Ray Solomonoff and The New Probability

Grace Solomonoff
Oxbridge Research, Cambridge, Mass. 02140, U.S.A.
prettyvivo@gmail.com   http://raysolomonoff.com *

July 25, 2026

Abstract

This is the story of Ray Solomonoff's life and his discovery of probabilistic induction and prediction. 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. 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 and invention is the story of a great adventure.

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 1900s 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 understanding of probability and his participation in major events during this unique era right up to his death in 2009. His greatest achievement was the invention of Algorithmic Probability in his General Theory of Induction. His vision of probability and machines that think will be part of our future.

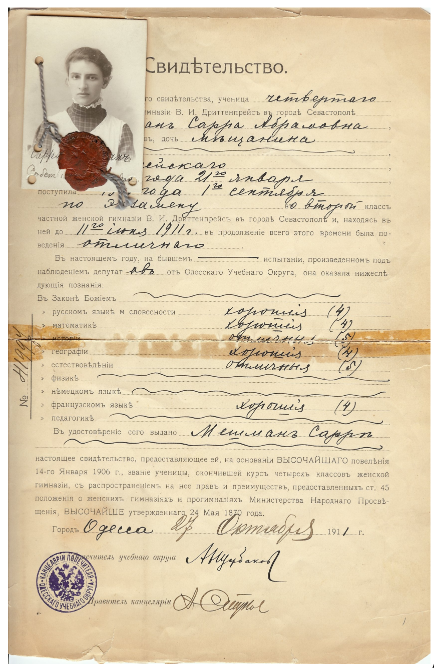
*Thanks to Donald Rosenthal and Julie Sussman for their invaluable help. Some of this report appeared in (open sources) Algorithms, 3(3), 2010, and Oxbridge Research. Thanks to Oxbridge Research (open source), Marcus Hutter, Henry Tirri, Juergen Schmidhuber, Alex Solomonoff for permission to reprint pictures. * This is enlarged, corrected version of the report in D.L.Dowe (Ed.) Solomonoff Festschrift, LNAI 2013, Springer-Verlag Berlin, Heidelberg, 2013.

2 Early Years

Ray's mother, Sarah Mashman, was born in Sevastopol, a port city located on the Black Sea coast of the Crimean peninsula.

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

Sevastopol, a wealthy city and main base for the Russian Imperial Black Sea Fleet 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 was quite unique, for she was allowed to attend the Catholic high school at a time when there was a strict quota, the numerus clausus, limiting the number of Jewish students in education. She graduated with honors in 1911.



Свидѣтельство.

...го свидѣтельства, ученица четвертаго
...гимназіи В. И. Дриттенпрейсъ въ городѣ Севастополѣ
...анъ Сарра Абрамовна,
...въ, дочь мѣщанина
...ейскаго
...года 21го января
поступила ... года 1го сентября
по экзамену во второй классъ
частной женской гимназіи В. И. Дриттенпрейсъ въ городѣ Севастополѣ и, находясь въ ней до 11го іюня 1911 г. въ продолженіе всего этого времени была поведенія отличнаго.

Въ настоящемъ году, на бывшемъ ——— испытаніи, произведенномъ подъ наблюденіемъ депутатовъ отъ Одесскаго Учебнаго Округа, она оказала нижеслѣдующія познанія:

Въ Законѣ Божіемъ	———
» русскомъ языкѣ и словесности	хорошія (4)
» математикѣ	хорошія (4)
» исторіи	отличныя (5)
» географіи	хорошія (4)
» естествовѣдѣніи	отличныя (5)
» физикѣ	———
» нѣмецкомъ языкѣ	———
» французскомъ языкѣ	хорошія (4)
» педагогикѣ	———

№ 11994

Въ удостовѣреніе сего выдано Мешманъ Сарръ настоящее свидѣтельство, предоставляющее ей, на основаніи ВЫСОЧАЙШАГО повелѣнія 14-го Января 1906 г., званіе ученицы, окончившей курсъ четырехъ классовъ женской гимназіи, съ распространеніемъ на нее правъ и преимуществъ, предоставленныхъ ст. 45 положенія о женскихъ гимназіяхъ и прогимназіяхъ Министерства Народнаго Просвѣщенія, ВЫСОЧАЙШЕ утвержденнаго 24 Мая 1870 года.

