Ray Solomonoff and the New Probability

Grace Solomonoff*

Oxbridge Research
Cambridge, Mass. 02140
trovaxo@yahoo.com http://world.std.com/˜rjs

Abstract. This is the story of Ray Solomonoff's Life and Times. He was there at Dartmouth in 1956 when Artificial Intelligence was first given its name, and took part in the major events during this unique era right up to his death in 2009. His invention in 1960 of Algorithmic Probability, with its multiple descriptions of data, led to better ways of handling data and prediction for machine learning. The theorems that are part of his discovery lie at the heart of Algorithmic Information Theory. His discovery of the Universal Prior validated the use of Baye's Rule for prediction, which he developed in his prediction system called Solomonoff Induction. Ray championed probability in AI during the decades it was unpopular and lived to see a renaissance in systems that learn and reason using probability. The story of his life is the story of a great adventure.

Keywords: Solomonoff, Artificial Intelligence, Algorithmic Probability, Algorithmic Information Theory, Inductive Inference, Kolmogorov Complexity, Baye's Rule, Solomonoff Induction

1 Introduction

This is the story of Ray Solomonoff's life, beginning with the immigration of his parents from Russia to the U.S. in the early 1900's and Ray's birth in 1926. There was a cheerful dedication in Ray's life: from his childhood enthusiasm for math and science, to his contribution in 1956 to the birth of Artificial Intelligence, his invention of a new probability and his participation in major events during this unique era right up to his death in 2009. His greatest achievements were the invention of Algorithmic Probability and Solomonoff Induction. His vision of probability and machines that think will be part of our future.

2 Early Years

Ray's mother, Sarah Mashman, was born in Sevastopol, a port city in Ukraine, located on the Black Sea coast of the Crimea peninsula. She graduated from high school in 1911.

* Thanks to Donald Rosenthal and Julie Sussman for their invaluable help. A version of this report is in Springer, Some of this report appeared in (open sources) Algorithms, 3(3), 2010, and Oxbridge Research. Thanks to Oxbridge Research (open source), Marcus Hutter, Henry Tirri, Juergen Schmidhuber, Alex Solomonoff for permission to reprint pictures.

Before the Russian Revolution of 1917 there was a part of Russia called the Pale of Settlement: It included present-day Poland, Lithuania, the Crimea and the Ukraine. Jewish people were restricted to that area.

During around 1890-1910, the Czarist government was coming apart and blaming it on the Jews, and anti-Jewish persecution was particularly intense at this time. About 1/3 or 1/2 of all the Jews in Russia emigrated, mostly to the U.S. Ray's parents were part of this wave.

Sevastopol, a wealthy city, was exempt from the Ukrainian Pale. Only the fact that Sarah's father was the local blacksmith, with specialized knowledge, enabled his family to live there. Sarah immigrated to New York about three years after high school. She got a job as a nurse's aide, and on the side began an amateur career of acting.

Rays father, Phillip Julius Solomonoff, born in 1897, would never tell where he was from, but Vilna (in Lithuania) is on his passport. By 1914 he had come to New York and was first in his class at the Baron de Hirsch Trade School, training to be a plumber.

Inspired with revolutionary fervor in 1917, he and cousin Esoc got permission from the Russian Consulate General for a Special Government Mission and went to "visit revolution." They returned soon after discovering that the shooting was with real bullets. Julius joined a Polish ship, where he jumped ship and ended up back in New York.

Julius was an illegal alien all his life. He met Sarah, and they lived in New York, where in 1922 their first son George was born.

In 1924 they moved to Cleveland, Ohio, a Midwest city, called the largest small town in the U.S.

And Ray was born July 25, 1926.

1. Ray and his father. 2. Ray with his brother George. Ray is on the left, already with his penetrating gaze, 3. The obligatory picture of Ray on a horsie.

Ray grew up during the depression. His father was a mechanic but his work never paid enough to cover the rent. His parents had to move frequently, which was a trauma to them; but Ray and George thought it was great — they got to meet new kids, have new adventures.

When they finally settled down. Ray built a lab in his parents' cellar. To vent the smoke from his experiments, he drilled a hole through the wall behind some bushes. His parents never found it, and it remained unseen when the house was sold and was there until the day the place was torn down, around 1998. But his parents were really tolerant anyway. Sarah once described sweeping the rug and hearing a multitude of tiny explosions from some grains of something scattered about from a experiment gone awry.

In later life Ray made quite a few inventions, for friends or home use. For example he made a Hurry clock for Marvin Minsky — a clock that was labeled "Hurry"; he removed a gear so it ran very fast — I guess you'd know you were always late.

