



## From AI to ML, from Logic to Probability

2018 Workshop on Finance, Insurance, Probability and Statistics (FIPS 2018)  
King's College, London, UK

Dr. Paul A. Bilokon

Imperial College London  
Kensington, London SW7 2AZ

Thalesians Ltd  
Level39, One Canada Square, Canary Wharf, London E14 5AB

11 September, 2018

## The Dartmouth Workshop

- In **summer 1956** a group of researchers gathered at a workshop organised by **John McCarthy**, then a young Assistant Professor of Mathematics, at **Dartmouth College** in Hanover, New Hampshire.
- The attendees included:
  - Marvin Minsky (1927–2016)
  - Trenchard More
  - Nathaniel Rochester (1919–2001)
  - Oliver Selfridge (1926–2008)
  - Claude Shannon (1916–2001)
  - Herbert Simon (1916–2001)
  - Ray Solomonoff (1926–2009)
- The stated goal was ambitious:

*The study is to proceed on the basis of the conjecture that **every aspect of learning** or any other feature of **intelligence** can, in principle, be so precisely described that a **machine** can be made to **simulate it**. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.*
- Thus the field of **Artificial Intelligence (AI)** was born.

## The term “Artificial Intelligence”

It was **John McCarthy** who coined the term **Artificial Intelligence**, which he defined as *the science and engineering of making intelligent machines, especially intelligent computer programs.*



Here we see him hosting a series of four simultaneous computer chess matches carried out via telegraph against rivals in Russia ten years later, in 1966.

## The participants in 1956

Do you recognise them?



## The participants in 1956

Do you recognise them?



From left: Trenchard More, Nathaniel Rochester, Ray Solomonoff, Marvin Minsky, John McCarthy, Oliver Selfridge, and Claude Shannon.

## The participants reunited at the July 2006 AI@50 conference

Do you recognise them?



## The participants reunited at the July 2006 AI@50 conference

Do you recognise them?



From left: Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff.

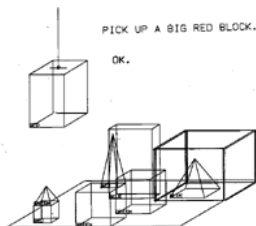
## Scruffy AI and blocks world

- **Marvin Minsky** (1927–2016) was on the MIT faculty from 1958.
- **Seymour Papert** (1928–2016) joined the faculty in 1963.
- Minsky and Papert concentrated on
  - just getting systems to do interesting things — “scruffy AI”. [HP15]*
- Their students worked on problems in limited **domains**,
  - application areas not requiring a broad general knowledge. Successful examples included the domains of integral calculus, geometry, and algebra as well as a famous series of problems in the “blocks world”, a simplified world consisting of some toy blocks sitting on a table. [HP15]*



# SHRDLU

- **Terry Winograd** (b. 1946) was at the time a PhD student of Seymour Papert.
- In 1968–70 he wrote **SHRDLU** [Win71, Win72], a program for understanding natural language.
- It conducted a simple dialog (via teletype) with a user about a **blocks world** shown on an early display screen (DEC-340 attached to a PDP-6 computer).



- Ever wondered how SHRDLU got its **name**? The answer is here:  
<http://hci.stanford.edu/winograd/shrdlu/name.html>

## The dialogue that was used as a SHRDLU demo

---

Person: PICK UP A BIG RED BLOCK.

Computer: OK. (does it)

Person: GRASP THE PYRAMID.

Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.

Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE  
↪ BOX.

Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM  
↪ HOLDING.

Computer: OK. (does it)

Person: WHAT DOES THE BOX CONTAIN?

Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.

Person: WHAT IS THE PYRAMID SUPPORTED BY?

Computer: THE BOX.

Person: HOW MANY BLOCKS ARE NOT IN THE BOX?

Computer: FOUR OF THEM.

Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?

Computer: YES, THE RED CUBE.

Person: IS IT SUPPORTED?

Computer: YES, BY THE TABLE.

Person: CAN THE TABLE PICK UP BLOCKS?

Computer: NO.

Person: CAN A PYRAMID BE SUPPORTED BY A BLOCK?

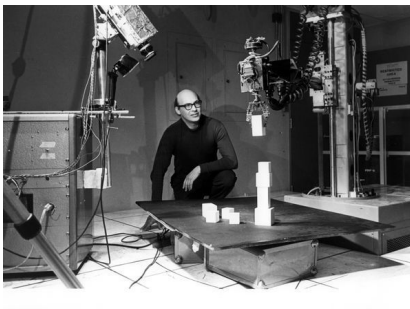
Computer: YES.

...

---

## Vision, mechanical manipulation, robotics

- Their work was not all about **simulation**. Here Minsky is pictured with a **physical** system consisting of a video camera and a robotic arm used to manipulate a physical blocks world:



- Further details of this work can be found in [MP71].

## Formal logic and LISP (i)

- **John McCarthy** came to MIT around the same time as Minsky.
- He advocated the use of **formal logic** in artificial intelligence.
- He invented **LISP**, and co-developed it with R. Brayton, D. Edwards, P. Fox, L. Hodes, D. Luckham, K. Maling, D. Park, S. Russell [McC60]:

*A programming system called **LISP (for LISt Processor)** has been developed for the IBM 704 computer by the Artificial Intelligence group at M.I.T. The system was designed to facilitate experiments with a proposed system called the Advice Taker, whereby a machine could be instructed to handle **declarative as well as imperative sentences** and could exhibit **"common sense"** in carrying out its instructions. The original proposal [McC58] for the Advice Taker was made in November 1958. The main requirement was a programming system for manipulating expressions representing formalized declarative and imperative sentences so that the Advice Taker system could make deductions.*

*In the course of its development, the LISP system went through several stages of simplification and eventually came to be based on a **scheme for representing the partial recursive functions of a certain class of symbolic expressions**.*

- McCarthy's work [McC60] was influenced by that of Allen Newell, J. Cliff Shaw and Herbert A. Simon on **Logic Theorist** [NS57], "the first artificial intelligence program" [Cre93], which would eventually prove 38 of the first 52 theorems in Alfred North Whitehead's and Bertrand Russell's *Principia Mathematica* [WR10].

## Formal logic and LISP (ii)

- He made use of **partial functions**:

A **partial function** is a function that is defined only on part of its domain. Partial functions necessarily arise when functions are defined by computations because for some values of the arguments the computation defining the value of the function may not terminate.

- **Propositional expressions and predicates**:

A **propositional expression** is an expression whose possible values are  $T$  (for truth) and  $F$  (for falsity). We shall assume that the reader is familiar with the propositional connectives  $\wedge$  ("and"),  $\vee$  ("or"), and  $\sim$  ("not")... A **predicate** is a function whose range consists of the truth values  $T$  and  $F$ .

- **Conditional expressions**, "a device for expressing the dependence of quantities on propositional quantities".

- **Recursive function definitions**:

By using conditional expressions we can, without circularity, define functions by formulas in which the defined function occurs. For example, we write

$$n! = (n = 0 \rightarrow 1, T \rightarrow n \cdot (n - 1)!)$$

## Formal logic and LISP (iii)

- Finally, McCarthy made extensive use of Alonzo Church's  $\lambda$ -calculus, [McC60]:

*It is usual in mathematics — outside of mathematical logic — to use the word “function” imprecisely and to apply it to forms such as  $y^2 + x$ . Because we shall later compute with expressions for functions, we need a distinction between **functions** and **forms** and a **notation** for expressing this distinction. This distinction and a notation for describing it, from which we deviate trivially, is given by Church [Chu41].*

*Let  $f$  be an expression that stands for a function of two integer variables. It should make sense to write  $f(3, 4)$  and the value of this expression should be determined. The expression  $y^2 + x$  does not meet this requirement;  $y^2 + x(3, 4)$  is not a conventional notation and if we attempted to define it we would be uncertain whether its value would turn out to be 13 or 19. Church calls an expression like  $y^2 + x$  a **form**. A form can be **converted into a function** if we can determine the correspondence between the variables occurring in the form and the **ordered list of arguments** of the desired function. This is accomplished by Church's  $\lambda$ -notation.*