Городъ Одесса 27 Октября 1911 г.

Попечитель учебнаго округа А. Щукаковъ

Правитель канцеляріи

Канцеляріи Попечителя Одесскаго Учебнаго Округа

1. Sarah Mashman in High School. 2. Sarah's Diploma.

During around 1890-1910, the Czarist government was coming apart and blaming it on the Jews, and anti-Jewish persecution became particularly intense. 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.

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.

Ray's father, Phillip Julius Solomonoff, would never tell where he was from, but Vilna (in Lithuania) is on his passport. He got inspired with revolutionary fervor and in 1917 when the Russian revolution started, he and cousin Esoc went to “visit revolution.” They returned soon after discovering that the shooting was with real bullets.



1. later photo of Julius. 2. Julius and Sarah 3. Sarah in New York

Julius joined a Polish ship, and made his way to the U.S., where he jumped ship and ended up in New York. Julius was an illegal alien all his life. But he studied at the Baron de Hirsch Trade School in New York City to be a plumber and was first in his class. 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.

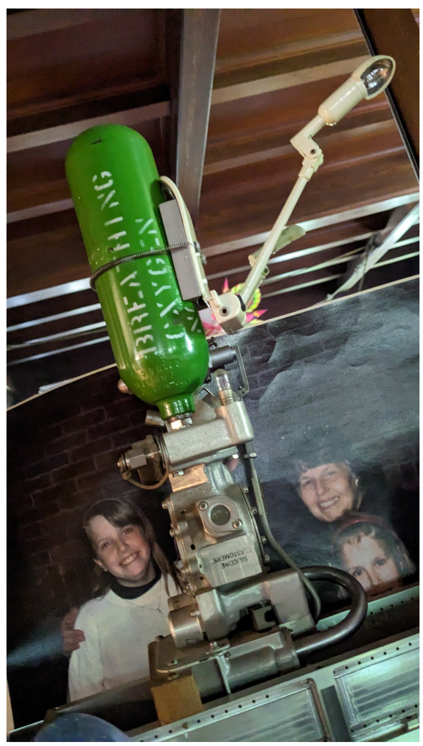


1. Ray and his father. 2. Ray with his brother George. Ray is on the left, already with his penetrating gaze. 3. Ray on a horsie.

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 close to the ground, 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 built a tic-tac-toe machine out of doorbells, a clock labeled "HURRY" — he removed a gear so it ran extra fast, the worlds biggest book-reading light, etc. etc.



The book-light at its Minsky home

But his greatest pleasure was in theories and discovery.
He wrote:

> —a few words about motivation. Motivation in science is roughly of two kinds: In one, the motivation is discovery itself — the joy of

> "going where no one has gone before" — the excitement of creating new universes and exploring them.
>
> Another kind of motivation is the achievement of a previously defined larger goal, and there may be many subsidiary discoveries on the path to this goal. In my own case both kinds of motivation were very strong. 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 "goal 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? [17]

And so by 1942, around the age of 16, Ray was completely 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: 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."

By then Ray was convinced that thinking machines were feasible. In those days, the most common term for his study was cybernetics, due to the mathematician Norbert Wiener's famous book *Cybernetics* tying aspects of human thinking to machine operations [21]. The name "Artificial Intelligence" would not come into vogue until a special event in 1956.

Claude Shannon's paper in 1948, and subsequent developments in information theory over the next few years, influenced Ray during, as well as after, college. The Shannon version of information is that it can be quantified. It is part of a set of possible messages and the quantity of information is related to its probability of being sent. The less likely the message, the more information it carries. It's an engineer's view — the transmission of information which is the basis of how we communicate.

Later, Ray wrote:

> Most important was the idea that information was something that could be quantified, and that the quantity of information was closely related to its probability. It suggested to me what I called at that time "The Information Packing Problem"— How much data could one pack into a fixed number of bits, or conversely how could one store a certain body of data using the least number of bits? The idea was that the amount of data one could pack into a certain number of bits was related to the redundancy or information content of the data. Since information content was related to probability, inverting a solution to the information packing problem could give one probabilities. Unfortunately, it was always possible to pack an arbitrary data string into 1 bit, using a suitable definition — a clearly inappropriate solution. [17]

When Ray did discover Algorithmic Probability, it would solve this problem.