But his greatest pleasure was in theories and discovery

He wrote "I first experienced the pure joy of mathematical discovery when I was very young — learning algebra. I didn't really see why one of the axioms was 'obvious', so I tried reversing it and exploring the resultant system. For a few hours I was in this wonderful never-never land that I myself had created! The joy of exploring it only lasted until it became clear that my new axiom wouldn't work, but the motivation of the joy of discovery continued for the rest of my life."

"The motivation for the discovery of Algorithmic Probability was somewhat different. It was the result of 'motivated discovery' — like the discovery of the double helix in biology, but with fewer people involved and relatively little political skullduggery. The goal I set grew out of my early interest in science and mathematics. I found that while the discoveries of the past were interesting to me, I was even more interested in how people discovered things. Was there a general technique to solve all mathematical problems? Was there a general method by which scientists could discover all scientific truths?"[Sol97]

And so by 1942, around the age of 16, Ray became captivated by the idea of finding a general technique to solve all problems. He felt already that scientists used probability in induction when they invented theories to account for data. This was when he first began to study induction. Even at this age, probability was part of his thinking.

He organized notebooks using key letters: TM meant Thinking Machine. His later notes all use key letters such as ALP for Algorithmic Probability, TS for Training Sequences, even HR for horse racing.

He went to Glenville High School, noted for its famous graduates. After graduation there was a hiatus in his studies, for World War II intervened. The draft began in 1940, and in November 1944 Ray joined the Navy to train in and then teach radio and was stationed in Gulfport, Mississippi.

After the war, in 1946 he went to the University of Chicago on the GI bill. The University was at its height: Enrico Fermi, Rudolf Carnap, Nicolas Rashevsky, Anatole Rapoport were among Ray's teachers.

3 From the University to the Birth of AI

In 1950 in a letter Ray writes "for the last 4 or 5 years cybernetics has been my chief scientific interest" ... "about a month and a half ago I worked out a method of devising a machine that would 'think'." But "the bubble broke. One found that the machine wouldn't work as well as expected" ... "But it was all a very wonderful adventure — somewhat disturbing at times, but nonetheless wonderful." At that time he also tried to get together a group of scientists interested in thinking machines, but did not succeed. By then Ray was convinced that thinking machines were feasible (the name Artificial Intelligence did not yet exist).

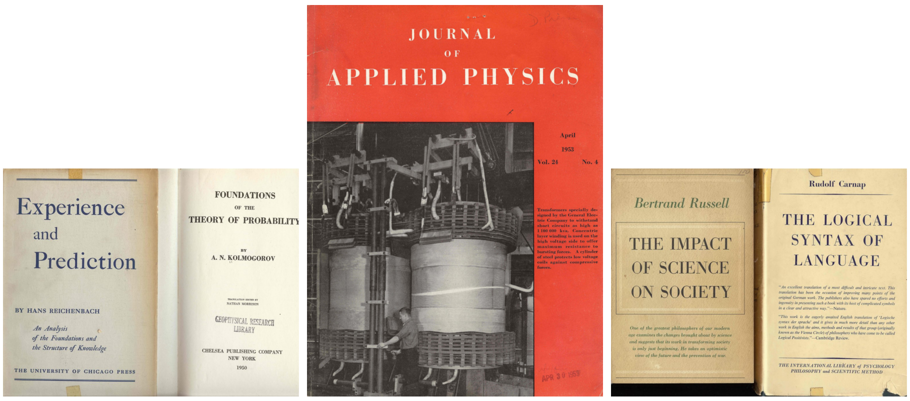
Claude Shannon's paper in 1948, and subsequent developments in information theory over the next few years, greatly influenced Ray during college. Shannon thought that information was something that could be quantified and that the quantity of information was related to its probability.

Rudolf Carnap was Ray's professor of Philosophy and Probability. Carnap was a "logical positivist." He felt that logical syntax should provide a system of concepts, a language, by the help of which the results of logical analysis would be exactly formulable. Problems, such as what language to use, presented too many difficulties to be successful, but the idea of combining language, information and probability was part of what led to Ray's discovery.

Ray studied mathematical biology with Rapoport and Rashevsky. His first published reports were three papers on neural nets, two with Anatole Rapoport in 1950-52, that are regarded as the earliest statistical analysis of networks.

Here are some of the items Ray was reading during college:

"Foundations of the Theory of Probability" (Kolmogorov); "Experience and Prediction" (Reichenbach); "Journal of Applied Physics"; "The Impact of Science on Society" (Russell)...