*If  $\mathcal{E}$  is a form in variables  $x_1, \dots, x_n$ , then  $\lambda((x_1, \dots, x_n), \mathcal{E})$  will be taken to be the function of  $n$  variables whose value is determined by substituting the arguments for the variables  $x_1, \dots, x_n$  in that order in  $\mathcal{E}$  and evaluating the resulting expression. For example,  $\lambda((x, y), y^2 + x)$  is a function of two variables, and  $\lambda((x, y), y^2 + x)(3, 4) = 19$ .*

## Criticism of the Logistic Approach

- Minsky was critical of the use of logic for representing knowledge. In an appendix to a widely disseminated preprint of [Min75], entitled *Criticism of the Logistic Approach*, which was removed from the published version, Minsky wrote:

*Because logicians are not concerned with systems that will later be enlarged, they can design axioms that permit only the conclusions they want. In the development of intelligence, the situation is different. One has to learn which features of situations are important, and which kinds of deductions are not to be regarded seriously.*

- Thus McCarthy's approach diverged from Minsky's and in 1963 McCarthy left MIT to start the *Stanford Artificial Intelligence Laboratory*. [HP15]
- As an alternative to formal logic, Minsky advocated an approach based on *frames* [Min75].
- Minsky's approach wasn't without its critics either, but...  
*Widely criticized as a trivial combination of *semantic nets* [Ric56] and *object-oriented programming* [DMN70, BDMN73], Minsky's frames paper served to place *knowledge representation* as a *central issue for AI*. [MMH98, p. 23]*

## Early artificial neural networks

**Artificial neural networks** are not a new idea: they originate from earlier work [PK05, Section 1.4]:

- As early as 1873, researchers such as the logician Alexander Bain [Bai73] and psychologist William James [Jam90] were imagining man-made systems based on neuron models.
- Warren McCulloch and Walter Pitt showed that neurons were **Turing-capable** and developed a **logical calculus of ideas immanent in nervous activity** [MP43], which Stephen Cole Kleene recognised as related to **finite automata** [Kle56].
- Donald Olding Hebb considered the role of the neurons in **learning** and developed a learning rule based on **reinforcement** to strengthen connections from important inputs — **Hebbian learning** [Heb49]. Hebb stated what would become known as **Hebb's postulate**:

*When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.*

According to [Med98],

*From a neurophysiological perspective, **Hebbian learning** can be described as a time-dependent, local, highly interactive mechanism that increases synaptic efficacy as a function of pre- and post-synaptic activity.*



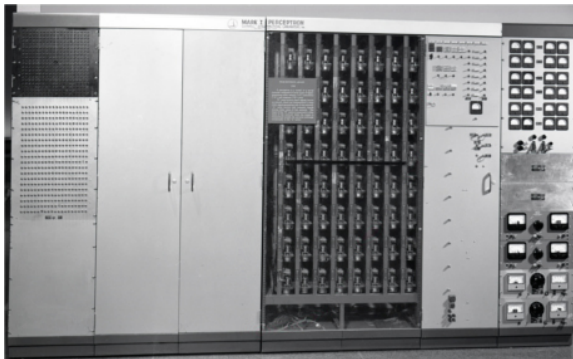
## Connectionist versus symbolic/structural AI

- Belmont G. Farley and Wesley A. Clark [FC54] and Nathaniel Rochester, John H. Holland, L. H. Haibt and W. L. Duda [RHHD56] **simulated Hebbian networks** — interconnected networks of simple units — on computers.
- Hebb also introduced the term **connectionism**, which would later be used to describe the approaches to UI based on interconnected networks of simple units.
- Other approaches to UI, such as those pioneered by Minsky, Papert and McCarthy may be described as **structural** or **symbolic**.

## The perceptron

- Working on pattern classification, Frank Rosenblatt (1928–1971) of the Cornell Aeronautical Laboratory invented the **perceptron** [Ros57, Ros60].
- It was first implemented on IBM 704 and then as a custom-built machine, the **Mark I Perceptron**.
- That machine had an array of 400 photoresistors, randomly connected to the “neurons”. The weights were encoded in potentiometers and weight updates were carried out by electric motors [Cor60, Bis07].
- Around the same time another early **feedforward** neural network algorithm was produced by Bernard Widrow and his first PhD student, Ted Hoff: the least mean squares (LMS) algorithm, also known as the Widrow–Hoff rule [WH60].
- In the next year, 1961, Widrow and his students developed the earliest learning rule for feedforward networks with multiple adaptive elements: the Madaline Rule I (MRI) [Wid62].
- Applications of LMS and MRI were developed by Widrow and his students in fields such as pattern recognition, weather forecasting, adaptive control, and signal processing. The work by R. W. Lucky and others at Bell laboratories led to first applications to adaptive equalisation in high-speed modems and adaptive echo cancellers for long-distance telephone and satellite circuits.

## Mark I Perceptron (i)



The Mark I Perceptron on exhibition at the National Museum of History and Technology, March 1968.

## Mark I Perceptron (ii)

According to the manual [Cor60],

*The **Mark I Perceptron** is a **pattern learning** and **recognition** device. It can learn to **classify plane patterns into groups** on the basis of certain geometric similarities and differences. Among the properties which it may use in its discriminations and generalizations are position in the retinal field of view, geometric form, occurrence frequency, and size.*

*If, of the many possible bases of classification, a particular one is desired, it can generally be transferred to the perceptron by a **forced learning** session or by an **error correction training** process. If left to its own resources the perceptron can still divide up into classes the patterns presented to it, on a classification basis of its own forming. This formation process is commonly referred to as **spontaneous learning**.*

*The **Mark I** is intended as an experimental tool for the direct study of a limited class of perceptrons. It is sufficiently flexible in configuration and operation to serve as a **model for any of a large number of perceptrons** possessing a single layer of non-cross-coupled association units.*

## The rise and fall of the perceptron (i)

- During a 1958 press conference, Rosenblatt made rather **strong statements** that were reported by The New York Times as follows:

*WASHINGTON, July 7 (UPI) — The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.*

- These comments caused **skepticism** among some researchers.
- In 1969, Minsky and Papert published *Perceptrons: An introduction to computational geometry* [MP69]. The book used mathematics, notably topology and group theory, to prove some results about the capabilities and limitations of simple networks of perceptrons.
- It contained some positive, but also negative results:
  - A single perceptron is incapable of implementing some predicates, such as the **XOR logical function**.
  - Predicates such as **parity** and **connectedness** also cause serious difficulties for perceptrons.

## The rise and fall of the perceptron (ii)

- The publication of the book led to the “XOR affair” [Dek13]:

*the story that circulates goes like this: “Marvin Minsky, being a proponent of structured AI, killed off the connectionism approach when he co-authored the now classic tome, Perceptrons. This was accomplished by mathematically proving that a single layer perceptron is so limited it cannot even be used (or trained for that matter) to emulate an XOR gate. Although this does not hold for multi-layer perceptrons, his word was taken as gospel, and smothered this promising field in its infancy.”*

*Marvin Minsky begs to differ, and argues that he of course knew about the capabilities of artificial neural networks with more than one layer, and that if anything, only the proof that working with local neurons comes at the cost of some universality should have any bearing.*

- Indeed, the earlier work of Warren McCulloch and Walter Pitts [MP43] had already shown that neural networks were Turing capable.
- Critics of the 1969 book posed other arguments that its publication, either intentionally or unintentionally, led to a decline in neural networks research for a decade.

