of curiosity and innocence. His openness to possibility and discovery was the key to his path in life. Don was the least arrogant and most delightfully funny person you could hope to meet. His view of the world and the way he communicated his observations were simple, crisp, and deep. He carried his success in science and in high-tech industry in a graceful way. During his whole life, he retained the happy curiosity of a kid. The world for him was a garden of wonders. He was always able, in an apparently effortless way, to come up with new and refreshingly counterintuitive observations of things and of people.

Work in my group continued to focus on the applications of machine learning and in particular, regularization. The most interesting topics were in object recognition. Work with Roberto Brunelli at IRST used machine learning techniques for a few object recognition tasks, such as gender classification from face images. Face identification was a somewhat more difficult task. We did some preliminary work showing that approaches based on the geometry of faces, the dominant approach dating back to police identikits and Takeo Kanade's PhD thesis, were far inferior to the correlation of filtered images (through a Laplacian of a Gaussian filter, similar to the tuning of ganglion cells). We also found that using a set of small templates rather than a single full-face template delivered better performance. These are simple but robust findings that resonated through later research on face recognition.

Kah-Kay Sung developed one of the first ever systems to learn to detect faces in images, which is now in every digital camera and phone. Kah-Kay started this project on his own and initiated a direction of research that continued for decades in my group. Kah-Kay was also the first of several great students in my lab that were from Singapore, a city-state whose people I admire. With Thomas Vetter, we looked at the role of symmetry in object recognition; symmetry can be regarded as guaranteeing additional virtual examples, thereby decreasing the sample complexity. We also studied shape-based categorization in terms of "nice" classes of objects. These are objects that can be represented in terms of a relatively small number of basic shapes. Faces are a good example. Empirically, almost every face can be represented as the linear combination of about 100 individual faces in 3-D. Several mathematical properties follow easily for "nice" classes. This approach was combined with regularization networks on real images by David Beymer to do image-based graphics and example-based recognition.

Amnon Shashua, who was a very independent student and postdoctoral fellow doing groundbreaking work in the geometry of computer vision, helped in some of this work. He is one of the smartest people I know and a very good friend. Perhaps because I found geometry difficult to understand, we interacted more on the business side, trying to help Takeo Miyazawa, a Japanese friend, with his start-up in Japan. Work on the geometry of vision, to which Amnon was a main contributor, was the last hurrah of computer vision as a field independent of machine learning. Thus Amnon was not directly involved in object detection, which was our first foray into machine learning for vision. Ironically, a company he started later in Israel, MobilEye, is at the moment the poster child for commercial success in machine learning. The chipset that the company makes provides vision for cars. One of its tasks? Pedestrian detection! I bet that the driver-less car of the future will not come from Google but from MobilEye. I am very proud of MobilEye and Amnon.

The lab took occasional diversions related to the main themes of learning and vision. Collaboration with a research center in Bari led to work with Nicola Ancona on optical flow, leading to efficient hardware implementation based on Green's theorems. A recurrent topic in computational vision was whether certain tasks had to precede others, like edge detection before object recognition. Psychophysics shows that, in certain cases, this is so but in others the opposite is true. To provide a computational demonstration, we considered the problem of extracting the boundaries of an object, such as a face, from the image. It turns out that recognition helps, and we proved it by providing a working algorithm. This demonstration is directly related to the arguments in the magic theory of 2011 about modularity and class-specific modules.

Learning Theory and More Applications

Five years had passed since we published our papers with Federico Girosi on learning and regularization, and it was time for an up-to-date review that connected spines to learning, Gaussian kernels to sigmoids and integrals, and unsupervised density estimation with Parzen windows to regularization. Partha Niyogi was a great PhD student who had been working with me but mainly with Federico. Together with Federico, he brought together techniques from approximation theory and statistics to address the key question of generalization performance of learning algorithms. They showed that in order to understand the complexity of the problem, one needed results from both approximation and statistical learning theory. It is a paper that got many mathematicians involved in learning theory, presenting them with a clean description of a challenging problem.

Before Partha left for ATT in 1997, he and Federico and I wrote a paper on virtual examples, showing a kind of equivalence between information in new examples and information in the regularizer. Together with Kah-Kay Sung, who died prematurely, Partha and I teamed up to compare radial basis functions (RBFs) and support vector machines (SVMs) in a paper that was probably and wrongly biased in support of SVMs. Eventually, Partha went on to join the faculty in CS and statistics in Chicago and to get early tenure. After Partha left MIT, I had the good fortune to work with him and see him quite a bit more. He was one of us, a small number of researchers including Steve Smale, Felipe Cucker, Federico Girosi, and me, who made supervised learning theory an important part of mainstream mathematics. He carried on a similar mathematization program for manifold learning. Tragically, Partha died in 2010. His death stopped the final formulation of an ambitious framework to analyze high-dimensional data in terms of geometry, sparsity, and hierarchy. As a person and a scientist he was unique. He was curious about all things scientific, and he was never afraid of learning new tools, no matter how complex, if he thought they were needed.

During those same years, Federico discovered Vladimir Vapnik and his pioneering work in the theory of learning. We invited him to give a talk and hosted him and his linguist girlfriend for a week at MIT. Vladimir was writing his book and asked Federico for help. At lunch at Legal Seafood, Federico told Vladimir about reproducing kernel Hilbert spaces (RKHS) and clarified some important aspects of this work.

Minds of Friends

My love for science was born with a fascination for the scientists behind the great discoveries. There are many reasons for this. Above all is my interest in the problem of intelligence and what is behind the mind of a genius. I also realized very soon that science is fun and interesting not only because of the highs induced by scientific discoveries—these are equivalent to winning a sporting event. It is also fun because the process of research makes it easy to interact with smart people.

It was also clear to me early on that science is social—the papers and especially the books are just a history of events, recounting and inevitably deforming what really happened. The history of our ideas is like our memories. Unlike computer memory, retrieving one of our memories also means modifying it.

Over the years, I have interacted and collaborated with prominent people from diverse fields. I am presently collaborating with Steve Smale, a Field medalist mathematician. My lab has also spawned people destined to fame (my first graduate student was the well-known neuroscientist Christof Koch) who themselves are innovators in diverse fields: industry, engineering, finance, business, computer science, robotics, Web search engines, and more. For fun, I have listed all the Nobel Prize winners whom I have met during my career. I know some of them quite well. The list includes David Hubel, Thorston Wiesel, Eric Kandel, Bob Horvitz, Max Delbrück, Niels Jerne, Susumu Tonegawa, Phil Sharp, Richard Feynman, Bernard Katz, Hadlan Hartline, Alan Hodgkin, John Eccles, Francis Crick, Janni Nusslein, Don Glaser, Dan Kahnemann, Bob Merton, Paul Samuelson, Jim Watson, Gerry Edelman, Rita Levi-Montalcini, Dennis Gabor, Franco Modigliani, Bob Solow, Manfred Eigen, Sidney Brenner, Erwin Neher, Bert Sakmann, Frank Wilczek, and David Baltimore. Most of them are very smart; I know that at least two or three of them were extraordinarily smart: Francis Crick, Paul Samuelson, and Richard Feynman.

Apart from the list here, I have had the extraordinary luck to meet several extraordinary scientists and human beings, some of whom are no longer with us. Antonio Borsellino and David Marr are among them. Werner Reichardt is another great person. It is impossible for me to speak about Werner Reichardt only as a scientist. He was the most influential person in my life, both scientifically and personally. He was first and foremost my scientific mentor. Werner was a loyal friend whose word could always be fully trusted. He was a gentleman of science educated in the tradition of the great German schools. I always admired his intellectual honesty and his courage; he was someone you could trust in the small and great needs of life. I will never forget the long discussions with him in his office and at his home, mainly about scientific work and sometimes about his passion to make a difference in the politics of science and in the Max Planck Society—about which he cared deeply.