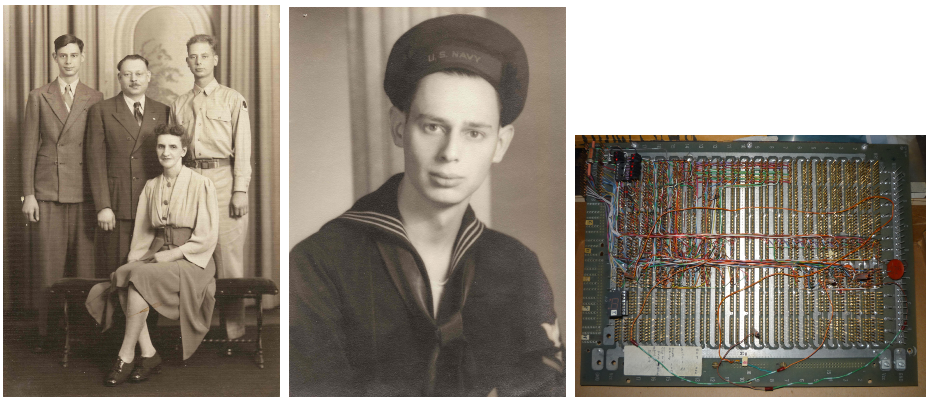
Here are some more of the items Ray was reading during college:

Actually he was quite influenced by science fiction. His favorite movie was 'The Shape of Things to Come', that tells of a visionary people who would save a dying world through technology.

In the 40's and 50's Science Fiction also represented something that was happening in the U.S., and in Europe also, a view that amazing things were possible through technology. The government began sponsoring projects that used information technology.

Meanwhile, the computer was rapidly becoming important. In 1941 the electronic computer had been introduced to the public. Ray was fascinated by them. Later, in the early 1970's Ray even built a computer, hand-wired it — sort of tapestry-like.

These were some of the early inventions and ideas that influenced Ray.



1. A family portrait when Ray was in high school (Ray on the left). 2. Ray in the Navy. 3. The computer that Ray made.

4 The Beginnings of AI

After college, Ray got a job in New York, working at Technical Research Group. He met Marvin Minsky at a conference. Minsky and Ray became lifelong close friends, and Ray often visited Cambridge, Massachusetts, where Minsky lived.

This was when Minsky and John McCarthy introduced him to another great idea: The Turing machine. It was a revelation.

The Turing machine was a concept of Alan Turing's, a British mathematician who lived 1912-1954. In 1936 Turing described a "universal computing machine" that could theoretically be programmed to solve any problem capable of solution by a specially designed machine. The "Turing machine" foreshadowed the digital computer.

Consider a computer as a finite state machine operating on a finite symbol set. A program tape on which a binary program is written feeds from left to right into the machine. The machine looks at the tape, changes its state according to the program, writes some output, and then gets the next part of the tape. A special Turing machine, the universal Turing machine, has an infinite memory, an input tape that goes only one way, and a work tape — this machine could mimic any other computer.

Ray wrote "It gave me a quick intuitive grasp of many ideas that I had before found incomprehensible. It is not unusual for the translation of a problem into a new language to have this wonderful effect."[Sol97]

In Cambridge, Ray had frequent discussions with Marvin Minsky and John McCarthy and others about thinking, mathematics and machines. Then in 1956 John McCarthy organized a little group to brainstorm on machine intelligence. They picked Dartmouth, New Hampshire for their venue. The principal participants would include John, Ray, Marvin, Tenchard More, Alan Newell, Herb Simon, and 4 other researchers.

McCarthy decided to give this new science a name: And so in the summer of 1956, the science named Artificial Intelligence was born.

The idea was that the 10 participants would brainstorm during 2 months at Dartmouth, and then in 10 months would come up together with some Great AI Idea. Oh sure! You have 10 mad scientists with 10 totally different orientations toward AI. What do you get? (10 totally different Great AI Ideas.)

From most of these ideas came logic-based, deterministic programs. Ray presented a paper on how machines could be made to improve themselves, by using unsupervised learning from examples. This paper was not probabilistic, but was new in that it replaced semantic with symbolic representation.

In 2010 Marvin wrote me: "In the early 1950's I developed many theories about ways to make neural networks learn — but none of these seemed likely to lead to high-level kinds of idea. But once Ray showed me his early ideas about symbolic 'inductive inference scheme', these completely replaced my older approach. So Ray's ideas provided the start of many, later, steps toward Artificial Intelligence."

The first question in Ray's outline for work in Thinking Machines at the Dartmouth Summer was "What is Probability?" Ray advocated a probabilistic approach to machine intelligence at this first meeting on AI in 1956, and continued to push on this dream for decades when such a view was controversial.

The Dartmouth Summer had difficulties; the Ford Foundation gave less funding than McCarthy hoped for, and some of the attendees spent little time there.