## The rise and fall of the perceptron (iii)

In his review of the book's 1988 expanded edition, Jordan B. Pollack, a proponent of connectionism, writes [Pol89] that

*Minsky and Papert surrounded their 1969 mathematical tract with **fairly negative judgements** and **loaded terms**, such as the following quotes, which have been used as evidence [DD88, RZ85] that they actually intended to stifle research on perceptron-like models.*

- *Perceptrons have been widely publicized as “pattern recognition” or “learning” machines and as such have been discussed in a large number of books, journal articles, and voluminous “reports”. Most of this writing... is **without scientific value**. (p. 4)*
- *We do not see that **any good can come of experiments** which pay no attention to limiting factors that will assert themselves as soon as the small model is scaled up to a usable size. (p. 18)*
- *[We] became involved with a somewhat therapeutic compulsion: to dispel what we feared to be the first shadows of a **“holistic” or “Gestalt” misconception** that would threaten to haunt the fields of engineering and artificial intelligence... (p. 20)*
- *There is **no reason to suppose** that any of these virtues **carry over to the many layered version**. Nevertheless, we consider it to be an important research problem to elucidate (or reject) our intuitive judgement that the extension is sterile. (p. 231)*

## The rise and fall of the perceptron (iv)

Pollack continues:

*Despite these pronouncements, in 1988, Minsky and Papert wish to **deny their responsibility**, or, at least, their **intentionality**, in bringing about the decade-long connectionist winter:*

- *One popular version is that the publication of our book **so discouraged research** on learning in network machines that a promising line of research was interrupted. Our version is that progress had already come to a **virtual halt** because of the **lack of adequate basic theories**. (p. xii)*



## The rise and fall of the perceptron (v)

Pollack argues that

*the **real problem** which **terminated the research viability** of perceptron-like models was the problem of **scaling**. Minsky and Papert asserted that as such learning models based on gradient descent in weight space were scaled up, they would be impractical due to local minimal extremely large weights and a concurrent growth in convergence time.*

*So, were they responsible for killing Snow White? No, since **intention** and **action** are separable, they were no more responsible than Bill, who, intending to kill his uncle, is "so nervous and excited [when driving] that he accidentally runs over and kills a pedestrian, who happens to be his uncle" [Sea80] If Minsky and Papert did not **intend to stifle** the field of neural networks, then, perhaps, they would act in accordance with their new motto: "**We see no reason to choose sides**" (p. xiv).*

but agrees that

*Perceptrons, and its authors, certainly have their places **assured in history**.*

## Connectionist winter

- Whatever the reason, neural networks became **unpopular** in the 1970s and few research groups continued research in this subject.
- Stephen Grossberg developed a self-organising neural network model known as **Adaptive Resonance Theory (ART)** [Gro76a, Gro76b].
- Teuvo Kohonen worked on **matrix-associative memories** [Koh72] and **self-organisation of neurons** into topological and tonotopical mappings of their perceived environment [Koh82, Koh88].

## The discovery of backpropagation

- In 1971 Paul John Werbos developed a method of training multilayer neural networks through **backpropagation** of errors. It was described in his 1974 PhD thesis at Harvard University *Beyond Regression: New Tools for Prediction and Analysis in Behavioral Sciences* [Wer74]. This work later appeared in extended form in his book *The Roots of Backpropagation* [Wer94]. See also [Wer90].
- This was a **major extension of feedforward neural networks** beyond the MRI rule of [Wid62].
- The backpropagation technique was **rediscovered** by D. B. Parker in 1985 and appeared in his technical report at MIT [Par85].
- At around the same time, during his PhD, in 1985, Yann LeCun proposed and published (at first, in French) a different version of the backpropagation algorithm [LeC88].
- This work received little attention until backpropagation was **refined** and **popularised** by David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams [RHW86].
- Backpropagation made it feasible to train **multilevel** neural networks with **high degrees of nonlinearity** and with **high precision**. See [WL90] for a review and example applications.

## The Hopfield network

- In 1982 John Hopfield [Hop82] invented the **associative neural network**, now known as the **Hopfield network**.
- Hopfield's focus was on the **collective action** of the network and not of the **individual neurons**.
- Hopfield networks serve as **content-addressable ("associative")** memory systems with binary threshold nodes.
- They are **guaranteed to converge to a local minimum**, but may sometimes converge to a false pattern (wrong local minimum) rather than the stored pattern (expected local minimum).
- Hopfield modeled the functioning of the neural network as an **energy minimisation process**.
- The discovery of backpropagation and the Hopfield network **rekindled interest** in neural networks and revived this research area. For more detailed history, see [PK05, WL92].

## The resurgence of AI as ML, deep learning

- The recent resurgence of **Artificial Intelligence (AI)** as **Machine Learning (ML)** was facilitated by advances in **artificial neural networks**.
- A **deep neural network (DNN)** [GBC17] is an artificial neural network (ANN) with **multiple hidden layers** between the input and output layers. Such networks can model complex nonlinear relationships.
- **Backpropagation** is a major ingredient in making much work with deep neural networks feasible.
- Contributions by Geoffrey E. Hinton and others [Hin89, HS06, HOT06, Hin07] have enabled the pre-training of multilayer feedforward neural networks one layer at a time, treating each layer in turn as an unsupervised restricted Boltzmann machine, then fine-tuning it using supervised backpropagation.
- This — along with advances in software and hardware — has made it computationally feasible to train and apply DNNs.
- Applications of DNNs — **deep learning** — has been at the core of the renewed interest in machine learning.

## AI versus ML



Nidhi Chappell, Intel

*AI is basically the **intelligence** — how we make machines intelligent, while machine learning is the implementation of the compute methods that support it. The way I think of it is: **AI is the science** and **machine learning is the algorithms** that make the machines smarter. So the enabler for AI is machine learning.*

## ML and probability theory

- Modern books on machine learning [HTF11, GBC17] introduce **probability theory** as one of its foundations.
- In [GBC17], Section 3.1, *Why Probability?*, the following justification is given:

*Many branches of computer science deal mostly with entities that are entirely **deterministic and certain**. A programmer can usually safely assume that a CPU will execute each machine instruction flawlessly. Errors in hardware do occur but are rare enough that most software applications do not need to be designed to account for them. Given that many computer scientists and software engineers work in a **relatively clean and certain environment**, it can be surprising that machine learning makes heavy use of probability theory.*

*Machine learning must always deal with **uncertain quantities** and sometimes **stochastic (nondeterministic) quantities**. Uncertainty and stochasticity can arise from **many sources**. Researchers have made compelling arguments for quantifying uncertainty using probability since at least the 1980s. Many of the arguments presented here are summarized from or inspired by [Pea88].*

*Nearly all activities require some ability to reason in the presence of uncertainty. In fact, beyond mathematical statements that are true by definition, it is difficult to think of any proposition that is **absolutely true** or any event that is **absolutely guaranteed** to occur.*

## Random experiment and the sample space

- A **random experiment**  $\mathcal{E}$  is an experiment such that
  - 1 all possible distinct outcomes of the experiment are known in advance;
  - 2 the actual outcome of the experiment is not known in advance with certainty;
  - 3 the experiment can be repeated under identical conditions.
- The **sample space**,  $\Omega$ , is the set of all possible outcomes of a random experiment.
- A subset  $A \subseteq \Omega$  of the sample space is referred to as an **event**.
- The empty set  $\emptyset \subseteq \Omega$  is referred to as the **impossible event**.
- The sample space itself,  $\Omega \subseteq \Omega$ , is referred to as the **certain event**.



## Example of a random experiment

- The random experiment  $\mathcal{E}$  consists in a *single toss* of an *unbiased coin*.
- The possible outcomes of this experiment are:
  - $\omega_1 = \text{"heads"}$ ,
  - $\omega_2 = \text{"tails"}$ .
- The sample space is thus  $\Omega = \{\omega_1 = \text{"heads"}, \omega_2 = \text{"tails"}\}$ .
- There are exactly *four* events —  $2^{|\Omega|} = 2^2 = 4$  subsets of  $\Omega$ :
  - $H = \{\omega_1\} = \text{"heads (obverse) comes up"}$ ;
  - $T = \{\omega_2\} = \text{"tails (reverse) comes up"}$ ;
  - $\emptyset = \{\} = \text{"nothing comes up"}$  — if we *do* perform the experiment  $\mathcal{E}$ , this will *never* occur, so this is indeed the **impossible event**;
  - $\Omega = \{\omega_1, \omega_2\} = \text{"either heads or tails comes up"}$  — if we *do* perform the experiment  $\mathcal{E}$ , this is *guaranteed* to occur, so this is indeed the **certain event** (we disregard the possibility of the coin landing on its edge — the third side of the coin; otherwise we'd need a separate outcome in  $\Omega$  to model this possibility).