When Francis Crick died, there were many articles about him. They were mostly memories of the charismatic personality, the brilliant mind, and the great scientist, and the stories of his discoveries in molecular biology. After all, the names of Watson and Crick will be with us as long as Einstein's and Planck's. There were ample reasons for Francis to have been arrogant, but my memories of him are quite different. Arrogant? The reverse was true. I will remember forever what I would call intellectual generosity. He did not suffer fools who made claims based on flimsy evidence. Any scientific theory must be based on hard facts. He was, however, extremely patient with curious people. I remember him after a talk at MIT. He must have been quite tired, yet he patiently and gently answered what seemed an endless series of questions from students and random people who stayed on after the huge audience had left. He was a gentleman of science. I remember the many working conversations at breakfast in Odile and Francis's kitchen overlooking the pool and more conversations next to it. It was a civilized, though deceptively relaxed, style of research. Most of all, I will remember forever his joy for life and for science—a wonderful gift from Francis and Odile to their friends. I am not sure he knew but visiting them, in the last few years in their house in La Jolla, was for me—two decades their junior—a unique way to refresh my mind, recharge my batteries, and to regain a sense of adventure and fun in research and in life. His courage was exemplary. He did not seem to worry about his disease and certainly he did not bother his friends. Of course, he knew well that he was given the great gift of a wonderful life. His enthusiasm for science and for conversation with friends continued unabated. A few weeks before he died, he talked to me over the phone—full of wonder and excitement about a new puzzle he was working on, a new paper he was just writing.

Don Glaser was another remarkable friend. Don passed away on February 28, 2013, at his home in the Berkeley Hills. Don won the Nobel Prize in Physics in 1960 for the discovery of the ingenious bubble chamber, made when he was 26. Later he co-founded Cetus, one of the very first biotech companies, acquired years later by Novartis. I always admired the grace with which Don carried his success in science and in the high-tech industry. Somehow, for his whole life, he managed to retain the happy curiosity of a child. For him, the world was a garden of wonders. He was always able, in an apparently effortless way, to come up with refreshingly counterintuitive observations of things and of people.

Partha's life was much too short. He was a student of mine. He was generous, super smart, and a great and trusted friend. Partha was a pure intellectual with a strong family tradition in social and political commitments. He cared deeply about people and about science. He was a scholar with a broad and deep knowledge of mathematics and linguistics. He was my guide in many questions about science and math. Among all statisticians I know, he was the one I most admire and respect. In our field of machine learning, he knew more than anybody else and understood better than anybody else. His mind was as clear and crisp as a September day.

The New Century: 2000–10

Models of Visual Cortex

Just before the turn of the century (1997 or so), I was doing a six-month sabbatical at the Salk Institute, where I met very often with Francis Crick. Nothing specific came out of those meetings, but I began to dream about the visual cortex as a complex system adapting to the data. A bright new student had come on the scene, and I tried hard to convince him to choose MIT for his graduate studies. Max Riesenhuber was the son of a previous German minister of research and a fellow of the Studienstiftung des deutschen Volkes, just as Christof had been. Max's task, after a somewhat failed project with Peter Dayan, was to simulate a two- or three-layer network consisting of filters and invariance operations. The conjecture was that such a system would work for object recognition by replicating the invariance properties for paperclips found by Nikos Logothetis in inferotemporal cortex (IT) neurons in the monkey. My main intuition came while running on the beach in La Jolla early in the morning outside of the condo that I was renting, right behind the La Jolla Shores Hotel. The intuition was inspired by the success of the strategy we had used for face detection, and more recently in pedestrian detection with Constantine Pageorgiou, which had so much impressed Jitendra Malik. In those systems, we used a scanning procedure to perform position invariant detection. The focus of processing was shifted in the x and y axes in lexicographic order, in this way scanning the whole image and performing classification at each step. Lexicographic scanning does not make much sense for the brain but is natural in computer vision, though apparently it became standard only later. My intuition was that the cortical equivalent of scanning might be a max operation performed in a hierarchical way over larger and larger neighborhoods. Max got the system to work. It was a toy system, but it managed to reproduce all of the basic properties found by Nikos. In addition, it worked much better than I had expected. I had begun working with object recognition in the early 1980s as a short project to disprove what I thought was a simplistic theory—the idea that a hierarchy of simple and complex visual cells in the visual cortex pass information up to hyper-complex cells. But in 1997, when we found that hierarchical processing worked better than expected, I got stuck on working on feedforward models of how the brain recognizes objects.

At the time, we did not know that Fukushima had implemented a similar system much earlier (even if it was not as faithful to the physiology as ours was). The work with Max, which led to a family of models dubbed HMAX by Mike Tarr, began a research thread in my group that is active today. The immediate aftereffect of the surprising success of the model was our attempt to get physiologists to do experiments to verify or falsify the model or to simply answer the questions it suggested. This is the type of interplay between theory and experimentation that I had learned from Werner Reichardt and in which I very much believe. This research was a main motivation for embarking in a big effort to get funding from NIH for a Conte Center in Neuroscience to fund collaborations on the topic of object recognition. This effort led to a breakup with Nikos Logothetis, who was angry to discover that Earl Miller was describing projects in which he was involved without asking his permission. It also led to my collaboration with Earl about processing stages beyond area IT for the categorization step described by HMAX. We also began collaborating with David Ferster and

Ilan Lampl, convincing them to do an experiment with complex cells in cat visual cortex (V1) that seemed to provide support for a max-like operation assumed by our model.

It was natural to try to extend this model to dorsal stream visual processing in the parietal cortex. That meant working with image sequences. This was the project that Martin Giese, a postdoctoral fellow recommended by von Seelen, had been working on with a computer vision project. It became an interesting class of models combining motion and shape in the recognition of objects and motion sequences and relating them to cortical areas.

Some of the ideas generated by those models were reviewed in a paper with Emilio Bizzi, a scientist and friend whom I respect enormously but with whom I never before had the opportunity to collaborate. The paper expanded on the ideas presented earlier at Cold Spring Harbor regarding how the brain might work. A few ideas were, I think, particularly interesting. The first was that Gaussian-like tunings might play a key role in generalization. The second idea was that Gaussian tunings could be obtained via normalized dot products. To me, this meant that normalization operations, instead of being computationally trivial, might play a key role in learning and generalization. Furthermore, very similar circuits, for instance those based on silent or shunting inhibition, could implement both normalized dot products and max-like operations, as shown later in more mathematical detail by Minjoon Kouh in his thesis.

Videorealistic Speech Synthesis

The most successful project in videorealistic speech synthesis began with Steve Librande and was developed by Tony Ezzat, first in his master's thesis and then in his PhD work with a system called Mary101. The system is still impressive today in its capacity to synthesize realistic video sequences of a specific person saying things that were never actually said. The system was tested in a version of the Turing test and it passed it: naïve observers could not decide whether the video sequence was real or synthetic. The paper was accepted at SigGraph, which was a major accomplishment because acceptance at SigGraph was then more difficult than *Nature* or *Science*. The *Boston Globe* learned about videorealistic speech synthesis from me when science writer Gareth Cook asked me about the new McGovern Institute. Gareth then wrote an article that appeared on the front page of the *Globe* emphasizing the ethical rather than the technological issues. I was a bit irritated initially, but it ended up being great PR. Other newspapers and magazines, hundreds of them across the globe, picked up the news. Television crews came. Tony and I found ourselves on the "Today Show" chatting (and flirting) with Katie Couric. I think that was the only occasion in which even non-scientist friends thought that I was doing something important!

Genomics and Machine Learning

When Sayan Mukherjee appeared at the Center for Biological and Computational Learning (CBCL) for a summer job, he came with negative recommendations from his previous boss and looked like a Tamil Tiger member. We had a thoughtful discussion with Federico on whether to take the calculated risk of accepting him. We did accept him, and I think it is one of the best risks I have ever taken. Life was not easy with Sayan. He consistently rejected anything that was organized. As a graduate student in the department of brain and cognitive sciences, he refused to have anything to do with neuroscience. I tried to get him to work in computer vision but had to give up. As a last desperate attempt to get him working, I took him with me to see my old friend Jill Mezirov, then working with Eric Lander at the Whitehead on the Human Genome project. Sayan fell in love with genomics and statistics, and the rest is history. One of the best compliments I ever received was at Duke a few years ago when some faculty members found out that I was Sayan's advisor. They said, "You should send us many more guys like Sayan!" Together with Gadi, Sayan created a special CBCL atmosphere that benefited other great students such as Shasha Rakhlin and Gene Yeo. It was a combination of heavy-duty mathematics, even if not always correct, a deeply felt anti-Bayes attitude, and total disinterest in neuroscience.