Of the programs from that summer, the most well known is Newell and Simon's Logical Theorist, and soon after that, their General Problem Solver. Their focus, as was that of McCarthy, was logic based and highly specialized. This led to "expert system" programs — for example, aids to medical diagnosis, or minerals prospecting. Really at this point there came a schism between the specialized, deterministic and the more generalized kinds of programs. The more rigidly logical projects were immediately applicable, and so AI became known by these programs. Probabilistic based programs were largely ignored. They were not well integrated into the world of computing, while logic based programs were ideal for "if-then" and "do loops".

5 The Discovery of Algorithmic Probability

Calvin Mooers was an attendee at the MIT Information Theory Symposium in September 1956. Ray's paper from the Dartmouth Summer Project, "An Inductive Inference Machine," was circulated there. Calvin liked Ray's ideas and invited him to come work with him in information retrieval. He got the perfect government contract for Ray to work with him. In creating his proposal he wrote Ray: "There seems to be developing a good possibility that would permit me to put you to work on doing long-range thinking on 'inductive inference machines' and whatever else they may touch upon. This would be a program of purely speculative thinking, and I would set up the support so *that* is understood. While the money would come from an interest in information retrieval, I would want the thinking untrammeled, and my thesis is that what you are doing, and what I am doing in information retrieval, have a common meeting ground that will develop in due time." And he got the grant! Don't you wish you got a grant like that!

Ray received government funding during the next few years. He never would again.

So Ray moved up to Cambridge, to work full time on his ideas about induction and machine intelligence.

Here he is at Marvin's house — and here — his thinking is pretty untrammeled...

Over the next few years Ray worked on machine intelligence, programming computers, use of the Turing machine, and coding methods. He still didn't have a really good way of combining them

Ray had met Noam Chomsky in the mid 50's and read Chomsky's works on context-free language. Ray realized this language could be a basis for solving Carnap's problem. Working at Calvin's Zator Company, Ray expanded this language into a stochastic form for all types of patterns: it was a probabilistic language. Normally one thinks of a language where either something is a sentence or is not a sentence. In a probabilistic language something has a probability of being a sentence. This provided the breakthrough he needed.

Using the Universal Turing Machine as the operator, Ray established that the generator of these patterns could be put in a binary, probabilistic form; the grammar of this language was somewhat like descriptions where simple descriptions were more likely than complex ones. He called this Algorithmic Probability, and he used it in a General Theory of Inductive Inference. In this discovery, in 1960, Ray became a founder of Algorithmic Information Theory.[Vit09]

Prior to this discovery, the usual method of calculating probability was based on frequency: take the ratio of favorable results to the total number of trials. Ray seriously revised this definition of probability. Algorithmic Probability is based on the length of random programs (algorithms) input into a universal Turing machine that produce a given sequence of symbols as output — the shorter programs being most likely.

His paper, "A Preliminary Report on a General Theory of Inductive Inference" is his first known publication of Algorithmic Probability and was published by Calvin's company in February 1960,[Sol60a] with a revision in November 1960.[Sol60b] He published a more complete exposition in two 1964 papers for the Journal of Information and Control.[Sol64a],[Sol64b]

In a letter in 2011, Marcus Hutter wrote: "Ray Solomonoff's universal probability distribution M(x) is defined as the probability that the output of a universal monotone Turing machine U starts with string x when provided with fair coin flips on the input tape. Despite this simple definition, it has truly remarkable properties, and constitutes a universal solution of the induction problem."(See also [RH11])

Algorithmic Probability combines several major ideas; of these, two might be considered more philosophical and two more mathematical.

The first is related to the idea of Occam's Razor: the simplest theory is the best. Ray's 1960 paper states "We shall consider a sequence of symbols to be 'simple' and have high a priori probability if there exists a very brief description of this sequence — using of course some stipulated description method. More exactly, if we use only the symbols 0 or 1 to express our description, we will assign the probability of 2^{-N} to a sequence of symbols, if its shortest possible binary description contains N digits."[Sol60b][Sol60a]

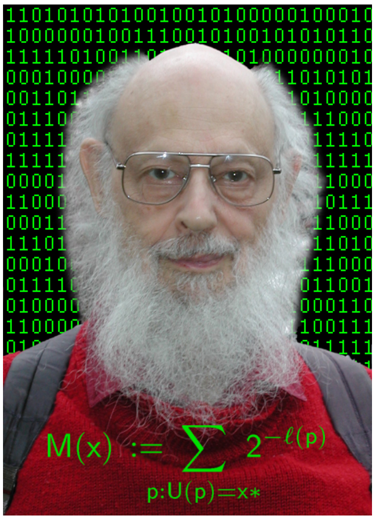
The second idea is similar to that of Epicurus: it is an expansion on the shortest code theory; if more than one theory explains the data, keep all of the theories. Ray writes "Equation 1 uses only the 'minimal binary description' of the sequence it analyzes. It would seem that if there are several different methods of describing a sequence, each of these methods should be given *some* weight in determining the probability of that sequence."[Sol60b][Sol60a]

$$\mathrm{P(x)}_M = \sum_{i=1}^{\infty} 2^{-|s_i(x)|}$$

This is the formula he developed to give each possible explanation the right weight. (The probability of sequence x with respect to Turing Machine M is the total sum of 2 to the minus length of each string s that produces an output that begins with x.)