## The classical interpretation of probability

- Let  $A$  be an event associated with an experiment  $\mathcal{E}$  so that  $A$  either occurs or does not occur when  $\mathcal{E}$  is performed.
- *Assume* that  $\Omega$  is finite.
- Furthermore, *assume* that all outcomes in  $\Omega$  are *equally likely*.
- Denote by  $M(\cdot)$  the number of outcomes in an event; thus  $M(A)$  is the number of outcomes in  $A$ ,  $M(\Omega)$  the number of outcomes in  $\Omega$ .
- Then the probability of  $A$  is given by

$$\mathbb{P}(A) = \frac{M(A)}{M(\Omega)}.$$

## The classical interpretation of probability: an example

- Let us continue our example where the random experiment  $\mathcal{E}$  consists in a single toss of an unbiased coin.
- For the event  $H = \{\omega_1\}$ , according to the classical interpretation of probability,

$$\mathbb{P}(H) = \frac{M(H)}{M(\Omega)} = \frac{1}{2}.$$

- But what if  $\Omega$  is *not finite*?
- And what if the coin is *biased*?

## The frequentist interpretation of probability

- Let  $A$  be an event associated with an experiment  $\mathcal{E}$  so that  $A$  either occurs or does not occur when  $\mathcal{E}$  is performed.
- Consider a **superexperiment**  $\mathcal{E}^\infty$  consisting in an infinite number of independent performances of  $\mathcal{E}$ .
- Let  $N(A, n)$  be the number of occurrences of  $A$  in the first  $n$  performances of  $\mathcal{E}$  within  $\mathcal{E}^\infty$ .
- Then the probability of  $A$  is given by

$$\mathbb{P}[A] = \lim_{n \rightarrow \infty} \frac{N(A, n)}{n}.$$

- This interpretation of probability is known as the **long-term relative frequency (LTRF)** (or **frequentist**, or **objectivist**) [Wil01, page 5].
- The claim is that, in the long term, as the number of trials approaches infinity, the relative frequency will converge *exactly* to the true probability.
- It requires that the probabilities be estimated from *samples*.
- Unknown quantities, such as means, variances, etc., are considered to be *fixed* but *unknown*.

## Question

Can you use the frequentist interpretation of probability to compute the probability of the existence of extraterrestrial life?

## Bayesian interpretation of probability

- In **Bayesian (subjectivist, epistemic, evidential)** interpretation, the probability of an event is the *degree of belief* that that event will occur.
- This degree of belief can be determined on the basis of
  - empirical data,
  - past experience, or
  - subjective plausibility.
- Bayesian probability can be assigned to *any statement*, whether or not a random experiment is performed.
- Unknown quantities, such as means, variances, etc., are regarded to follow a *probability distribution*, which expresses our degree of belief about that quantity at a particular time.
- On arrival of *new information*, the degree of belief can be *updated*.

## The axiomatic interpretation of probability

- Andrey Nikolaevich Kolmogorov (1903–1987): “The theory of probability as a mathematical discipline can and should be developed from axioms in exactly the same way as Geometry and Algebra.” [Kol33]
- **Kolmogorov’s axioms** of probability:
  - **First axiom:** For any event  $E$ ,  $\mathbb{P}[E] \in \mathbb{R}$ ,  $\mathbb{P}[E] \geq 0$ . (The assumption of finite measure.)
  - **Second axiom:**  $\mathbb{P}[\Omega] = 1$ . (The assumption of unit measure.)
  - **Third axiom:** For any countable collection of disjoint events  $E_1, E_2, \dots$ ,  
 $\mathbb{P}[\bigcup_{i=1}^{\infty} E_i] = \sum_{i=1}^{\infty} \mathbb{P}[E_i]$ . (The assumption of  $\sigma$ -additivity.)
- Consistency:
  - The LTRF and Bayesian interpretations motivated Kolmogorov’s axioms and are *consistent* with them.
  - The LTRF interpretation reappears in the axiomatic interpretation as a theorem — the **Strong Law of Large Numbers**.
  - The axioms describe how probability *behaves*, not what probability *is*... Or is Kolmogorov saying that what probability is is defined by the way it behaves? (“When I see a bird that walks like a duck and swims like a duck and quacks like a duck, I call that bird a duck.” — Indiana poet James Whitcombe Riley, around 1916.)

## History: Andrey Nikolayevich Kolmogorov (1903–1987)



**Andrey Nikolaevich Kolmogorov** was one of the founders of modern (measure-theoretic) probability theory. Its foundational axioms, often referred to as Kolmogorov axioms, first appeared in a German monograph entitled *Grundbegriffe der Wahrscheinlichkeitrechnung* in the *Ergebnisse der Mathematik* in 1933 [Kol33]. A Russian translation by G. M. Bavli was published in 1936, which was used to produce an English translation [Kol56]



## Consequences of the axioms

- **Null empty set:**  $\mathbb{P}[\emptyset] = 0$ .
- **Complement rule:** for any event  $A$ ,  $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$ .
- **Difference rule:** for any events  $A, B$ , if  $A \subseteq B$ ,  $\mathbb{P}[B \setminus A] = \mathbb{P}[B] - \mathbb{P}[A]$ .
- **Monotonicity rule:** for any events  $A, B$ , if  $A \subseteq B$ , then  $\mathbb{P}[A] \leq \mathbb{P}[B]$ .
- **The upper bound on probability is 1:** for all  $A$ ,  $\mathbb{P}[A] \leq 1$ .
- **Inclusion-exclusion rule:** for any events  $A, B$ ,  

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$
- **Bonferroni inequality:** for any events  $\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$ .
- **Continuity property:** If the events  $A_1, A_2, \dots$  satisfy  $A_1 \subseteq A_2 \subseteq \dots$  and  $A = \bigcup_{i=1}^{\infty} A_i$ , then  $\mathbb{P}[A_i]$  is increasing and  $\mathbb{P}[A] = \lim_{i \rightarrow \infty} \mathbb{P}[A_i]$ . If the events  $B_1, B_2, \dots$  satisfy  $B_1 \supseteq B_2 \supseteq \dots$  and  $B = \bigcap_{i=1}^{\infty} B_i$ , then  $\mathbb{P}[B_i]$  is decreasing and  $\mathbb{P}[B] = \lim_{i \rightarrow \infty} \mathbb{P}[B_i]$ .
- **Borel–Cantelli Lemma:** For any events  $A_1, A_2, \dots$ , if  $\sum_{i=1}^{\infty} \mathbb{P}[A_i] < \infty$ , then  $\mathbb{P}[\bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} A_j] = 0$ .<sup>1</sup>
- The rest of probability theory!

---

<sup>1</sup>The event  $\mathbb{P}[\bigcap_{i=1}^{\infty} \bigcup_{j=i}^{\infty} A_j]$  is sometimes referred to as “ $A_i$  infinitely often” or as the limit superior of the  $A_j$ ,  $\limsup_{j \rightarrow \infty} A_j$ .

## Frequentist vs Bayesian interpretation of probability

- The frequentist approach is (arguably) *objective*.
- The Bayesian approach is (arguably) *subjective*.
- The frequentist approach uses only *new data* to draw conclusions.
- The Bayesian approach uses *both new and past data, and belief*, to draw conclusions.

## Probability theorists or logicians?

- On the face of it, **probability theory** developed independently of **logic**...
- However, some of the great **probability theorists** of the 20th century either started off, or became **logicians**!
- **Andrey Nikolayevich Kolmogorov** wrote *On the principle of the excluded middle* in 1925 *On the Interpretation of Intuitionistic Logic* [Kol32] in 1931, before many of his probability-theoretic papers, and around the same time as *Grundbegriffe der Wahrscheinlichkeitsrechnung* [Kol33].
- Kolmogorov would later — in 1953 — worked on the generalisation of the concept of **algorithm** [Kol53].
- He was Head of the **Mathematical Logic Group (Kafedra)** at Moscow State University from 1980 until the end of his life in 1978.
- **Norbert Wiener's** PhD thesis completed at Harvard University in 1913 was entitled *A comparison Between the Treatment of the Algebra of Relatives by Schroeder and that by Whitehead and Russell* [Wie13] and his supervisors were the philosopher Karl Schmidt and Josiah Royce, the latter being among the founding fathers of the Harvard school of logic, Boolean algebra, and foundations of mathematics.