Object Detection and Machine Learning

Until 2005 or so, I was involved in two informal subgroups, one focused on the brain and one focused on computer science and machine learning applications. In those years, the main application domain was object detection. Following Kah-Kay's work on face detection, I decided to tackle people detection. It was a lucky choice because of the contacts I had as a consultant with Daimler Benz (of Mercedes Benz). In discussions with their technical management, I learned that cars and pedestrians were important objects for a camera-driven car. At the time, I was looking for object classes for research and had bought a few rubber toys including dinosaurs and other animals, but I was unable to come up with interesting uses for them. Pedestrian detection for cars seemed a quite good application for object detection. I asked Pawan Sinha for suggestions, and he came up with a set of templates that to me looked like Haar wavelets. Thus, we used a wavelet representation of images and templates obtained from examples of pedestrians and trained a classifier to detect people. In the hands of Constantine Papageorgiou, the system worked quite well. He spent a month in Germany implementing a faster and coarser version that ran on a PC in the trunk of an experimental Mercedes. Anuj Mohan, another great student who died tragically years later in Silicon Valley, improved on the system and made it more robust to occlusions by using an architecture comprising a hierarchy of SVM classifiers.

Learning Theory

Another five years had passed since the updated review of our progress on regularization for learning, and it was time for a third review. Federico had left for a second PhD and a second career, so I led the effort. I sweated blood for that paper in *Advances in Computational Mathematics*; neither Theos nor Massi helped much. It was a nice paper that ended with a number of interesting contributions. The most obvious one was to show that SVMs are a special case of regularization. This fact had been ignored by Vladimir Vapnik, who knew but perhaps did not want others to know, and by the rest who did not want to understand.

Somewhat later in 2000, I was invited to a workshop in Hong Kong celebrating the 70th birthday of Steve Smale. On the occasion of the symposium, his collected work was published in three volumes, more than 1,600 pages, and it was all math. The reason for publishing all of Steve's work was that Steve is one of the greatest mathematicians of our time. I quote from the preface, "Many mathematicians have contributed powerful theorems in several fields. Smale is one of the very few whose work has opened up to research vast areas that were formerly inaccessible."

The reason I was invited to the workshop is that a couple of years earlier Steve had started to work in the field of statistical learning. His first paper published in the *AMS* in 2001 was circulating at the time of the Hong Kong workshop as a preprint with the title "The Mathematical Foundations of Learning." It had the stated goal of bringing learning theory into the mainstream of mathematics. I was quite proud that below the title there was a quotation from a review paper by Shelton and me that said, "The problem of learning is arguably at the very core of the problem of intelligence, both biological and artificial." I had met Steve in Tübingen and again in Berkeley, but we did not really know one another. The workshop in Hong Kong and our common interest in learning theory brought us together in a friendship and in many reciprocal visits.

Francis Crick and Steve Smale are two of my scientific heroes. I admire both of them not only because of their scientific achievements but also because of their refreshingly young attitude toward life. Steve is not only a great mathematician but has given to his friends and collaborators the wonderful gift of a contagious joy for life and for science.

Stability in Learning

I have a long-standing interest in the epistemological resonances of learning theory, an area of thinking that had been raised earlier by Vapnik in the following way. The process of learning from example is a metaphor or perhaps a cartoon of science itself. Get experimental data (i.e., the examples), which can be thought of as input-output pairs, and then infer a theory or model that explains the data and that can predict the future. Without prediction, astrology and even economics would be scientific. What makes a field scientific is that a model inferred from data must hold for new data (i.e., the model must generalize and be predictive). Note that, from this point of view, models and theories are what make data useful—they help us to predict and to plan. Note the close relationship between prediction and compression: theories are a compressed version of the data because they allow prediction of much data without having to memorize them. What interests me most, because of my work on ill-posed problems, is the observation that scientific theories are "stable" for the same reason. Because they are supposed to predict future data, they should not change when new data come in. In this sense, the theory within a scientific domain (e.g., mechanics) should change only a little bit—mostly in response to new experiments. For example, Newton's laws should change little, apart from low-probability events such as relativity.

On the surface, most of these statements appear to be tautologies. But in mathematics, tautologies that are proven formally become theorems, and the conditions under which they are valid are often important and far from obvious. In this spirit, around the year 2000, I realized that in my learning class we were motivating the use of regularization because learning is an ill-posed problem; that is, it is a problem for which the solution either does not exist or is not unique or is unstable. At the same time, I was also using regularization to obtain predictivity and generalization properties. Were well-posedness and generalization equivalent? At a certain level, they have to be. More precisely, when we formulate a learning problem, we require that the solution is a function that fits the data and generalizes; we also require that the solution is unique and is stable against perturbations in the data. These seem to be separate requirements, but are they? It is easy to see that ensuring uniqueness and existence is simple: the key condition is stability. So the question becomes: Do specific definitions of stability and of predictivity exist for which it is possible to prove that they are equivalent? It turned out that a few applied mathematicians had addressed stability properties in learning for somewhat different reasons. A paper by Bousquet and Elisseef, for which I was a (very positive) reviewer for NIPS, derived generalization bounds from a very strong definition of stability. There were also elegant papers by Partha Niyogi.

Together with Sayan, Rif, and Partha, I spent more than a year getting deep into technicalities and proving results that often turned out to be wrong. I did not sleep well and had nighmares about incorrect theorems. The usual routine was as follows. I formulated a new theorem and half-proved it. Sayan found a way to finish the proof, and Rif showed that it was wrong, usually by using a clever counterexample. Eventually, we got a few results right, at which point Partha distilled the whole into a few nice crystals. It was for us like a long sailing trip through storms and adversities that bound us for life. I am quite proud of the end product. Stability is necessary and sufficient for predictability! A paper was published in *Nature*; I think it is quite rare to have a paper in *Nature* on a theorem; however, it was pretty much ignored by the scientific community. I still think it is important and that eventually people may realize that. In fact, I am dreaming that in one of the future projects it may be possible to show that well-posedness, and in particular stability, are the connection between computer science and physical science, between bits and atoms, tying together notions such as models of computation, intelligence, scientific theories, generalization, compression, and the arrow of time.

Development of HMAX and the elusive theory

The model of the ventral stream developed initially with Max Riesenhuber was becoming more faithful to cortical data and more complex in the hands of Thomas Serre, a very effective PhD student. In terms of object recognition, it seemed to perform on par with the best computer vision algorithms. I became really interested when Lior Wolf, then a postdoctoral fellow in my computer science subgroup, told me quite seriously that HMAX was doing very well. In the following months, I realized that he was right. For the first time in my life, I saw a model that captured what we knew about a segment of neuroscience that performed better than engineered systems. Because of this, I no longer have separate computer science and neuroscience subgroups. We may be at that point in time when we know enough about the brain to actually inform engineering to guide some of this research.

For me this is a new paradigm, one maintaining that a computational model based on how the brain works is more powerful than one based on engineering principles; moreover, it can be used to test hypotheses about the brain itself. In fact, I was quoted as saying, "It's a virtuous loop, because the model that is based on what neuroscientists know about how the brain works is also helping neuroscientists learn even more about the how the brain works. You need theories to drive experiments, and you need experiments to test the theories. I think we are now at the stage where we can use brain-based models to design and interpret neuroscience experiments for studying human behavior, and also mental disorders, and maybe someday improving human intelligence." A study with Serre and Oliva involved flashing a variety of real-world photos and asking the model whether an animal was present. This was a difficult task because the animal could be a lizard, tiger, or bee, and it could be a lone profile or a detail in a crowded landscape. In both cases, the model performed just as well as human subjects, who were shown the same images. The model even made the same kinds of errors that the people made on ambiguous images. The model outperformed other computer programs designed by engineers.