Closely related is the third idea of its use in a Bayesian framework. The universal prior is taken over the class of all computable measures; no hypothesis will have a zero probability. Using program lengths of all programs that could produce a particular start of the string, x, Ray gets the prior distribution for x, used in Bayes rule for accurate probabilities to predict what is most likely to come next as the start is extrapolated. The Universal Probability Distribution functions by its sum to define the probability of a sequence, and by using the weight of individual programs to give a figure of merit to each program that could produce the sequence. [Sol60b][Sol64a]

The fourth idea shows that the choice of machine, while it could add a constant factor, would not change the probability ratios very much. These probabilities are machine independent; this is the invariance theorem that is considered a foundation of Algorithmic Information Theory.[Sol60b][Sol64b]



Here is another beautiful picture of this formula in a little different format: The photoshopped picture is by Marcus Hutter, using a photo by Jürgen Schmidhuber.

In 1965 the great Russian mathematician Andrey Kolmogorov published the first idea (simplicity) and fourth idea (machine independence) in the journal Problems of Information Transmission.[Kol65] Kolmogorov had also been working with Turing machines, and had been giving lectures at Moscow University on the subject of complexity. Like Ray, he revised the frequentist view of probability. He wrote "The basic discovery, which I have accomplished independently from and simultaneously with R. Solomonoff, lies in the fact that the theory of algorithms enables us to eliminate this arbitrariness by the determination of a 'complexity' which is almost invariant (the replacement of one method by another leads only to the supplement of the bounded term)"[Vit09]

Paul Vitányi writes that Kolmogorov's introduction of complexity "was motivated by information theory and problems of randomness. Solomonoff introduced algorithmic complexity independently and earlier and for a different reason: inductive reasoning. Universal a priori probability, in the sense of a single prior probability that can be substituted for each actual prior probability in Bayes's rule was invented by Solomonoff with Kolmogorov complexity as a side product, several years before anybody else did."[Vit09]

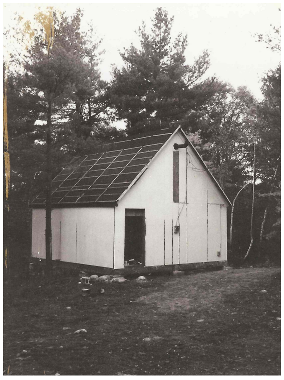
Vitányi notes: "we will associate Solomonoff's name with the universal distribution, and Kolmogorov's name with the descriptional complexity."[LV08] Kolmogorov was interested in the information content of a string, while Ray was interested in the predictive power of a string.

With respect to thinking machines: If a machine is working on a set of problems, the algorithm it uses may only work for some problems. If you can use probability in the best possible way to estimate how likely each one in a whole set of algorithms is, then if one program (algorithm) doesn't work, you have a good method, a probabilistic way, to search for another.

Ray's ideas were ahead of his time, especially here in America. He gave some lectures at local universities. Often students didn't know what he was talking about (he wasn't too good at explaining this stuff anyway!). Students would fall asleep or quietly escape out the side door.

Never mind; meanwhile he bought some land in New Hampshire and built a house. Since he didn't know anything about building a house, he got a book. He built the house as a cube, since that was easiest. He had real sloping roof, to let the snow off.

He didn't know about heating, but he knew about electric light bulbs, and in those days a light bulb was 80% heat and only 20% light, so he heated the house with light bulbs. His friend Al Jenks told me recently that Al would bring his friends over, ostensibly to meet Ray, but actually to show them that Ray really did heat his house entirely with light bulbs — two long rows along the ceiling from wall to wall.

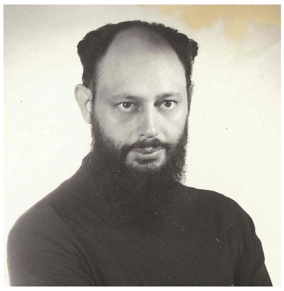


Pictures: 1. the house. 2: house decor. 3. A later picture by Alex Solomonoff of Ray working on the house.

It had no windows, but you could pull out long sections of the wall from floor to roof and put in screens during the summer.

And it was shortly after he finished the house that he met...ME.





Pictures: 1. My baby picture. 2. A few years after meeting Ray. 3. Ray, a few years before meeting me.

In 1969, I met Ray at a party at a friend's house. We often went to the house Ray built in New Hampshire.