## Stochastic processes

- **Probability space:**  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\Omega$  is a set,  $\mathcal{F}$  is a  $\sigma$ -algebra of its subsets and  $\mathbb{P}$  is a measure on  $(\Omega, \mathcal{F})$  such that  $\mathbb{P}(\Omega) = 1$
- **Real-valued random variable  $X$ :** an  $(\mathcal{F}, \mathcal{B}\mathbb{R})$ -measurable function  $X : \Omega \rightarrow \mathbb{R}$
- **Law of the random variable  $X$ :** the image measure of  $\mathbb{P}$  under  $X$ ,  $\mathbb{P}_X : \mathbb{R} \rightarrow [0, 1]$ ,  
 $\mathbb{P}_X(B) := \mathbb{P} \circ X^{-1}(B)$
- **Stochastic process  $X$ :** a parametrised (by some indexing set  $T$  representing **time**) collection of random variables,  $\{X_t\}_{t \in T}$ , defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  and assuming values in the same measurable space
- Can also be viewed as a random variable on  $(\Omega, \mathcal{F}, \mathbb{P})$  taking values in  $(C(T, S), \mathcal{B}C(T, S))$
- **Law of the stochastic process  $X$ :** the pushforward probability measure  $\mathbb{P} \circ X^{-1} : \mathcal{B}(C[T, S]) \rightarrow [0, 1]$

## Brownian motion and Wiener measure

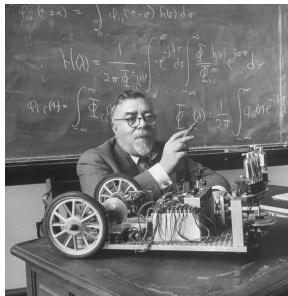
- **Brownian motion**: the stochastic process  $W$  such that
  - $W_0 = 0$
  - $t \mapsto W_t$  a.s. everywhere continuous
  - independent increments with  $W_t - W_s \sim N(0, t - s)$
- **Wiener measure** is the law of  $W$
- The Wiener measure of a **basic point-open** set of continuous functions from  $[0, 1]$  to  $\mathbb{R}$ , i.e. a set of the form  $\{f \mid a_i < f(t_i) < b_i, 0 = t_0 < t_1 < \dots, < t_n = 1\}$ , is given by

$$\frac{1}{\sqrt{\pi^n \prod_{i=1}^n (t_i - t_{i-1})}} \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} e^{-\sum_{j=1}^n \frac{(x_j - x_{j-1})^2}{t_j - t_{j-1}}} dx_n \dots dx_1,$$

where  $x_0 := 0$ .

- Brownian motion was studied extensively by **Albert Einstein** and its law was constructed by **Norbert Wiener** [Wie23]
- Probably the most important stochastic process, a paradigmatic **martingale**
- Ubiquitous in stochastic analysis [Øks10, KS91] and mathematical finance [Shr04]

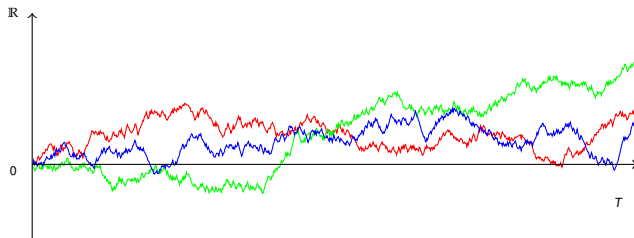
## History: Norbert Wiener (1894–1964)



**Norbert Wiener** produces the first construction of the law of the Brownian motion and publishes it in 1923 [Wie23]

## The trajectories of the Brownian motion

- The following graph shows three **trajectories** or **realisations** of  $W$
- Each trajectory corresponds to a particular  $\omega \in \Omega$
- We shall assume  $T = [0, 1]$



## Brownian motion as the limit of the symmetric random walk

- For  $n \in \mathbb{N}^*$ , let

$$X_n = \begin{cases} +1, & \text{with probability } \frac{1}{2}, \\ -1, & \text{with probability } \frac{1}{2}, \end{cases}$$

thus each  $X_n$  is a Bernoulli random variable

- Let  $Y_0 := 0$  and, for  $n \in \mathbb{N}^*$ , let  $Y_n := \sum_{i=1}^n X_i$
- We have thus constructed a real-valued discrete time stochastic process  $Y_n$ . This process is called a **symmetric random walk**
- For a given  $N \in \mathbb{N}^*$ , define the stochastic process  $Z$ , which we shall refer to as the **scaled symmetric random walk**:  $Z_t^{(N)} = \frac{1}{\sqrt{N}} Y_{Nt}$  for all  $t \in \{0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N}{N}, \frac{N+1}{N}, \dots\} =: T^{(N)}$ , i.e. such  $t$  that make  $Nt$  a nonnegative integer, ensuring that  $Y_{Nt}$  is well defined
- We can turn  $Z^{(N)}$  into a continuous time stochastic process by means of linear interpolation: for  $t \in [0, +\infty)$ , define

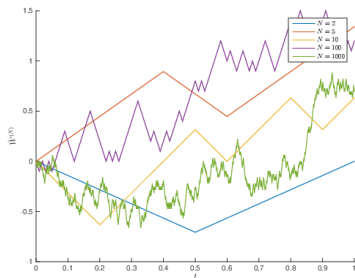
$$\hat{W}_t^{(N)} := Z_{\frac{n}{N}}^{(N)} + \left(t - \frac{n}{N}\right) \left[Z_{\frac{n+1}{N}}^{(N)} - Z_{\frac{n}{N}}^{(N)}\right],$$

where  $n \in \mathbb{N}^0$  is such that  $\frac{n}{N} \leq t < \frac{n+1}{N}$  (clearly it is unique, so  $\hat{W}_t^{(N)}$  is well defined).

- One can prove, using the CLT, that, for  $s, t \in [0, +\infty)$ ,  $s \leq t$ , the distribution of  $\hat{W}_t^{(N)} - \hat{W}_s^{(N)}$  approaches normal with mean 0 and variance  $t - s$  as  $N \rightarrow +\infty$



## The trajectories of the symmetric random walk



Several sample paths of the scaled symmetric random walk process,  $\hat{W}^{(N)}$ , generated using different arrays of random variates (each sample path corresponds to a different  $\omega \in \Omega$ ) and different values of  $N$ . The time is restricted to  $[0, 1]$ .

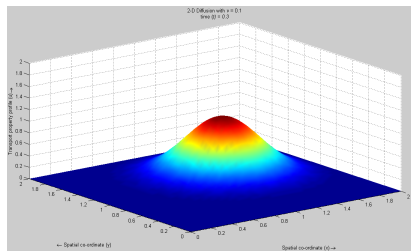
## Brownian motion and the heat equation (and other PDEs)

- Let  $u(x, t)$  be the temperature at location  $x$  at time  $t$ .
- The heat equation is given by

$$\frac{\partial}{\partial t} u(x, t) = \frac{1}{2} \Delta_x u(x, t).$$

- It can be written in terms of Brownian motion using the Feynman-Kac formula:

$$u(x, t) = \mathbb{E} [u(W_t + x, 0)].$$



## Aleatory versus epistemic uncertainty

- Let us start with a quote from a paper on reliability engineering by **Der Kiureghian** and **Ditelvsen** [DKD07]:

*While there can be many sources of uncertainty, in the context of modeling, it is convenient to categorize the character of uncertainties as either aleatory or epistemic. The word **aleatory** derives from the Latin “alea”, which means the rolling of dice. Thus, an aleatoric uncertainty is one that is presumed to be the intrinsic randomness of a phenomenon. Interestingly, the word is also used in the context of music, film and other arts, where a randomness or improvisation in the performance is implied. The word **epistemic** derives from the Greek “episteme”, which means knowledge. Thus, an epistemic uncertainty is one that is presumed as being caused by lack of knowledge (or data)*