Though Thomas was the key person in the new effort around HMAX, a number of other students and postdoctoral fellows contributed greatly, among them Gabriel Kreiman, Minjoon Kouh, Charles Cadieu, and Ulf Knoblich. The model could reproduce human performance in object recognition in conditions under which eye movements and detailed inspection were not allowed (i.e., rapid image presentations). The model was consistent with results from all the main cortical areas involved in vision, and it predicted several properties, including population invariant coding of object identity and category, in studies I had planned and for which I provided funding. (The experiments were eventually done by Chou Hung and Jim DiCarlo and analyzed by Gabriel Kreiman). A similar model for the dorsal stream matched human performance for classifying actions. This last feat required a daunting amount of computational power because it was processing not individual images but sequences of movements each with hundreds of image frames.

The model was working well. I think feedforward architectures such as HMAX are close to explaining feedforward processing in the visual cortex. Achieving human performance under general unrestricted conditions is close to achieving human intelligence, which requires computations over longer times, top-down connections, memory retrival, and nonvisual areas. I bet that two-thirds of the circuitry in the visual cortex is involved in the management and execution of non-feedforward processing. A small step in the direction of modeling recurrent connections is related to what is commonly called attention. We added more layers of complexity to the model, incorporating the influence of attention when scanning a scene cluttered by many objects. Other computer vision models become inefficient or fail at this challenging data-crunching task, but our model adopted the same strategy that human subjects use (as measured by tracking their eye movements)—homing in on areas of the image most likely to contain objects of interest and ignoring the rest. This move makes information processing more efficient and validates neuroscience research focused on how attention impacts perception.

On the theoretical side, the inability to understand why HMAX-type architectures work so well had become embarrassing. Finally, Steve Smale came up with a clean implicit neural equation and derived kernel formalizing hierarchical architectures in a way that should make it possible to study them mathematically. Sometimes models and simulations are not enough, and theories are needed. Unfortunately, this brave attempt led to a clean mathematical formulation but little more.

Phenotyping Behavior

The field of genomics has made staggering progress in recent years. Because of the genetically engineered mice used as models of mental diseases in our building at MIT and elsewhere, we understand the genome in a quantitaitve reproducible way. The behavioral phenotype on the other hand is understood only qualitatively most of the time. Ironically, phenotyping is the bottleneck in today's genomics. If we could phenotype behavior as quantitatively and easily as we obtain the genotype, we would have a gold mine of data among which we could find important, subtle correlations. With this motivation in mind, we decided with Thomas Serre and Hueihan to adapt the model of the dorsal stream to categorize behavior of a mouse in a cage. The system recognizes and categorizes sets of behaviors by inventorying how often and when a laboratory mouse eats, grooms, explores, and so on. Again, the model was as accurate as human subjects. There is still much work to be done to make this system more robust to changes in illumination, and especially to extend it to groups of mice and their social interactions. It is quite feasible in a few years to be able to replace all the specific boxes and tasks used to measure specific behavioral traits with a general purpose vision system and video analysis.

The New Century: 2010 and Beyond

Intelligence in Brains, Minds, and Machines

The idea of an Intelligence Initiative was initiated by Josh Tenenbaum and by me in discussions with Rafael Reif, Mriganka Sur, and Marc Kastner at MIT. From the earliest stages, Mriganka and Marc strongly supported the idea and pushed Josh and me to make something happen. Trying to spread an I^2 infection at MIT, we discussed the idea with a number of MIT faculty, MIT friends, and agencies such as the National Science Foundation (NSF). A workshop at the American Academy of Arts and Sciences elicited a surprising amount of enthusiasm among MIT faculty. At the last moment, I sent in a proposal for a MIT150 workshop, which was accepted. The organization of the workshop was a lot of work, and it happened only because of my administrative assistant Kathleen and her incredible organization and energy. It was a big success. The idea of I^2 is simple: the problem of intelligence is the greatest problem in science and technology. The last serious attempt to do something about it was at the time of AI, 50 years ago. Much has changed since then in terms of computer power, neuroscience, and technologies. It is now time to try again! After a period of seed funding for projects across the institute, the initiative is now formalized in a Center of Brains, Minds, and Machines funded by the NSF.

I believe that the next few years will be a golden age for smart machines and smart applications such as Siri and future versions of Siri. This technology is the marketplace outcome of technologies developed in laboratories around the world, including ours, that have been focused on machine learning over the last two decades. However, we will not be able to build "really intelligent" machines unless we make substantial progress in a new cycle of basic research. I use the term "really intelligent" in the sense of human intelligence, that is, in the sense of being able to be like one of us and to pass a "full" Turing test. This is what we ultimately want to achieve in the Center for Brains, Minds, and Machines.

The Magic Theory

Two years ago, around March 2011, a number of ideas concerning a revision of the theory that we had tried to develop with Steve Smale began to converge in my head. The focus then had been on features; I decided to make a 180-degree turn and focus on invariances. The idea that the main computational goal of the ventral stream was to compute an invariant representation suddenly became very plausible. I asked Joel for a plausibility demonstration; how well can a simple classifier generalize to the concept of dog and of cat from a single example of a dog and a cat? Normally the answer would be "not at all." Generalization performance is 50 percent (i.e., chance). Suppose however that an oracle rectifies the images so that they are always given from a standard viewpoint. Then suddenly generalization performance from a single example is quite good!

Many other related ideas began to click into place. Invariances could be easily learned during development, the tuning of V1 simple cells emerges from it (assuming Hebbian learning), the modularity of cortex can be predicted, and the obscure properties of face cells suddenly make sense. Around this time, I began to use the term "magic theory" because so many things suddenly made sense. Every day seemed to bring a new idea, a new twist, a new connection, and a new question. Of course, there were quite a few wrong turns in the process, such as the initial emphasis on factorization of invariances. It has been quite exhilarating for me. Unlike other projects, this one is quite big and involves a large number of ideas and mathematical tools. I managed to take one problem at the time and to solve it or decide that it was not important to making continuous and real progress. Initially, it was my personal project. I wrote a long technical report that appeared online in July 2011. Then I brought in several members of the group to help: Jim Mutch, Joel Leibo, Fabio Anselmi, Lorenzo Rosasco, and Andrea Tacchetti. I am a physicist. What I find most intriguing and potentially elegant about these theoretical results is that the computational goals, the hierarchical architectures, and that several of the tuning properties of cells in the ventral stream may follow from symmetry properties of the visual world.

References

Ancona, N. and T. Poggio. Optical Flow from 1-D Correlation: Application to a Simple Time-to-Crash Detector, *International Journal of Computer Vision*, 14, 131–146, 1995.

Anselmi F., J.Z. Leibo, L. Rosasco, J. Mutch, A. Tacchetti, and T. Poggio. Magic Materials: A Theory of Deep Hierarchical Architectures for Learning Sensory Representations, *CBCL paper*, Massachusetts Institute of Technology, Cambridge, MA, April 1, 2013.

Bertero, M., T. Poggio, and V. Torre. Ill-Posed Problems in Early Vision, *Proceedings of the IEEE*, 76, 869–889, 1988.

Beymer, D. and T. Poggio. Image Representation for Visual Learning, *Science*, 272, 1905–1909, 1996.

Borsellino, A. and T. Poggio. Convolution and Correlation Algebras, *Kybernetik*, 13(2), 113–122, 1973.

Borsellino, A. and T. Poggio. Holographic Aspects of Temporal Memory and Optomotor Responses, *Kybernetik*, 10(1), 58–60, 1972.

Brunelli, R., D. Falavigna, T. Poggio, and L. Stringa. Automatic Person Recognition by Acoustic and Geometric Features, *Machine Vision and Applications*, 8(5), 317–325, 1995.

Brunelli, R. and T. Poggio. Caricatural Effects in Automated Face Perception, *Biological Cybernetics*, 69, 235–241, 1993.

Brunelli, R. and T. Poggio. Face Recognition: Features Versus Templates, *IEEE PAMI*, 15, 1042–1052, 1993.

Brunelli, R. and T. Poggio. HyperBF Networks for Gender Classification. In: *Proceedings of the Image Understanding Workshop*, (IU Meeting in San Diego, CA, January, 1992), Morgan Kaufmann, San Mateo, CA, 311–314, 1992.