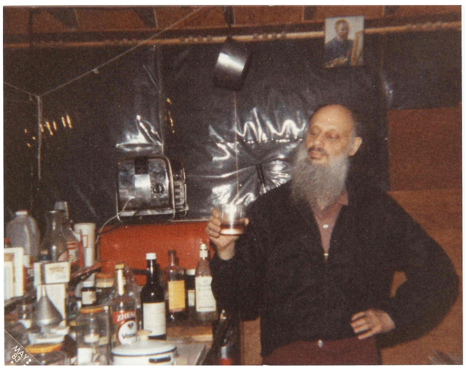
I wrote poetry and saw that good poems often have good metaphors. A metaphor is an alternate description of something; it helps you see in a different way. The goal is to make the subject of the poem meaningful or beautiful. When Ray told me what he was doing, just like those students, I didn't know what he was talking about, but his multiple theories and multiple descriptions seemed like a mathematical analogue to poetry. It felt like we were on some kind of similar adventure. But even if our goals had been different, we shared interests and feelings — our life has been this way forever.

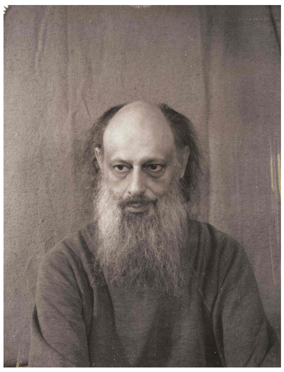
We had an apartment in Harvard Square, Cambridge, which we kept through the 90's. Harvard Square was where you met everybody as you wandered around with your backpack. We were there during the Vietnam War protests. From the building roof we watched the Weathermen break bank windows. We were there during the days of the flower children. Ray was a friend of the poet Allen Ginsberg. Whenever we were at Ginsberg's poetry readings he would hand Ray his watch to monitor the time.

Our apartment had a lot of stuff: electronics and computer "toys" and building materials; but mainly our apartment had books — stacks and stacks of papers and books. All the walls had bookshelves, even the ceiling had bookshelves hanging from it across the hallway.

Ray read voraciously about all kinds of things, but was focused on his prediction theory, talking to others, often at the AI Lab at MIT. It was a very exciting time.

In 1970 since the military had stopped funding civilian science, Ray formed his own one-man company, "Oxbridge Research," an extremely eclectic organization.





On the right, a portrait of Ray in 1984. On the left, a portrait of Ray in Oxbridge Research's Chemical Experimentation Lab, making our most important invention: the liquid helium zonk. The zonks are a whole class of drinks. One example: take a banana, two kinds of rum, lemon and honey and blend them in a blender, then freeze into a subzero mush. Other inventions were the pickled herring cocktail and studies of the brightness of various drinks with mixes that fluoresced in ultraviolet light.

In 1975 Chaitin, who had developed a version of descriptional (Kolmogorov) complexity later in the 1960's, wrote about complexity in the Scientific American. In 1978 Leonid Levin and Peter Gács came to America and were very supportive and interested in Ray's probability theory.

6 The Guerrilla Workshop

However — the dominant researchers used deductive, logic-based methods; deterministic expert systems were the popular products of AI. Many scientists felt uncertainty just wouldn't work in this field. The turning point, at least in the U.S.— was in 1985.

The first workshop on Uncertainty in AI (UAI) was in 1985. How did it happen? The program note was sort of like the Declaration of Independence.

It says: "This workshop came about as a result of a panel discussion at the AAAI conference at Austin, Texas, U.S., in 1984. This panel was on the problem of the representation of uncertainty in Artificial intelligence." ... "Some speakers implied that more than one number was necessary to represent uncertainty, while others stated that numbers should not be used at all! Except for a valiant rearguard defense by Judea Pearl, everyone on the panel agreed that probability as a representation of uncertainty either was misguided or inadequate for the task. Several of us who have been using probability within AI, as well as engineers and physicists, know this conclusion to be false, and our outrage at this denigrating of probability was the spur that triggered this workshop."

This workshop has continued ever since, and was a starting point that led to a revolution in mainstream AI. It is now the annual Conference in Uncertainty in AI (UAI), hosted by AAAI. For this workshop Ray gave a paper on how to apply the universal distribution to problems in AI: the best order of searching for solutions is $T1/P1$, where $T1$ is the time needed to test the trial and $P1$ the probability of success of that trial — the shorter codes getting the higher probability. This is based on the search technique invented by Leonid Levin,[Lev73] so Ray called it Lsearch.

I remember that first meeting: Ray called it a guerrilla workshop. But it marked an acceptance of many aspects of probability by mainstream AI. There still is conflict in AI circles over the use of probability. But the logic and probabilistic reasoning are moving closer. For example, I believe John McCarthy shifted his focus from deductive logic in his development of what he calls 'circumscription', which is more probabilistic. Things like fuzzy logic and Bayesian work lie in the probabilistic area.