Ray also had papers by Norbert Wiener. Wiener's version of information looked at the particular message itself, how it changes over time and causes changes. He wrote that information is "the content of what is exchanged with the outer world as we adjust to it, and make our adjustment felt upon it" [22, p. 18].

Ray's discovery would also consider the content of a message, and would discover a way to quantify that.

He was also influenced by his own teachers at the University of Chicago (UChicago).

The University was at a peak of brilliance then. Ray kept the notes from Enrico Fermi's students on Nuclear power (The compendium was titled Unclear Notes). Ray's teacher, Nicholas Rashevsky, founded the field of Mathematical Biology, 1936. Anatole Rapoport, in 1948, wrote about connectivity in nets.

Ray's first published reports were three papers on neural nets, two with Anatole Rapoport in 1950-52 and his own in 1953. These are regarded as the earliest statistical analysis of networks.

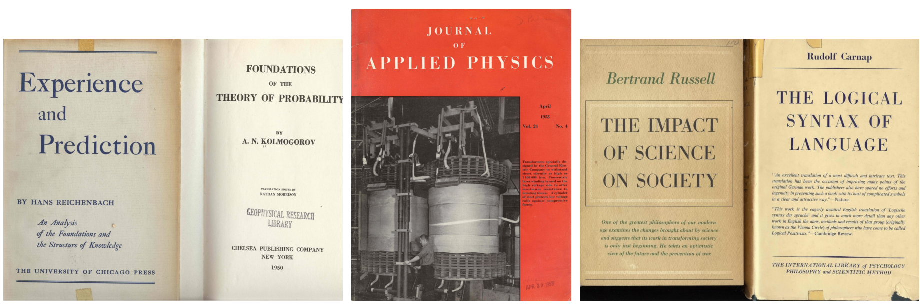
But most important for Ray was Rudolf Carnap, Ray's professor of Philosophy and Probability.

Carnap was a "logical positivist." (Meaning only comes via true observation or mathematically correct logic). By 1950, Carnap had expanded beyond purely logical syntax; he felt that the formal logical structure of a language could provide a precise system of concepts, through which the results of logical analysis would be exactly formulable [1].

Later he tried to figure out how to describe everything in the entire universe, including itself, by a long digital string.

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 of Algorithmic Probability.

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



1. Books by Hans Reichenbach, Andrey Kolmogorov. 2. Journal of Applied Physics. 3. Books by Bertrand Russell, Rudolf Carnap

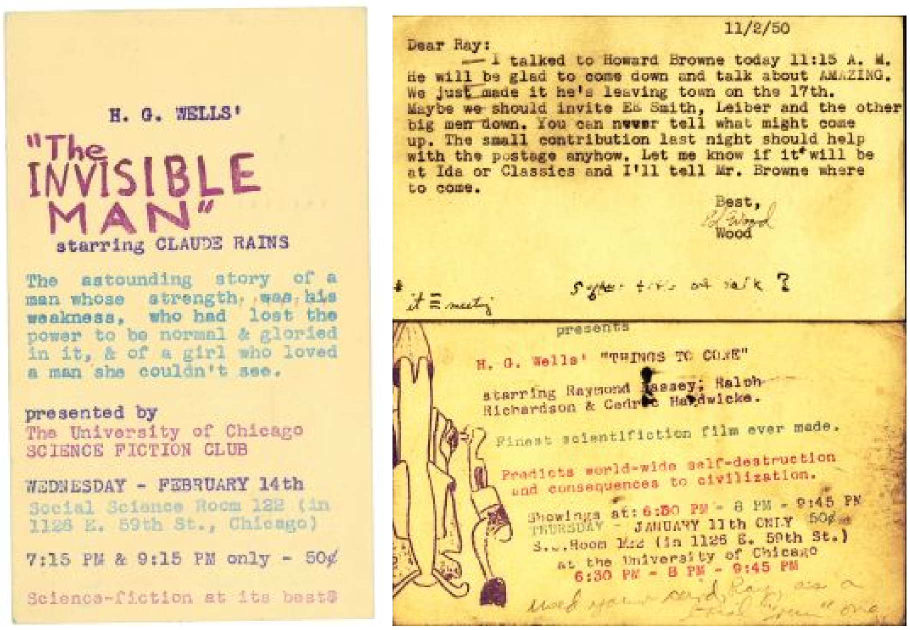
Here are some more of the items Ray was reading during college:

1. Science Fiction from early 1950s. 2. First issue of Flash Gordon. 3. Science Fiction fanzines, 1950.

Actually he was quite influenced by science fiction; Hugo Gernsback had started publishing "Amazing Stories" in 1926 and John Campbell's "Astounding" magazine, followed in 1930. Sci Fi fandom had recently been born, and

Ray was president of the newly formed, always cash-strapped, Sci Fi club at the University.

The UChicago Sci Fi Club featured talks! Movies! Postcards announced *The Invisible Man*, starring Claude Rains in “The astounding story of a man whose strength was his weakness, who had lost the power to be normal & gloried in it...!” and H.G. Wells *Things to Come*, the “Finest scientifiction movie ever made!” (note the word scientifiction — still in use then, as the terms science fiction, and sci fi weren’t always used yet.)



Thrilling sci fi postcards

These two movies, rather than freaking out with Big Bugs, Alien Monsters, are notable for showing the promise and danger when humans achieve technology breakthroughs.

His favorite movie was ‘The Shape of Things to Come’, that tells of a visionary people who would save a dying world through technology.

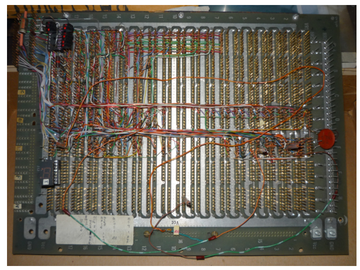
But Ray was also aware of the dangers. In his Junior year he even called for a conference of scientists to talk about the possibilities and dangers of cybernetics. There was no response. Only now, with the advent of AI Large Models, like Claude, Gemini, ChatGPT, is this call being heard.

Science Fiction represented something that was happening in the U.S. and in Europe also; an awareness of amazing things that were possible through this technology. The government sponsored many new projects.

As part of all this, the computer was rapidly becoming important. In 1941 electronic computers 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

Soon after college, Ray got a job in New York, working at Technical Research Group. In 1952 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 computing machine that could theoretically be programmed to solve any problem capable of a solution. He described two versions. One, later called a Turing Machine used an early version of a program, that was meant to solve some specific type of problem. The second, the Universal Turing Machine, was more like current computers. Given the right instructions it could mimic any of the "plain vanilla" Turing Machines.

This new concept brought the solving of a problem right into the world of machines. It used operational terms — not like math problems written on a paper but a step by step process – a program — that could direct the machine in solving a problem.

This step by step, operational view is foundational to the development of symbolic logic.

Turing also used his machine concept to prove that it is logically impossible to know whether the machine's program will ever finish. For example, it could get stuck in an infinite loop, or search for something forever. This was later called the halting problem.

Ray noted:

> I learned to understand and appreciate Turing machines - both universal and non- universal. Up until that time, I had only a poor understanding of formal logic and the limits imposed by Gödel's Theorems. The translation of formal logic and recursive function theory into theorems about Turing machines was a real revelation for me. 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. [17, p 8]

In Cambridge, Ray had frequent discussions with Marvin Minsky and John McCarthy and others about thinking, mathematics and machines.

And so, in 1955, John McCarthy, who was an Assistant Professor of Mathematics at Dartmouth, recognizing the incredible potential of this — whatever it was — and the possibilities for some kind of coherence about it, decided it was time to act.

He got together with Marvin Minsky, who was at MIT, Nat Rochester at IBM, and Claude Shannon, already well known for his work in information theory.

They picked Dartmouth, New Hampshire for their venue. The Math department there was headed by the remarkable John Kemeny, later president of Dartmouth and co-designer of the early programming language, Basic.