Brunelli, R. and T. Poggio. Template Matching: Matched Spatial Filters and Beyond, *Pattern Recognition*, 30(5), 751–768, 1997.

Bülthoff, H.H., J. Little, and T. Poggio. A Parallel Algorithm for Real Time Computation of Optical Flow, *Nature*, 337, 549–553, 1989.

Bülthoff, H.H., T. Poggio, and C. Wehrhahn. 3D Analysis of the Flight Trajectories of Flies (Drosophila Melanogaster), *Z. Naturforsch*, 35c, 811–815, 1980.

Cadieu, C., M. Kouh, A. Pasupathy, C. Connor, M. Riesenhuber, and T. Poggio. A Model of V4 Shape Selectivity and Invariance, *Journal of Neurophysiology*, 98, 1733–1750, 2007.

Chikkerur, S. and T. Poggio. Approximations in the HMAX Model, MIT-CSAIL-TR-2011-021/CBCL-298, Massachusetts Institute of Technology, Cambridge, MA, April 14, 2011.

Crick, F. H., D.C. Marr, and T. Poggio. An Information-Processing Approach to Understanding the Visual Cortex. In: *The Organization of the Cerebral Cortex*, E.O. Schmitt, F.G. Worden, and G.S. Dennis (eds.), pp. 505–533, Cambridge, MA: MIT Press, 1980.

Cucker, F. and R. Wong (eds.), *The Collected Papers of Stephen Smale*. Singapore University Press, World Scientific Publishing, 2000.

Dahan, E., A.J. Kim, A.W. Lo, T. Poggio, and N. Chan, Securities Trading of Concepts (STOC), *Journal of Marketing Research*, 48(3), 497–517, 2011.

De Mori, R. and T. Poggio. Information Retrieval e Biblioteche Digitali, *Technology Review: Edizione Italiana*, 13, 1, 2000.

Edelman, S. and T. Poggio. Models of Object Recognition, *Current Opinion in Neurobiology*, 1, 270–273, 1991.

Edelman, S. and T. Poggio. A Network that Learns to Recognize 3D Objects, *Nature*, 343, 263–266, 1990.

Evgeniou, T., M. Pontil, C. Papageorgiou, and T. Poggio. Image Representations and Feature Selection for Multimedia Database Search, *IEEE Transactions in Knowledge and Data Engineering*, 15(4), 911–920, 2002.

Evgeniou, T., M. Pontil, and T. Poggio. Regularization Networks and Support Vector Machines, *Advances in Computational Mathematics*, 13(1), 1–50, 2000.

Evgeniou, T., M. Pontil, and T. Poggio. Statistical Learning Theory: A Primer, *International Journal of Computer Vision*, 38(1), 9–13, 2000.

Ezzat, T., G. Geiger, and T. Poggio. Trainable Videorealistic Speech Animation*, ACM SIGGRAPH 2002*, San Antonio, TX, July 2002.

Ezzat, T. and T. Poggio. Visual Speech Synthesis by Morphing Visemes, *International Journal of Computer Vision*, 38(1), 45–57, 2000.

Fahle, M., S. Edelman, and T. Poggio. Fast Perceptual Learning in Visual Hyperacuity, *Vision Research*, 35(21), 3003–3013, 1995.

Fahle, M. and T. Poggio. Visual Hyperacuity: Spatiotemporal Interpolation in Human Vision, *Proceedings of the Royal Society London B*, 213, 451–477, 1981.

Freedman, D.J., M. Riesenhuber, T. Poggio, and E.K. Miller. Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex, *Science*, 291, 312–316, 2001.

Freedman, D.J., M. Riesenhuber, T. Poggio, and E.K. Miller. Comparison of Primate Prefrontal and Inferior Temporal Cortices during Visual Categorization, *Journal of Neuroscience*, 23, 5235–5246, 2003.

Freedman, D.J., M. Riesenhuber, T. Poggio and E.K. Miller. Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex, *Cerebral Cortex*, 16(11), 1631–1644, 2006.

Freedman, D.J., M. Riesenhuber, T. Poggio, and E.K. Miller. Visual Categorization and the Primate Prefrontal Cortex: Neurophysiology and Behavior, *Journal of Neurophysiology*, 88, 930–942, 2002.

Gamble, E., D. Geiger, T. Poggio, and D. Weinshall. Integration of Vision Modules and Labeling of Surface Discontinuities, IEEE Trans., *Systems, Man and Cybernetics*, 19(6), 1576–1581, 1989.

Geiger, G. and T. Poggio. Asymptotic Oscillations in the Tracking Behavior of the Fly Musca Domestica, *Biological Cybernetics*, 41, 197–201, 1981.

Geiger, G. and T. Poggio. The Muller-Lyer Figure and the Fly, *Science*, 190, 479–480, 1975.

Geiger, G. and T. Poggio. On Head and Body Movements of Flying Flies, *Biological Cybernetics*, 25(3), 177–180, 1977.

Geiger, G. and T. Poggio. The Orientation of the Fly toward Visual Patterns: On the Search for the Underlying Functional Interactions, *Biological Cybernetics*, 19, 39–54, 1975.

Giese, M.A. and T. Poggio. Morphable Models for the Analysis and Synthesis of Complex Motion Pattern, *International Journal of Computer Vision*, 38(1), 59–73, 2000.

Giese, M.A. and T. Poggio. Neural Mechanisms for the Recognition of Biological Movements, *Nature Neuroscience Review*, 4, 179–192, 2003.

Girosi, F., M. Jones, and T. Poggio. Regularization Theory and Neural Networks Architectures, *Neural Computation*, 7(2), 219–269, 1995.

Girosi, F. and T. Poggio. Networks and the Best Approximation Property, *Biological Cybernetics*, 63(3), 169–176, 1990.

Girosi, F. and T. Poggio. Representation Properties of Networks: Kolmogorov's Theorem Is Irrelevant, *Neural Computation*, 1, 465–469, 1989.

Heimburger, L., T. Poggio, and W. Reichardt. A Special Case of Nonlinear Interactions in the Visual System of the Fly, *Biological Cybernetics*, 21, 103–105, 1976.

Heisele, B., P. Ho, J. Wu, and T. Poggio. Face Recognition: Component-based versus Global Approaches, *Computer Vision and Image Understanding*, 91(1/2), 6–21, 2003.

Heisele, B., T. Serre, and T. Poggio. A Component-Based Framework for Face Detection and Identification, *International Journal of Computer Vision*, 74(2), 167–181, 2007.

Heisele, B., T. Serre, S. Prentice, and T. Poggio. Hierarchical Classification and Feature Reduction for Fast Face Detection with Support Vector, *Pattern Recognition*, 36, 2007–2017, 2003.

Heisele, B., A. Verri, and T. Poggio. Learning and Vision Machines, *IEEE Visual Perception: Technology and Tools*, 90(7), 1164–1177, 2002.

Hung, C.P., G. Kreiman, T. Poggio, and J.J. DiCarlo. Fast Readout of Object Identity from Macaque Inferior Temporal Cortex, *Science*, 310, 863–866, 2005.

Hurlbert, A. and T. Poggio. Rendere le Macchine (e l'Intelligenza Artificiale) in Grado di Vedere, *Sistemi Intelligenti*, 1, 75–103, 1989.

Hurlbert, A. and T. Poggio. Synthesizing a Color Algorithm from Examples, *Science*, 239, 482–485, 1988.

Hutchinson, James, PhD, February 1994; S.M., June 1986. PhD Thesis title: A Radial Basis Function Approach to Financial Time Series Analysis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.

Hutchinson, J.M., A. Lo, and T. Poggio. A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks, *Journal of Finance*, 49(3), 851–889, 1994.

Isik, L., J.Z. Leibo, J. Mutch, S.W. Lee, and T Poggio. A Hierarchical Model of Peripheral Vision, MIT-CSAIL-TR-2011-031/CBCL-300, Massachusetts Institute of Technology, Cambridge, MA, June 2011.

Isik, L., J.Z. Leibo. and T. Poggio. Learning and Disrupting Invariance in Visual Recognition, MIT-CSAIL-TR-2011-040/CBCL-302, Massachusetts Institute of Technology, Cambridge, MA, September 10, 2011.