- **Domain theorists** are usually concerned with **epistemic** uncertainty: e.g. the “approximate” or “partial” reals  $[a, b] \in \mathbb{IR}$ ,  $a < b$ , represent the partial knowledge about some perfect real  $r \in [a, b] \subseteq \mathbb{R}$  at a given stage of the computation [Sco70a, AJ94, ES98]. However, the probabilistic power domain can handle **both** kinds of uncertainty
- **Probability theorists**, as we shall see, are concerned with **both** kinds of uncertainty. How they handle them depends on their interpretation of probability

## Classical probability theory incorrectly propagates ignorance

- If **probability theory** can express both **aleatory** and **epistemic** uncertainty, why bother with domain theory?
- Under **Laplace's Principle of Insufficient Reason**, the uncertainty about a parameter must be modelled with a uniform distribution, assigning equal probabilities to all possibilities. Bayesians refer to these as **uninformative priors**, **not very informative priors**, etc.
- Surely the assertion “The value of  $X$  lies in the interval  $[a, b]$  (but its probability distribution is unknown)” contains **strictly less** information than “The value of  $X$  is uniformly distributed on  $[a, b]$ ”?
- Inability to distinguish between the two in classical probability theory leads to **problems**, as described by **Ferson** and **Ginzburg** [FG96]:
  - Classical probability theory incorrectly propagates ignorance
  - Second-order Monte Carlo methods require unjustified assumptions
  - *Probability theory and interval analysis can (and should) be combined*
- We employ **domain theory** to this end to construct **partial stochastic processes**. Partial stochastic processes are to classical stochastic processes what partial reals are to classical reals

## Domain theory



Dana Scott (b. 1932)

- **Domain theory** was introduced by **Dana Scott** in the late 1960s and early 1970s as a **mathematical theory of computation**. According to Scott [Sco70b],  
*the theory is based on the idea that **data types** can be **partially ordered** by a relation similar to that of **approximation**, and as a result can be considered as **complete lattices**.*
- In the same work, Scott argues that the theory ought to be **mathematical** rather than **operational** in its approach. The mathematical meaning of a procedure ought to be the function from elements of the data type of inputs to elements of the data types of the outputs. The operational meaning will generally provide a trace of the whole history of its computation.
- One of the first applications of the theory was the construction of the first mathematical model for the untyped  $\lambda$ -calculus [Sco70b].

## Domain theory



Yuri Leonidovich Ershov  
(b. 1940)

- In the USSR, **Yuri Leonidovich Ershov** carried out extensive work on domain theory.
- Part of it was **independent** of and **contemporary** with Scott's work.
- Elsewhere Ershov **answered many questions** that were posed by Scott but were left unanswered [GHK<sup>+</sup>03].
- Therefore in the literature the Scott domains also sometimes called **Scott–Ershov domains**, as in [Bla00], for example.

## Computational models for classical spaces



Abbas Edalat

- **Abbas Edalat** applied domain theory to produce **computational models** of **classical mathematical spaces**.
- This research project started in **1993** and is still **ongoing**.
- The idea is to use domain theory to **reconstruct** some basic mathematics.
- This is achieved by **embedding** classical spaces into the set of **maximal elements** of suitable **domains**.
- **Applications** have included the dynamical systems, measures and fractals [Eda95b] and integration [Eda95a].

## Elements of domain theory (i)

- **Poset**  $(D, \sqsubseteq)$ : a set  $D$  with a binary relation  $\sqsubseteq$  which is reflexive, anti-symmetric, and transitive
- **Supremum**  $x \in D$  of a subset  $A \subseteq D$ : an upper bound of  $A$  s.t. whenever  $y$  is any other upper bound of  $A$ ,  $x \sqsubseteq y$ . We write  $x = \bigsqcup A$
- A nonempty  $A \subseteq D$  is **directed** if, for all  $a, b \in A$ , there exists  $c \in A$  with  $a \sqsubseteq c$  and  $b \sqsubseteq c$



## Elements of domain theory (ii)

- A **directed-complete** poset (**dcpo**): each of its directed subsets has a supremum
- A **bounded-complete** poset: each of its subsets that has an upper bound has a supremum
- $D$  dcpo,  $x, y \in D$ .  $x$  **approximates**  $y$  ( $x \ll y$ ) if  $\forall$  directed  $A \subseteq D$ ,  $y \sqsubseteq \bigsqcup A \Rightarrow x \sqsubseteq a$  for some  $a \in A$
- $B_x$  a **basis** for  $D$ :  $\forall x \in D$ ,  $B_x := \downarrow x \cap B$  contains directed subset with supremum  $x$
- **( $\omega$ -) continuous** dcpo: dcpo with a (countable) basis
- **Domain**:  $\omega$ -continuous dcpo
- **Scott domain**: bounded complete domain

## Scott topology

- We can define **topologies** on dcpos.
- As Abramsky and Jung point out [AJ94], in domain theory we can tie up **open sets** with the concrete idea of **observable properties** (see [Smy92]).
- Let  $(D, \sqsubseteq)$  be a dcpo. A subset  $G$  of  $D$  is said to be **Scott open** if it satisfies the following two conditions:
  - 1 the subset  $G$  is an upper set, i.e.  $\uparrow G = G$ , and
  - 2 if  $A \subseteq D$  is a directed subset with  $\bigsqcup \uparrow A \in G$ , then there is some  $x \in A$  such that  $\uparrow x \subseteq G$ .
- Condition (2) is equivalent to saying that  $G$  has a non-empty intersection with  $A$  whenever  $A$  is directed and its supremum is in  $G$ .
- In words, Scott open sets can be described as upper (Condition (1)) and inaccessible by directed suprema (Condition (2)).
- The collection  $\mathcal{T}_S(D)$  of all Scott open sets of the dcpo  $(D, \sqsubseteq)$  is a topology, so  $(D, \mathcal{T}_S(D))$  is a topological space.
- We call the collection  $\mathcal{T}_S(D)$  of all Scott open sets of the dcpo  $(D, \sqsubseteq)$  the **Scott topology** of  $D$ .
- Unlike the usual (Euclidean) topology, this topology is non-Hausdorff. Such topologies are considered in great depth in Jean Goubault-Larrecq's recent text [GL13].
- M. B. Smyth [Smy92] explains that the Scott topology can be seen as a topology of **positive information**, whereas the Lawson topology can be seen as a topology of **positive-and-negative information**. The computational content of the Lawson topology is further discussed in [JES06].

## Topology: intuition (i)

- An open set containing a point  $x$  is called a **neighbourhood** of that point.
- Thus an open set is a neighbourhood of *each* of its points.
- A neighbourhood of a point  $x$  can be thought of a set of points that are “**sufficiently close**” to  $x$ . Different neighbourhoods specify different degrees of closeness.
- For example, if we take the real line,  $\mathbb{R}$ , with its usual (Euclidean) topology, then the intervals

$$(x - 1, x + 1), \left(x - \frac{1}{2}, x + \frac{1}{2}\right), \left(x - \frac{1}{3}, x + \frac{1}{3}\right), \dots, \left(x - \frac{1}{i}, x + \frac{1}{i}\right), \dots$$

are all neighbourhoods of  $x \in \mathbb{R}$  of increasing “**degree of closeness**”.

- Remember that  $X$  itself is open, so a neighbourhood of all of its points. Somehow the open set  $X$  encodes the “lowest” “degree of closeness”. In this loose sense, all the points in  $X$  are “close”.

## Topology: intuition (ii)

- Intuitively, “putting together” two neighbourhoods — two “degrees of closeness” — also gives a “degree of closeness”.
- Therefore the *union* of any (arbitrary) family of open sets is again an open set: for each point belonging to the union, a neighbourhood of that point is a subset of the union, so the union itself is a neighbourhood of that point.
- What about the *intersection*?
- Consider two open sets,  $O_1, O_2 \in \mathcal{T}$ . Consider some  $x \in O_1 \cap O_2$ . The elements of  $O_1$  are precisely all the points in  $X$  that are close to  $x$  to some “degree of closeness 1”. The elements of  $O_2$  are precisely all the points in  $X$  that are close to  $x$  to some “degree of closeness 2”. The elements of  $O_1 \cap O_2$  are precisely all the points in  $X$  that are close to  $x$  to both “degree of closeness 1” *and* “degree of closeness 2” — thus  $O_1 \cap O_2$  represents a stronger “degree of closeness” than either  $O_1$  or  $O_2$ . It is natural that  $O_1 \cap O_2$  should also be an open set.
- Inductively, any *finite* intersection of open sets should be an open set:  $(\dots(((O_1 \cap O_2) \cap O_3) \cap O_4) \cap O_5) \cap \dots) \cap O_n$  for some  $n \in \mathbb{N}^*$ .