Ray inspired many people, especially young people. Our nephew Alex would visit from Cleveland. He says "There was nothing like this for me in Cleveland. Ray talked about mathematics in a way that made it exciting." Ray influenced him to get his Ph.D. in mathematics, and we have remained close to Alex ever since.


Pictures: 1. Alex's Ph.D. graduation, May 1992. 2. Ray, me, our niece Nickie, and Alex in 2007.

7 Later work

Ray spent the rest of his life discovering, proving aspects, refining and enlarging his General Theory of Inductive Inference, with the goal of having machines that could solve hard problems. He stressed that machine intelligence did not need to emulate human intelligence, and probably that it would not.

He wrote about the problem of incomputability. His quotable quote: "Incomputability — it's not a bug, it's a feature!" Systems that are computable will not be complete. The incomputability is because some algorithms can never be evaluated because it would take too long. But these programs will at least be recognized as possible solutions. On the other hand, any computable system is incomplete. There will always be descriptions outside that system's search space which will never be acknowledged or considered, even in an infinite amount of time. Computable predictions models hide this fact by ignoring such algorithms. Minimum Description Length, for example, first cuts out some space and then chooses the shortest program. Minimum Message Length of Chris Wallace does acknowledge incomputability; it is closer to Algorithmic Probability.[WB68][WD99] It chooses only the shortest code to work with, however, while Algorithmic Probability uses as many as there is time for.

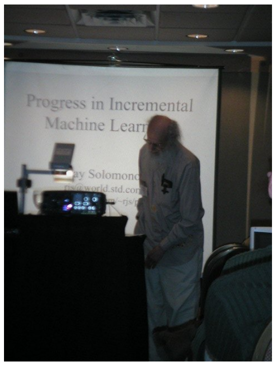
In other papers Ray wrote about how best to limit search by limits on time or computation cost. He developed methods for working with other types of data, not just sequences. He categorized problems into various types such as "inversion problems" and "time-limited optimization problems" and developed ways of dealing with the different types.

Throughout his career Ray was concerned with the potential benefits and dangers of AI, discussing it in many of his published reports. In 1985 he analyzed a likely evolution of AI, giving a formula predicting when it would reach the "Infinity Point." This Infinity Point is an early version of the "Singularity" later made popular by Ray Kurzweil.[Sol85]

During most of his life, Ray worked independently, developing his theories without any academic or industrial support. Paul Vitányi notes "it is unusual to find a productive major scientist that is not regularly employed at all."

However, he would meet frequently with Minsky, Shannon, and others at MIT, and with other researchers throughout the world. We went to many countries and conferences meeting amazing and dedicated people. There was Saarland University in Saarbrücken, Germany, where Wolfgang Paul invited Ray to do research in 1990-1991. There was the ISIS conference in Australia in 1996; after it the indefatigable David Dowe drove us for miles along the Great Ocean Road and we explored the ancient and beautiful rain forest there. Later we stayed with Paul Vitányi, who traveled with us by bus and car, and then we visited other areas. When Ray was a little boy, he thought about being a naturalist; Dr. Doolittle was one of his favorite books. How could he not be entranced by seeing Cassowaries? There was his visiting professorship at the Dalle Molle Institute for Artificial Intelligence in Lugano, Switzerland, run by Jürgen Schmidhuber. The researchers at that Institute were so cohesive. We had community meals — memorable spaghetti dinners with coffee from their super espresso machine. There was our visit, in 1998, to the dynamic Computer Research Learning Center, at Royal Holloway, University of London, which Alex Gammerman had just founded. Later Ray gave the first Kolmogorov lecture there, receiving the Kolmogorov Award. Ray was a visiting professor there until his death.

Here are some pictures from over the years:



Pictures: 1. Ming Li, his wife Jessie Zou, Ray and me, maybe in the 1990's. 2. and 3. Pictures by Jürgen Schmidhuber and Henry Tirri. at NIPS Workshop on Universal Learning Algorithms and Optimal Search, 2002

Ray was happy when the AGI08 conference occurred; this was first of the AGI conferences focused on Artificial General Intelligence, moving as far away as possible from narrowly focused and highly specialized programs.

Eric Horvitz, for many years the President of AAAI, notes Ray "advocated the probabilistic approach to machine intelligence at the first meeting on AI in 1956, continued to push on this dream for decades when such a view was controversial, and lived to see a renaissance in systems that learn and reason under uncertainty, relying on representations of probability — a perspective that is now at the foundation of modern AI research."

But of all his productive life, his greatest invention is Algorithmic Probability and his General Theory of Inductive Inference.