Isik, L., J.Z. Leibo, and T. Poggio. Learning and Disrupting Invariance in Visual Recognition with a Temporal Association Rule, *Front, Comput. Neurosci.*, 6:37. doi: 10.3389/fncom.2012.00037, June 25, 2012

Isik, L., E.M. Meyers, J.Z. Leibo, and T. Poggio. Preliminary MEG Decoding Results, MIT-CSAIL-TR-2012-010,CBCL-307, Massachusetts Institute of Technology, Cambridge, MA, April 20, 2012.

Jhuang, H., E. Garrote, J. Mutch, X. Yu, V. Khilnani, T. Poggio, A.D. Steele, and T. Serre. Automated Home-Cage Behavioral Phenotyping of Mice, *Nature Communications*, 1:68, doi: 10.1038/ncomms1064, September 7, 2010.

Jones, M. and T. Poggio. Multidimensional Morphable Models: A Framework for Representing and Matching Object Classes, *International Journal of Computer Vision*, 29(2), 107–131, 1998.

Jones, M., P. Sinha, T. Vetter, and T. Poggio. Top-Down Learning of Low-Level Vision Tasks, *Current Biology*, 7(12), 991–994, 1997.

Koch, C. and T. Poggio. The Biophysical Properties of Spines as a Basis for Their Electrical Function: A Comment on Kawato and Tsukahara, 1983, *Journal of Theoretical Biology*, 113, 225–229, 1985.

Koch, C. and T. Poggio. Electrical Properties of Dendritic Spines, *Trends in Neurosciences*, 6, 80–83, 1983.

Koch, C. and T. Poggio. A Simple Algorithm for Solving the Cable Equation in Dendritic Trees of Arbitrary Geometry, *J. Neuroscience Methods*, 12, 303–315, 1985.

Koch, C. and T. Poggio. A Theoretical Analysis of Electrical Properties of Spines, *Proceedings of the Royal Society London B*, 218, 455–477, 1983.

Koch, C., T. Poggio, and V. Torre. Nonlinear Interactions in a Dendritic Tree: Localization, Timing and Role in Information Processing, *PNAS*, 80, 2799–2802, 1983.

Koch, C., T. Poggio, and V. Torre. Retinal Ganglion Cells: A Functional Interpretation of Dendritic Morphology, *Proceedings of the Royal Society London*, 298, 227–264, 1982.

Koch, C., V. Torre, and T. Poggio. Computations in the Vertebrate Retina: Gain Enhancement, Differentiation and Motion Discrimination, *Trends in Neurosciences*, 9, 204–211, 1986.

Kouh, M and T. Poggio. A Canonical Neural Circuit for Cortical Nonlinear Operations, *Neural Computation*, 20(6), 1427–1451, 2008.

Kreiman, G., C.P. Hung, A. Kraskov, R.Q. Quiroga, T. Poggio, and J.J. DiCarlo. Object Selectivity of Local Field Potentials and Spikes in the Macaque Inferior Temporal Cortex, *Neuron*, 49, 433–445, 2006.

Kuehne, T., H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A Large Video Database for Human Motion Recognition, ICCV 2011.

Lampl, I., D. Ferster, T. Poggio, and M. Riesenhuber. Intracellular Measurements of Spatial Integration and the MAX Operation in Complex Cells of the Cat Primary Visual Cortex, *Journal of Neurophysiology*, 92, 2704–2713, 2004.

LeCun, Y., D.G. Lowe, J. Malik, J. Mutch, P. Perona, and T. Poggio. Object Recognition, Computer Vision, and the Caltech 101: A Response to Pinto et al., *PLoS Computational Biology*, Posted Online March 2008.

Leibo, J.Z., J. Mutch, and T Poggio. How Can Cells in the Anterior Medial Face Patch Be Viewpoint Invariant?, presented at COSYNE 2011, Salt Lake City, UT. Available from *Nature* Proceedings at dx.doi.org/10.1038/npre.2011.5845.1.

Leibo, J.Z., J. Mutch, and T Poggio. Learning to Discount Transformations as the Computational Goal of Visual Cortex, presented at FGVC/CVPR 2011, Colorado Springs, CO. Available from *Nature* Proceedings at dx.doi.org/10.1038/npre.2011.6078.1.

Leibo, J.Z., Mutch J., and T. Poggio. Why the Brain Separates Face Recognition from Object Recognition, Advances in Neural Information Processing Systems, Granada, Spain, December 2011.

Little, J., T. Poggio, and E.B. Gamble. Seeing in Parallel: The Vision Machine, *International Journal of Supercomputer Applications*, 2(4), 13–28, 1988.

Logothetis, N.K., J. Pauls, H. Bülthoff, and T. Poggio. View-Dependent Object Recognition by Monkeys, *Current Biology*, 4(5), 401–414, 1994.

Logothetis, N.K, J. Pauls, and T. Poggio. Shape Representation in the Inferior Temporal Cortex of Monkeys, *Current Biology*, 5(5), 552–563, 1995.

Marr, D., G. Palm, and T. Poggio. Analysis of a Cooperative Stereo Algorithm, *Biological Cybernetics*, 28(4), 223–239, 1978.

Marr, D. and T. Poggio. A Computational Theory of Human Stereo Vision, *Proceedings of the Royal Society London B*, 204, 301–328, 1979.

Marr, D., and T. Poggio. Cooperative Computation of Stereo Disparity, *Science*, 194, 283–287, 1976.

Marr, D. and T. Poggio. From Understanding Computation to Understanding Neural Circuitry. In: *Neuronal Mechanisms in Visual Perception*, E. Poppel, R. Held, and J.E. Dowling (eds.), pp. 470–488, *Neurosciences Res. Prog. Bull.*, 15, 1977.

Marr, D., T. Poggio, and E. Hildreth. Smallest Channel in Human Vision, *Journal of the Optical Society of America*, 70, 868–870, 1979.

Marr, D., S. Ullman, and T. Poggio. Bandpass Channels, Zero-crossings and Early Visual Information Processing, *Journal of the Optical Society of America*, 69, 914–916, 1979.

Marroquin, J., S. Mitter, and T. Poggio. Probabilistic Solution of Ill-posed Problems in Computational Vision, *Journal of American Statistical Association*, 82, 76–89, 1987.

Meyers, E.M., D. J. Freedman, G. Kreiman, E.K. Miller, and T. Poggio. Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex, *Journal of Neurophysiology*, 100, 1407–1419, 2008.

Mohan, A., C. Papageorgiou, and T. Poggio. Example-Based Object Detection in Images by Components, *IEEE (PAMI)*, 23(4), 349–361, 2001.

Mukherjee, S., P. Niyogi, T. Poggio, and R. Rifkin. Learning Theory: Stability Is Sufficient for Generalization and Necessary and Sufficient for Consistency of Empirical Risk Minimization, *Advances in Computational Mathematics*, 25, 161–193, 2006.

Mukherjee, S., P. Niyogi, T. Poggio, and R. Rifkin. Statistical Learning: Stability Is Sufficient for Generalization and Necessary and Sufficient for Consistency of Empirical Risk Minimization, CBCL Paper #223, Massachusetts Institute of Technology, Cambridge, MA, December 2002 [January 2004 revision].

Mukherjee, S., P. Tamayo, J.P. Mesirov, D. Slonim, A. Verri, and T. Poggio. Support Vector Machine Classification of Microarray Data, AI Memo 1676/CBCL Paper #182, Massachusetts Institute of Technology, Cambridge, MA, December 1999.

Nakajima, C., M. Pontil, B. Heisele, and T. Poggio. Full-Body Person Recognition System, *Pattern Recognition*, 36, 1997–2006, 2003.

Nielsen, K.R.K. and T. Poggio. Vertical Image Registration in Stereopsis, *Vision Research*, 24, 1133–1140, 1984.

Nishihara, H.K. and T. Poggio. Hidden Cues in Random-Line Stereograms, *Nature*, 300, 347–349, 1982.

Niyogi, P., F. Girosi, and T. Poggio. Incorporating Prior Information in Machine Learning by Creating Virtual Examples, *Proceedings of the IEEE*, 86(11), 2196–2209, 1998.