Other principal participants were also planned: Ray Solomonoff, Julian Bigelow, John Holland (who was unable to attend) for the full time. Allen Newell, Herb Simon, Oliver Selfridge, David Mackay (also unable to attend), Tom Etter, Trenchard More (as a substitute for Nat Rochester, who due to the pressure of his work could not attend all the sessions), would attend for shorter times, with quite a few others invited for various amounts of time.

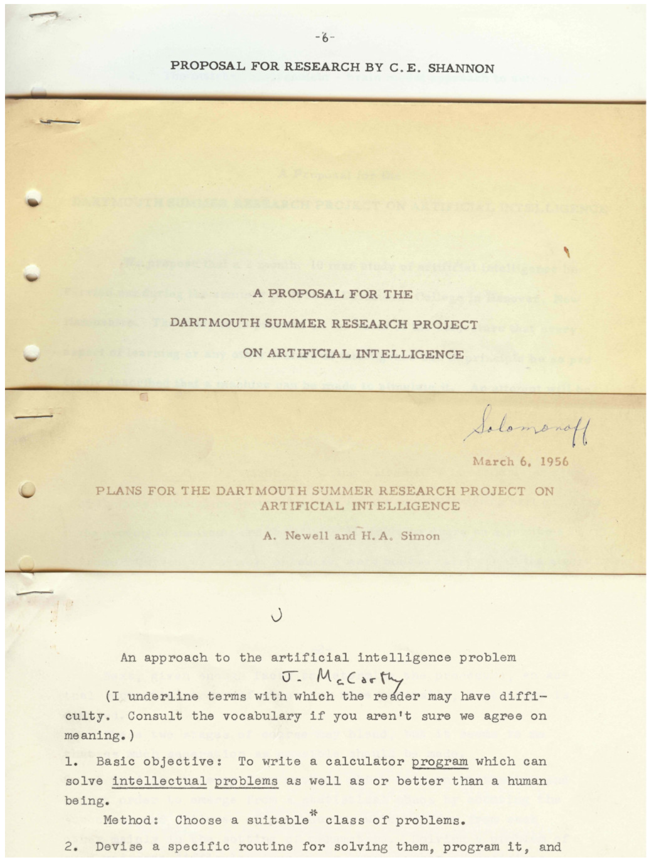
But you can't just invite people to "this new field." You have to call it something!

And that's when McCarthy boldly picked the name "Artificial Intelligence." He chose the name partly for its neutrality. He wanted to avoid alliance with a narrow recherché domain like automata theory [8, p.53].

He also wanted to avoid the term most commonly used: cybernetics. There were two main reasons for this: First, cybernetics was too heavily focused on analog feedback. Digital operations were much more relevant, especially as computers relied on digital switches, not on continuous variables like the rotation of a wheel.

Second, he would have to accept the extremely assertive Norbert Wiener as guru or have to argue with him [8, p. 53].

The basic idea was that the 10 participants would brainstorm during 2 months at Dartmouth, and then 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??



My proposal! Let's do MY proposal!

A lot of totally different Great AI Ideas! (Ray's collection includes 7 different proposals).

From most of these ideas came logic-based, deterministic programs using deduction. Ray, however, championed induction for prediction.

Induction and deduction are two principal philosophical views on how to solve a problem.

We use deduction to logically reason things out. e.g.,

My Acme Chatbot says 2 plus 2 is 5.
2 plus 2 is not 5.
Therefore I *deduce* that *certainly* my Chatbot is wrong (again!).

We use induction to draw probable conclusions from observing examples. e.g.,

I asked my Acme Robot 60 times to cook dinner.
60 times it just sat there.

If I ask it again, I *induce* that *probably* it will just sit there (like it did when I asked it to vacuum.)

The split between deduction and induction flowered into the two great Kingdoms: Artificial Intelligence and Machine Learning.

Ray's work, for many years, was considered a part of machine learning. Really it has been the advent of Large Models that properly brought the two together.

Ray wrote and presented a paper at Dartmouth, "An Inductive Inference Machine," on how machines could be made to improve themselves by induction, using unsupervised learning from examples.