## Topology: intuition (iii)

- What about *countable* intersections?
- Consider an example. Take  $x \in \mathbb{R}$ . The intervals

$$(x - 1, x + 1), \left(x - \frac{1}{2}, x + \frac{1}{2}\right), \left(x - \frac{1}{3}, x + \frac{1}{3}\right), \dots, \left(x - \frac{1}{i}, x + \frac{1}{i}\right), \dots$$

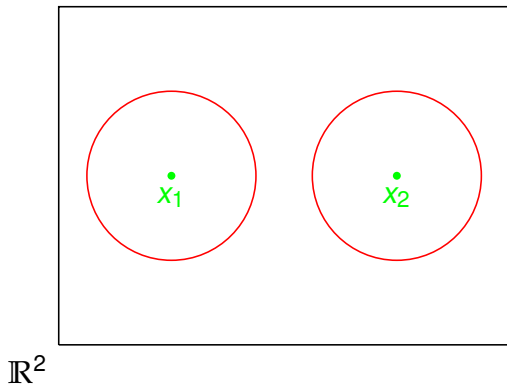
all contain  $x$  and consists of points that are “close” to  $x$ . Their *countable*, not *finite*, intersection

$$\bigcap_{i=1}^{\infty} \left(x - \frac{1}{i}, x + \frac{1}{i}\right) = \{x\}$$

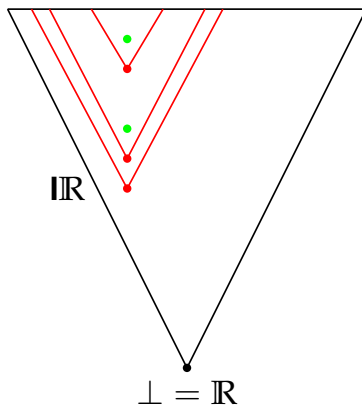
is precisely the **singleton**  $\{x\}$ .

- If we admit *countable* (let alone arbitrary!) intersections into a topology we end up with *too many* sets, since any subset of  $X$  can be written as an arbitrary union of singleton sets.
- If all singleton sets were open, all sets would be open, all sets would be closed, only finite sets would be compact, and each function  $f : \mathbb{R} \rightarrow X$  would be continuous. This isn't particularly meaningful!

# Hausdorff topologies



## Non-Hausdorff topologies



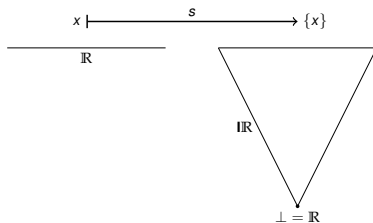
## Domain-theoretic computational models

- A **(domain-theoretic computational) model** of a topological space  $X$  is a continuous domain  $D$  together with a homeomorphism  $\phi : X \rightarrow S$ , where  $S \subseteq \text{Max}(D)$  is a  $G_\delta$  subset of the maximal elements  $\text{Max}(D)$  carrying its relative Scott topology inherited from  $D$
- Introduced by **Abbas Edalat** in [Eda97]



## Interval domain

- **Interval domain:**  $\mathbb{IR} := \{[a, b] \mid a, b \in \mathbb{R} \wedge a \leq b\}$
- Ordered by reverse subset inclusion
- For directed  $A \subseteq \mathbb{IR}$ ,  $\sqcup A = \bigcap A$
- $I \ll J \Leftrightarrow J \subseteq I^\circ$
- $\{[p, q] \mid p, q \in \mathbb{Q} \wedge p \leq q\}$  a countable basis for  $\mathbb{IR}$



## Overview of the domain-theoretic framework

- [BE17, BE14] introduces a **domain-theoretic framework** for continuous time, continuous state stochastic processes
- Their laws are embedded into the space of maximal elements of a **normalised probabilistic power domain** on the space of **continuous interval-valued functions** endowed with the relative Scott topology
- The resulting  $\omega$ -continuous bounded complete dcpo is used to define **partial stochastic processes** and characterise their **computability**
- For a given stochastic process, **finitary approximations** are constructed. Their lub is the process's law
- Applying this to Brownian motion and its law, the Wiener measure, a **partial Wiener measure** is constructed, giving a proof of the **computability** of the Wiener measure, alternative to the one by **Willem L. Fouché** [Fou00, DF13])

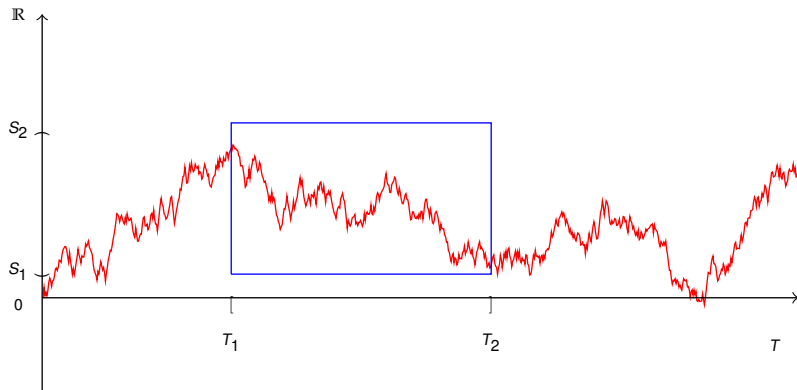
## Domain-theoretic function spaces

- Investigated by **Thomas Erker**, **Martín Escardó** and **Klaus Keimel** [EEK98]
- $X$ : locally compact Hausdorff space,  $\mathcal{O}(X)$ : its lattice of open subsets,  $L$ : bounded complete domain
- For  $O \in \mathcal{O}(X)$ ,  $s \in L$ , a **single-step function** is the continuous map

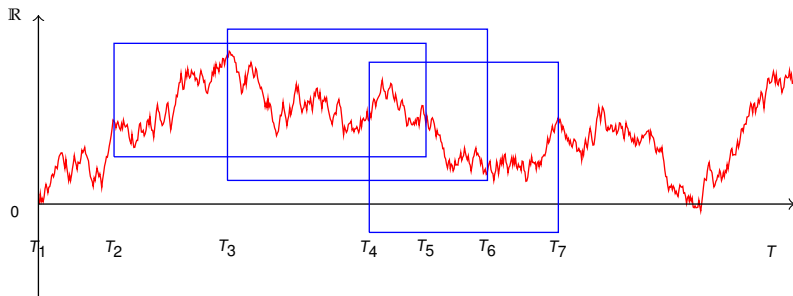
$$a\chi_O(x) = \begin{cases} a, & \text{if } x \in O; \\ \perp, & \text{otherwise} \end{cases}$$

- **Step function**: join of a bounded finite collection of single-step functions
- $[X \rightarrow L]$ : set of all continuous functions  $g : X \rightarrow L$ ; a bounded complete domain w.r.t. pointwise order induced by  $L$
- Basis: step functions

## Single-step function $\rightarrow$ subbasic compact-open set



## Step function $\rightarrow$ basic compact-open set



## Topology: definition

Let  $X$  be a set and  $\mathcal{T}$  a collection of subsets of  $X$ . Then  $\mathcal{T}$  is a **topology** on  $X$  iff:

- 1 both the empty set  $\emptyset$  and  $X$  are elements of  $\mathcal{T}$ ;
- 2 *arbitrary* unions of elements of  $\mathcal{T}$  are also elements of  $\mathcal{T}$ ;
- 3 *finite* intersections of elements of  $\mathcal{T}$  are also elements of  $\mathcal{T}$ .