Oren, M., C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. A Trainable System for People Detection. In: *Proceedings of the 1997 Image Understanding Workshop*, New Orleans, LA, 207–214, May 1997.

Palm, G. and T. Poggio. Stochastic Identification Methods for Nonlinear Systems: An Extension of the Wiener Theory, *SIAM Journal of Applied Mathematics*, 34, 524–535, 1978.

Palm, G. and T. Poggio. The Volterra Representation and the Wiener Expansion: Validity and Pitfalls, *SIAM Journal of Applied Mathematics*, 33, 195–216, 1977.

Palm, G. and T. Poggio. Wiener-like System Identification in Physiology, *Journal of Mathematical Biology*, 4(4), 375–381, 1977.

Papageorgiou, Constantine Ph.D., December 1999. Thesis title: A Trainable System for Object Detection in Images and Video Sequences, Massachusetts Institute of Technology, Electrical Engineering and Computer Science and Operations Research.

Papageorgiou, C. and T. Poggio. Trainable Pedestrian Detection. In: *Proceedings of International Conference on Image Processing*, Kobe, Japan, October 1999.

Papageorgiou, C. and T. Poggio. A Trainable System for Object Detection, *International Journal of Computer Vision*, 38(1), 15–33, 2000.

Poggio, G. and T. Poggio. The Analysis of Stereopsis, *Annual Review of Neuroscience*, 7, 379–412, 1984.

Poggio, T. (sections with J. Mutch, J.Z. Leibo, and L. Rosasco). The Computational Magic of the Ventral Stream: Toward a Theory, *Nature Precedings*, doi:10.1038/npre.2011.6117.1, July 16, 2011.

Poggio, T. Economic Models and Time Series: Learning from Examples, *Tactician (Citicorp Technology)*, 4(2), 28–31, 1991.

Poggio, T. The Levels of Understanding Framework, Revised, MIT-CSAIL-TR-2012-014, CBCL-308, Massachusetts Institute of Technology, Cambridge, MA, May 31, 2012.

Poggio, T. The Levels of Understanding Framework, Revised, *Perception*, 41, 1017–1023, 2012.

Poggio, T. Marr's Computational Approach to Vision, *Trends in Neurosciences*, 10, 258–262, 1981.

Poggio, T. On Holographic Models of Memory, *Kybernetik*, 12(4), 237–238, 1973.

Poggio, T. On Optimal Nonlinear Associative Recall, *Biological Cybernetics*, 19, 201–209, 1975.

Poggio, T. Review of Movements of the Eyes, *Quarterly Review of Biology, R.H.S. Carpenter*, 54, 118, 1979.

Poggio, T. A Theory of How the Brain Might Work. In: *Proceedings of Cold Spring Harbor Symposia on Quantitative Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 4, 899–910, 1990.

Poggio, T. A Theory of Nonlinear Interactions in Multi-Inputs (Nervous) Systems, *Exp. Brain Res.*, Supplement to Vol. 23, 163, 1975.

Poggio, T. Vision by Man and Machine, *Scientific American*, 250, 106–116, 1984.
Poggio, T. and E. Bizzi. Generalization in Vision and Motor Control, *Nature*, 431, 768–774, 2004.
Poggio, T. and R. Brunelli. A Novel Approach to Graphics, AI Memo 1354/CBIP Paper 71, Massachusetts Institute of Technology, Cambridge, MA, February 1992.
Poggio, T., M. Fahle, and S. Edelman. Fast Perceptual Learning in Visual Hyperacuity, *Science*, 256, 1018–1021, 1992.
Poggio, T., E. Gamble, and J. Little. Parallel Integration of Vision Modules, *Science*, 242, 436–440, 1988.
Poggio, T. and F. Girosi. Networks for Approximation and Learning, *Proceedings of the IEEE* (special issue: *Neural Networks I: Theory and Modeling*), 78(9), 1481–1497, 1990.
Poggio, T. and F. Girosi. Regularization Algorithms for Learning that Are Equivalent to Multilayer Networks, *Science*, 247, 978–982, 1990.
Poggio, T. and F. Girosi. A Sparse Representation for Function Approximation, *Neural Computation*, 10(6), 1445–14454, 1998.
Poggio, T, and F. Girosi. A Theory of Networks for Approximation and Learning, AI Memo 1140/CBIP Paper 31, Massachusetts Institute of Technology, Cambridge, MA, July 1989.
Poggio, T. and C. Koch. Ill-Posed Problems in Early Vision: From Computational Theory to Analog Networks, *Proceedings of the Royal Society London B*, 226, 303–323, 1985.
Poggio, T. and C. Koch. Synapses that Compute Motion, *Scientific American*, 256, 46–52, 1987.
Poggio T., J. Mutch, F. Anselmi, L. Rosasco, J.Z. Leibo, and A. Tacchetti. The Computational Magic of the Ventral Stream: Sketch of a Theory (and Why Some Deep Architectures Work), MIT-CSAIL-TR-2012-035, Massachusetts Institute of Technology, Cambridge, MA, December 29, 2012.
Poggio, T. and W. Reichardt. Considerations on Models of Movement Detection, *Kybernetik*, 13, 4, 223–227, 1973.
Poggio, T. and W. Reichardt. On the Representation of Multi-input Systems: Computational Properties of Polynomial Algorithms, *Biological Cybernetics*, 37(3), 167–186, 1980.
Poggio, T. and W. Reichardt. A Theory of Pattern Induced Flight Orientation of the Fly, Musca Domestica, *Kybernetik*, 12, 185–203, 1972.
Poggio, T. and W. Reichardt. Visual Control of Orientation Behavior in the Fly. Part II: Toward the Underlying Neural Interactions, *Quarterly Review of Biophysics*, 9, 377–438, 1976.
Poggio, T. and W. Reichardt. Visual Fixation and Tracking by Flies: Mathematical Properties of Simple Control Systems, *Biological Cybernetics*, 40, 101–112, 1981.
Poggio, T., W. Reichardt, and W. Hausen. A Neuronal Circuitry for Relative Movement Discrimination by the Visual System of the Fly, *Naturwissenschaften*, 68(9), 443–466, 1981.
Poggio, T., R. Rifkin, S. Mukherjee, and P. Niyogi. General Conditions for Predictivity in Learning Theory, *Nature*, 428, 419–422, 2004.

Poggio, T. and T. Serre. Models of Visual Cortex, *Scholarpedia*, 8(4), 3516, 2013.

Poggio, T. and C. Shelton. Learning in Brains and Machines, *Spatial Vision*, 13(2–3), 287–296, 2000.

Poggio, T. and C. Shelton. Machine Learning, Machine Vision and the Brain, *AI Magazine*, 20(3), 37–55, 1999.

Poggio, T. and S. Smale. The Mathematics of Learning: Dealing with Data, *Notices of the American Mathematical Society (AMS*), 50(5), 537–544, 2003.

Poggio, T. and V. Torre. A Volterra Representation of Some Neuron Models, *Biological Cybernetics*, 27(2), 113–124, 1977.

Poggio, T., V. Torre, and C. Koch. Computational Vision and Regularization Theory, *Nature*, 317, 314–319, 1985.

Poggio, T. and A. Verri. Introduction: Learning and Vision at CBCL, *International Journal of Computer Vision*, 38, 1, 5–7, 2000.

Poggio, T., S. Voinea, and L. Rosasco. Online Learning, Stability, and Stochastic Gradient Descent, Cornell University Library, arXiv:1105.4701v2 [cs.LG], May 25, 2011.

Poggio, T., H. Voorhees, and A. Yuille. A Regularized Solution to Edge Detection, *Journal of Complexity*, 4, 106–123, 1988.

Pomeroy, S.L., P. Tamayo, M. Gaasenbeek, L.M. Sturia, M. Angelo, M.E. McLaughlin, J.Y.H. Kim, L.C. Goumnerova, P.M. Black, C. Lau, J.C. Allen, D. Zagzag, M.M. Olson, T. Curran, C. Wetmore, J.A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J.P. Mesirov, E.S. Lander, and T.R. Golub. Prediction of Central Nervous System Embryonal Tumour Outcome Based on Gene Expression, *Nature* (Letters), 415, 436–442, 2002.