In his notebook at the Dartmouth Workshop, he also developed another report, "Induction and the Problem of Learning"

In the notebook, he wrote: "Actually, practically all of the induction problems, aka time series, are special cases of the following: Given a large object that may be composed of many related or unrelated parts: suppose that part of this object is unknown to you - the problem is to predict just what it is. At any rate the concepts useful in time series prediction are also useful for the general induction problem." [10, p. 7]

He used this idea for probabilistic prediction in two talks he gave during the summer.

These talks were consolidated in the report he circulated there, and at the IRE convention in Sept. 1956. In 1957 a short version was published at the IRE western conference [11].

An Inductive Inference Machine held ideas that now resonate in modern AI: the use of statistical probability to predict what comes next, and the idea of explaining the same pattern in different ways to predict what the next piece of data in the series might be.

This report also replaced semantic with symbolic representation.

In 2010 Marvin wrote: "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."[3]

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 report from that summer, the most well known is Newell and Simon's "Logical Theorist," and a year 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.

Ray's focus was on inductive inference. How can an intelligence (human or artificial) predict and understand the future based on past data. To do this, an intelligence needs to form hypotheses. The Universal Turing Machine provided a mathematically rigorous way to define and quantify hypotheses using the concept of program length.

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 [18].

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.

Although there are glimmers, such as the concept of using different methods to solve the same problem, in Ray's 1956 writings, the report, "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 [12], with a revision in November 1960 [13]. He published a more complete exposition in two 1964 papers for the Journal of Information and Control [14], [15].

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" [9].

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" [13][12].

The second idea is similar to that of Epicurus's principle: it is an expansion on the shortest code theory; if more than one theory explains the data, keep all of the theories. In 1960, Ray wrote "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"[13] [12].

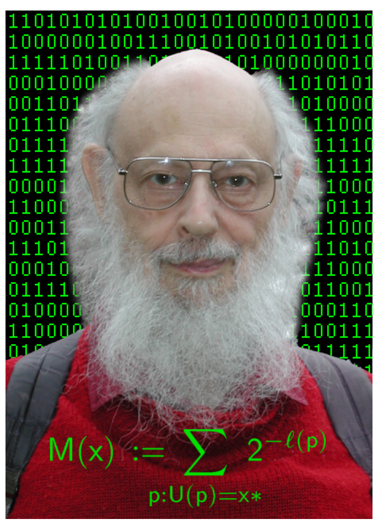
$$P(x)_M = \sum_{i=1}^{\infty} 2^{-|s_i(x)|}$$

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. He uses this with 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 (based on shortness) of individual programs to give a figure of merit to each program that could produce the sequence [13][14].

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 [13], [15].



Here is another beautiful picture of this formula in a 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 [4]. 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)" [18].

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"[18].

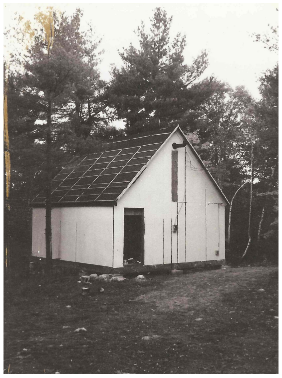
Vitányi notes: "we will associate Solomonoff's name with the universal distribution, and Kolmogorov's name with the descriptional complexity" [6]. Kolmogorov was interested in the information content of a string, while Ray was interested in the predictive power of a string.

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 (It wasn't that easy to explain 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.

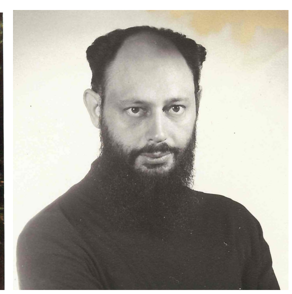


1. the house. 2: house decor. 3. A later picture by Alex Solomonoff of Ray working on the house.

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.





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 schlepped 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. Here is our logo:

Oxbridge Research
P.O. Box 559
Cambridge, Massachusetts 02238

Oxbridge Research requested and received technology grants, was part of a Stock Market Application group, and did many other projects. It was also a cover for our secret Trovaxo Laboratory.