Leonid Levin wrote that Ray "had a very powerful and general approach. In the future, his ideas will have more influence."

And his whole life Ray did what he loved doing. Vitányi notes "But from all the elder people (not only scientists) I know, Ray Solomonoff was the happiest, the most inquisitive, and the most satisfied. He continued publishing papers right up to his death at 83."[GV11]

Ray enjoyed life up to the end, organizing a gorgeous costume for Halloween at the end of October; and also continued with his serious side, his work on Algorithmic Probability and prediction, completing a paper for AGI 10, in late November 2009, just before he died.

In his last paper, for AGI10, Ray discussed what he called "The Guiding Probability Distribution" — a nice updating method, and a name that seems to me like something that was guiding him too. So he had a long happy life, which is something to celebrate. And even more than that, he had a shining vision he followed for his whole life, and it guided him right to the very end.

References

[Car50] R. Carnap. *Logical Foundations of Probability.* 1950.

[Cho56] A.N. Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, IT-2(3):113–124, Sept 1956.

[GV11] P. Gács and P. Vitányi. Raymond J. Solomonoff 1926 – 2009. *IEEE Information Theory Society Newsletter*, 61(11), March 2011.

[Kol65] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1(1):1–7, 1965.

[Lev73] L.A. Levin. Universal search problems. *Problemy Peredaci Informacii*, (9):115–116, 1973. Translated in Problems of Information Transmission 9, pp. 265–266.

[LV08] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, N.Y., third edition, 2008.

[RH11] S. Rathmanner and M. Hutter. A philosophical treatise of universal induction. *Entropy*, (13):1076–1136, 2011.

[Sha48] C.E. Shannon. The mathematical theory of communication. *Bell System Technical Journal*, (27):379–423, 623–656, 1948.

[Sol56] R.J. Solomonoff. An inductive inference machine. Dartmouth Summer Research Project on Artificial Intelligence, August 1956. A privately circulated report.

[Sol60a] R.J. Solomonoff. A preliminary report on a general theory of inductive inference. Technical Report V-131, Zator Co. and Air Force Office of Scientific Research, Cambridge, Mass., Feb 1960.

[Sol60b] R.J. Solomonoff. A preliminary report on a general theory of inductive inference. (revision of Report V-131). Technical Report ZTB-138, Zator Co. and Air Force Office of Scientific Research, Cambridge, Mass., Nov 1960.

[Sol64a] R.J. Solomonoff. A formal theory of inductive inference: Part I. *Information and Control*, 7(1):1–22, March 1964.

[Sol64b] R.J. Solomonoff. A formal theory of inductive inference: Part II. *Information and Control*, 7(2):224–254, June 1964.

[Sol75] R.J. Solomonoff. Inductive inference theory - a unified approach to problems in pattern recognition and artificial intelligence. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, volume 1, pages 274–280, Tbilisi, Georgia, U.S.S.R., Sept 1975.

[Sol78] R.J. Solomonoff. Complexity-based induction systems: Comparisons and convergence theorems. *IEEE Transactions on Information Theory*, IT–24(4):422–432, July 1978.

[Sol85] R.J. Solomonoff. The time scale of artificial intelligence; reflections on social effects. *Human Systems Management*, 5:149–153, 1985.

[Sol86] R.J. Solomonoff. The application of algorithmic probability to problems in artificial intelligence. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 473–491. Elsevier Science Publishers, B.V., 1986. Also in: M.Kochen and H.M. Hastings, Eds., Advances in Cognitive Science, AAAS Selected Symposia Series, AAAS, Washington, D.C., pp. 210-227, 1988.

[Sol97] R.J. Solomonoff. The discovery of algorithmic probability. *Journal of Computer and System Sciences*, 55(1):73–88, August 1997.

[Sol08] R.J. Solomonoff. Three kinds of probabilistic induction: Universal distributions and convergence theorems. *The Computer Journal*, 51(5):566–570, 2008. Christopher Stewart Wallace (1933-2004) memorial special issue.

[Tur37] A. M. Turing. On computable numbers, with an application to the entsheidungsproblem. *Proc. London Math. Soc.*, 42:230–265, March 1937.

[Vit09] P. Vitányi. Obituary: Ray Solomonoff, founding father of algorithmic information theory. CWI, Amsterdam, December 2009.

[WB68] C.S Wallace and D.M. Boulton. An information measure for classification. *The Computer Journal*, 11:185–194, 1968.

[WD99] C.S. Wallace and D. L. Dowe. Minimum message length and Kolmogorov complexity. *Computer Journal*, 42(4):270–283, 1999. special issue on Kolmogorov complexity.

A shorter version of this was published: D.L. Dowe (Ed.): Solomonoff Festschrift, LNAI 7070, pp. 3752, 2013.