Rachlevsky-Reich, B., I. Ben-Shaul, N. Tung Chan, A. Lo, and T. Poggio. GEM: A Global Electronic Market System, *Information Systems*, 24(6), 495–518, 1999.

Rakhlin, A., S. Mukherjee, and T. Poggio. Stability Results in Learning Theory, *Analysis and Applications*, 3(4), 397–417, 2005.

Ramaswamy, S., P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, and T.R. Golub. Multiclass Cancer Diagnosis Using Tumor Gene Expression Signatures, *PNAS*, 98(26), 15149–15154, 2001.

Reichardt, W. and T. Poggio. Figure-Ground Discrimination by Relative Movement in the Visual System of the Fly, *Biological Cybernetics*, 35, 81–100, 1979.

Reichardt, W. and T. Poggio. A Theory of the Pattern Induced Flight Orientation of the Fly, Musca Domestica, II, *Biological Cybernetics*, 18, 69–80, 1975.

Reichardt, W. and T. Poggio. Visual Control of Orientation Behavior in the Fly. Part I: A Quantitative Analysis, *Quarterly Review of Biophysic*s, 3, 311–375, 1976.

Reichardt, W., T. Poggio, and K. Hausen. Figure-Ground Discrimination by Relative Movement in the Visual System of the Fly. II: Toward the Neural Circuitry, *Biological Cybernetics*, 46, 1–30, 1983.

Riesenhuber, M. and T. Poggio. Are Cortical Models Really Bound by the "Binding Problem"?, *Neuron* 24, 87–93, 1999.

Riesenhuber, M. and T. Poggio. Hierarchical Models of Object Recognition in Cortex, *Nature Neuroscience*, 2, 1019–1025, 1999.

Riesenhuber, M., and T. Poggio. Models of Object Recognition, *Nature Neuroscience*, 3, 1199–1204, 1999.

Riesenhuber, M. and T. Poggio. Neural Mechanisms of Object Recognition, *Current Opinion in Neurobiology*, 12, 162–168, 2002.

Rifkin, R., S. Mukherjee, P. Tamayo, S. Ramaswamy, C.-H. Yeang, M. Angelo, M. Reich, T. Poggio, E.S. Lander, T.R. Golub, and J.P. Mesirov. An Analytical Method for Multi-Class Molecular Cancer Classification, *SIAM Reviews*, 45(4), 706–723, 2003.

Roy, J.E., Riesenhuber, M., T. Poggio, and E.K. Miller. Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience*, 30, 8519–8528, 2010.

Schoelkopf, B., K.K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik. Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers, *IEEE Transactions on Signal Processing*, 45(11), 2758–2765, 1997.

Serre, T., G. Kreiman, M. Kouh, C. Cadieu, U. Knoblich, and T. Poggio. A Quantitative Theory of Immediate Visual Recognition, *Progress in Brain Research*, 104(15), 6424–6429, 2007.

Serre, T., A. Oliva, and T. Poggio. A Feedforward Architecture Accounts for Rapid Categorization, *Proceedings of the National Academy of Sciences USA*, 104(15), 6424–6429, 2007.

Serre, T., L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Object Recognition with Cortex-Like Mechanisms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29, 3, 411–426, 2007.

Sinha, P. and T. Poggio. I think I Know that Face . . . , *Nature* (Correspondence), 384(6608), 404, 1996.

Sinha, P. and T. Poggio. Response to "Comment" article by Lamouret, Cornilleau-Peres, and Droulez, *Cognitive Sciences*, 1(2), 43–84, 1997.

Sinha, P. and T. Poggio. Role of Learning in Three-Dimensional Form Perception, *Nature*, 384(6608), 460–463, 1996.

Sinha, P. and T. Poggio. United We Stand: The Role of Head Structure in Face Recognition, *Perception*, 31(1), 133, 2002.

Smale, S., L. Rosasco, J. Bouvrie, A. Caponnetto, and T. Poggio, Mathematics of the Neural Response, *Foundations of Computational Mathematics*, 10, 67–91. 2010.

Sung, K.-K. and T. Poggio. Example-Based Learning for View-Based Human Face Detection. In: *Proceedings from Image Understanding Workshop*, (Morgan Kaufmann, San Mateo, CA), 843-850, November 13–16, 1994.

Sung, K.-K. and T. Poggio. Example-Based Learning for View-Based Human Face Detection, *IEEE PAMI*, 20(1), 39–51, 1998.

Sung, K.-K. and T. Poggio. *Finding Human Faces with a Gaussian Mixture Distribution-Based Face Model, Recent Progress in Computer Vision*, LNCS Series, New York: Springer-Verlag, 1995.

Tan, C. and T. Poggio. Faces as a Model Category for Visual Object Recognition, MIT-CSAIL-TR-2013-004, CBCL-311, Massachusetts Institute of Technology, Cambridge, MA, March 18, 2013.

Torre, V. and T. Poggio. On Edge Detection, *IEEE Trans. PAMI*, 8, 147-163, 1986.

Torre, V. and T. Poggio. A Synaptic Mechanism Possibly Underlying Directional Selectivity Motion, *Proceedings of the Royal Society London B*, 202, 409–416, 1978.

Verri, A. and T. Poggio. Motion Field and Optical Flow: Qualitative Properties, *IEEE Trans. PAMI*, 11, 490–498, 1989.

Verri, A. and T. Poggio. Regularization Theory and Shape Constraints, AI Memo 916, Massachusetts Institute of Technology, Cambridge, MA, September 1986.

Vetter, T., A. Hurlbert, and T. Poggio. View-Based Models of 3D Object Recognition: Invariance to Imaging Transformations, *Cerebral Cortex*, 5(3), 261–269, 1995.

Vetter, T. and T. Poggio. Linear Object Classes and Image Synthesis from a Single Example Image, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(7), 733–742, 1997.

Vetter, T. and T. Poggio. Symmetric 3D Objects Are an Easy Case for 2D Object Recognition, *Spatial Vision*, 8(4), 443–453, 1994.

Villa, S., L. Rosasco, and T. Poggio, On Learnability, Complexity and Stability, arXiv 1303.5976 March 24, 2013.

Voorhees, H. and T. Poggio. Computing Texture Boundaries from Images, *Nature*, 333, 364–367, 1988.

Wehrhahn, C. and T. Poggio. Real-Time Delayed Tracking in Flies, *Nature*, 261, 43–44, 1976.

Wehrhahn, C., T. Poggio, and T. Bülthoff. Tracking and Chasing in Houseflies (Musca): An Analysis of 3D Flight Trajectories, *Biological Cybernetics*, 45, 123–130, 1982.

Wilems R.M. Re-appreciating the why of cognition: 35 years after Marr and Poggio, *Frontiers in Psychology*, 2, 1–5, 2011.

Yeo, G., E. Van Nostrand, D. Holste, T. Poggio, and C.B. Burge. Identification and Analysis of Alternative Splicing Events Conserved in Human and Mouse, *Proceedings of the National Academy of Sciences (PNAS)*, 102(8), 2850–2855, 2005.

Yu, A.J., M.A. Giese, and T. Poggio. Biophysiologically Plausible Implementations of the Maximum Operation, *Neural Computation*, 14(12), 2857–2881, 2002.

Yuille, A. and T. Poggio. Fingerprints Theorems for Zero Crossing, *J. Optical Society America A*, 2, 683–692, 1985.

Yuille, A.L. and T. Poggio. Scaling Theorems for Zero Crossings, *IEEE Trans. PAMI*, 8, 15–25, 1986.

Zhang Y., E. Meyers, N. Bichot, T. Serre, T. Poggio, and R. Desimone. Object Decoding with Attention in Inferior Temporal Cortex, *Proceedings of the National Academy of Sciences*, 108(21), 8850–8855, 2011.

Zoccolan, D., M. Kouh, A. Pasupathy, C. Connor, M. Riesenhuber, and T. Poggio. Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporarl Cortex, *Journal of Neuroscience*, 27(45), 12292–12307, 2007.