On the left, Ray in Secret Trovaxo Chemical Experimentation Lab, developing important inventions, including the liquid helium zonk. On the right: our Senior Development Team putting secret Trovaxo experiments to rigorous testing

Meanwhile, in 1975 Gregory 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,[5] 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 best use of probability, even as different aspects of probability are now used in all Large Models. But in general logic and probabilistic reasoning have united. For example, in an interview with me, back in 2011, John McCarthy remarked that he had shifted his focus from deductive logic in his development of what he calls 'circumscription', to a more probabilistic system. Things like fuzzy logic and Bayesian work lie in the probabilistic area.

At the AI@50 conference in 2006, there was more discussion about whether AI should be logic-based or probability-based.[7, p. 88] "In defense of probability David Mumford argued that in the last 50 years, brittle logic has been displaced with probabilistic methods. Eugene Charniak supported this position by explaining how natural language processing is now statistical natural language processing. He stated frankly, 'Statistics has taken over natural language processing because it works.' "(Quote by James Moor)[7, p. 89] And the Dartmouth (America's oldest college newspaper) quoted Ray "It felt like it was magic," — "The idea that a machine could do things that before we thought only humans could do, that was sort of a breakthrough." — and the newspaper continues: "[Solomonoff] says he often thinks about the possible dark side of technology that could make weapons easier to produce. But he hopes people will manage these dangers and that AI could complement human knowledge and "solve really hard problems we haven't solved," like curing cancer or AIDS."


News

AI conference returns to College after 50 yrs.

By Alex Belser, The Dartmouth Staff

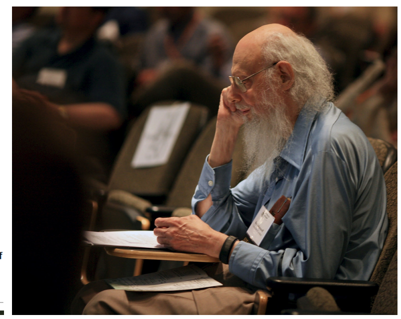
Published on Tuesday, July 18, 2006

Fifty years after a group of about 10 young scientists first met to start the nascent field of artificial intelligence, some of them returned for a fiftieth anniversary conference this weekend entitled AI@50, the Dartmouth Artificial Intelligence Conference: The Next Fifty Years.

The three-day program, which lasted from Thursday through Saturday, consisted of over 40 presentations including a few retrospective lectures about the history of the field, but primarily focused on the future,

Lauren Wool/The Dartmouth Staff

Pioneers of artificial intelligence reconvened at Dartmouth for the "AI@50" conference last weekend.



1. Dartmouth news July 2006. 2. Ray at the AI@50 conference.

Ray inspired many people, especially young people. Our nephew Alex came to us from Cleveland. He says "There was nothing like this for me in Cleveland. Ray talked about mathematics in a way that made it exciting." Alex became part of our family and Ray influenced him to get his Ph.D. in mathematics.

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.[19][20] It chooses only the shortest code to work with, however, while Algorithmic Probability uses as many as there is time for.

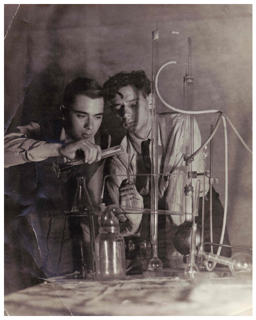
In other papers Ray wrote about how best to limit search by limits on time or computation cost. He developed methods for working with other types of data, not just sequences. He categorized problems into various types such as "inversion problems" and "time-limited optimization problems" and developed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

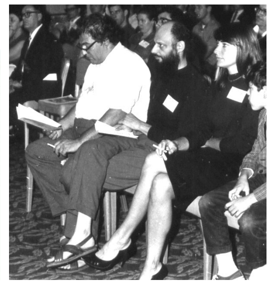
Throughout his career Ray was concerned with the potential benefits and dangers of AI, discussing it in some of his published reports. In 1985 he analyzed a likely evolution of AI, giving a formula predicting when it would reach the "Infinity Point." This Infinity Point is an early version of the "Singularity" later made popular by Ray Kurzweil.[16]

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."

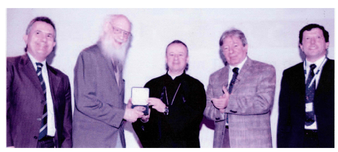
However, he would meet frequently with Minsky, Shannon, and others at MIT, and with other researchers throughout the world. We went to many countries and conferences meeting amazing and dedicated people. There was Saarland University in Saarbrücken, Germany, where Wolfgang Paul invited Ray to do research 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. Dolittle 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:

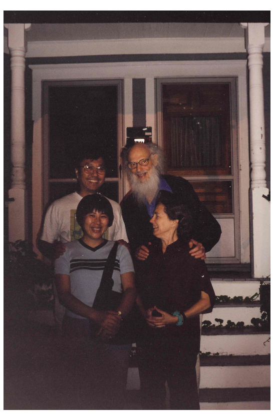


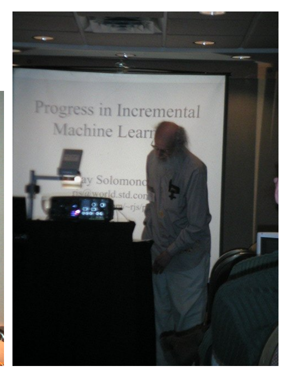


1. Ray and Irv in Ray's Lab, high school age. 2. Camp counsellor during College. 3. With Maggie Lettvin at Boskone I, Sci Fi convention 1965.

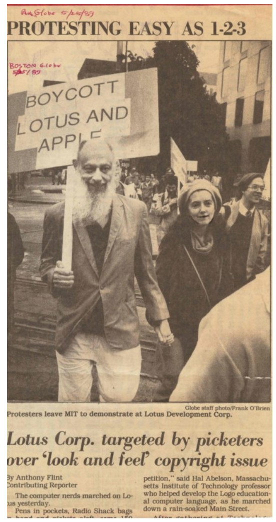




1. Around 1965, Ray, Mary and Rollo Silver, Marvin Minsky, (?), Ed Fredkin. 2. Ray in Lebanon, his sponsor, Fouad Chedid, on the right. 3. At Royal Holloway, UK: Rissanen, Vovk, Gammerman, Vapnik, Wallace and Ray





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



PROTESTING EASY AS 1-2-3

Globe staff photo/Frank O'Brien

Protesters leave MIT to demonstrate at Lotus Development Corp.

Lotus Corp. targeted by picketers over 'look and feel' copyright issue

By Anthony Flint
Contributing Reporter

The computer nerds marched on Lotus yesterday.

Pens in pockets, Radio Shack bags

petition," said Hal Abelson, Massachusetts Institute of Technology professor who helped develop the Logo educational computer language, as he marched down a rain-soaked Main Street.

1. Father Ray; picture by Alex. 2. Geeks protest Lotus keyboard layout copyright. 3. At our place in New Ipswich.

Ray was happy when the AGI '08 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" [2].

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

[1] Rudolf Carnap. *Logical Foundations of Probability*. University of Chicago Press, Chicago, 1950.

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

[3] Memorial Guests. *Ray Solomonoff Memorial Notes*. March 2010.

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

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

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

[7] J Moor. The dartmouth college artificial intelligence conference: The next fifty years. *AI Magazine*, 27(4):87–91, 2006.

[8] N. Nilsson. *The Quest for Artificial Intelligence*. Cambridge University Press, 2010.

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

[10] Ray Solomonoff. Induction and the problem of learning. http://raysolomonoff.com/dartmouth/notebook, 1956. handwritten manuscript.

[11] Ray Solomonoff. An inductive inference machine. In *IRE Convention Record, Section on Information Theory, Part 2*, pages 56–62, New York, 1957. Institute of Radio Engineers.

[12] 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.

[13] 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.

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

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

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

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

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

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

[20] 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.

[21] N. Wiener. *Cybernetics*. MIT Press, 1948.

[22] Norbert Wiener. *The Human Use of Human Beings: Cybernetics and Society*. Da Capo Press, New York, N.Y., 1988.