# Valuations

- **Valuation** on top. space  $(X, \mathcal{T})$ : map  $v : \mathcal{T} \rightarrow [0, \infty)$  s.t.
  - **Modularity**:  $v(G) + v(H) = v(G \cup H) + v(G \cap H)$
  - **Strictness**:  $v(\emptyset) = 0$
  - **Monotonicity**:  $G \subseteq H \Rightarrow v(G) \leq v(H)$  for all  $G, H \in \mathcal{T}$
- It is **probabilistic** if  $v(X) = 1$  and **continuous** if, for directed  $A \subseteq \mathcal{T}$ ,  
 $v(\bigcup_{G \in A} G) = \bigsqcup_{G \in A} v(G)$
- Unlike measures, valuations are defined on open, rather than measurable, sets.  
Favoured in computable analysis
- Nice properties and extension results. See **Mauricio Alvarez-Manilla** *et al* [AMESD00] and **Jean Goubault-Larrecq** [GL05]

## Probabilistic power domain

- **(Normalised) probabilistic power domain**  $\mathbf{P}(X)$ : set of continuous valuations (with  $v(X) = 1$ ) ordered pointwise: for  $v, v' \in \mathbf{P}(X)$ ,  $v \sqsubseteq v'$  iff for all open sets  $G \in \mathcal{T}$ ,  $v(G) \leq v'(G)$
- Introduced by **Nasser Saheb-Djahromi** [SD80] and studied extensively by **Claire Jones** and **Gordon Plotkin** [JP89, Jon90]
- For any  $b \in X$ , the **point valuation**  $\delta_b : \mathcal{O}(X) \rightarrow [0, \infty)$  defined by

$$\delta_b(O) = \begin{cases} 1, & \text{if } b \in O; \\ 0, & \text{otherwise.} \end{cases}$$

- Any finite linear combination  $\sum_{i=1}^n r_i \delta_{b_i}$  with  $r_i \in [0, \infty)$ ,  $1 \leq i \leq n$ , is a continuous valuation on  $X$  (called a **simple valuation**).
- If  $X$  is an  $\omega$ -continuous dcpo with  $\perp$ , then  $\mathbf{P}^1(X)$  is also an  $\omega$ -continuous dcpo with bottom element  $\delta_\perp$  and has a basis consisting of simple valuations



## Probabilistic power domain: an important result

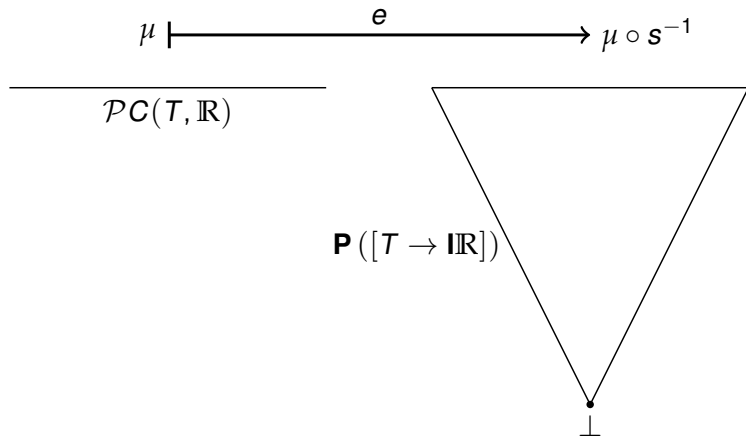
- Let  $D$  be an  $\omega$ -continuous domain. A valuation  $\nu$  in  $\mathbf{P}(D)$  is maximal in  $\mathbf{P}(D)$  (i.e.  $\nu \in \text{Max}(\mathbf{P}(D))$ ) iff  $\nu$  is supported in the set  $\text{Max}(D)$  of maximal elements of  $D$
- The “if” direction of this result was proved by **Abbas Edalat** [Eda95b, Proposition 5.18]
- The “only if” direction by **Jimmie D. Lawson** [Law98, Theorem 8.6]

## Domain-theoretic model for stochastic processes

- $\mathcal{PC}(T, \mathbb{R})$ : the space of probability measures on  $C(T, \mathbb{R})$  endowed with the weak topology
- $e : \mathcal{PC}(T, \mathbb{R}) \rightarrow \mathbf{P}([T \rightarrow \mathbb{IR}])$ ,  $e(\mu) = \mu \circ s^{-1}$ , embeds  $\mathcal{PC}(T, \mathbb{R})$  onto the set of maximal elements of  $\mathbf{P}^1([T \rightarrow \mathbb{IR}])$
- For a simple valuation  $\nu := \sum_{j=1}^n r_j \delta_{g_j}$ ,  $n \in \mathbb{N}^*$ , and  $l \in \mathbb{R}^+$  define the ***l*-mass of  $\nu$**  by  $m_l(\nu) := \sum_{j=1}^n \{r_j \mid |g_j| < l\}$
- Let  $\nu_1 \sqsubseteq \nu_2 \sqsubseteq \nu_3 \sqsubseteq \dots$  be an increasing chain of simple valuations in  $\mathbf{P}([T \rightarrow \mathbb{IR}])$  with  $\nu_i := \sum_{j=1}^{n_i} r_{ij} \delta_{g_{ij}}$ ,  $n_i \in \mathbb{N}^*$
- Define  $\nu := \bigsqcup_{n \in \mathbb{N}^*} \nu_n$
- Then the support of  $\nu$  is in the subspace of the embedded classical functions iff, for all  $n \in \mathbb{N}^*$ , there exists  $N \in \mathbb{N}^*$  such that

$$m_{1/n}(\nu_N) > 1 - 1/n \tag{1}$$

## The new picture



## Finitary approximation of a valuation

- $D$  is a bounded complete domain with a countable basis  $B := (b_1, b_2, \dots)$  closed under finite suprema
- Let  $\nu$  be any valuation on  $D$  and  $\nu^*$  its canonical extension to a measure on  $D$ . In particular,  $\nu^*$  could be the law of the stochastic process of interest
- We will show how to obtain  $\nu$  as a supremum of an increasing chain of simple valuations on  $D$
- Recursively define a sequence of finite lists of subsets of  $B$ : define  $A_0 := [a_1^0 := \perp]$ ; for  $n \in \mathbb{N}^0$ ,

$$A_{n+1} = [b_{n+1} \sqcup a_{i_1}^n, \dots, b_{n+1} \sqcup a_{i_{L_n}}^n, a_1^n, \dots, a_{K_n}^n],$$

where  $a_1^n, \dots, a_{K_n}^n$  are the elements of  $A_n$  in order, and  $a_{i_1}^n, \dots, a_{i_{L_n}}^n$  is the sublist of  $A_n$  consisting of those elements that have an upper bound with  $b_{n+1}$ . ( $L_n \leq K_n$ )

- For example,

$$A_1 = [b_1, \perp];$$

$$A_2 = \begin{cases} [b_2 \sqcup b_1, b_2, b_1, \perp] & \text{if } b_2 \sqcup b_1 \text{ exists,} \\ [b_2, b_1, \perp] & \text{otherwise} \end{cases}$$

## Finitary approximation of a valuation

- Further, for  $n \in \mathbb{N}^0$ ,  $v_n := \sum_{i=1}^{K_n} r_i^n \delta_{a_i^n}$ , where

$$r_i^n := v^* \left( \uparrow a_i^n \setminus \bigcup_{k=1}^{i-1} \uparrow a_k^n \right) \quad (2)$$

- The sequence of simple valuations  $(v_n)_{n \in \mathbb{N}}$  is an increasing chain, i.e., for all  $n \in \mathbb{N}$ ,  $v_n \sqsubseteq v_{n+1}$
- The supremum of the approximating chain  $(v_n)_{n \in \mathbb{N}}$  of simple valuations gives the approximated valuation:

$$\bigsqcup_{n \in \mathbb{N}} v_n = v$$

## Some details of the proof: monotonicity (i)

- To prove that the sequence is increasing, we use the modification [Eda95a] of the splitting lemma [JP89] for the normalised probabilistic power domain: we need to show the existence of the nonnegative numbers (called **transport numbers**)  $t_{i,j}^n$  for

$$i = 1, \dots, K_n, j = 1, \dots, K_{n+1},$$

such that, for a fixed  $i$ ,  $\sum_{j=1}^{K_{n+1}} t_{i,j}^n = r_i^n$ ; for a fixed  $j$ ,  $\sum_{i=1}^{K_n} t_{i,j}^n = r_j^{n+1}$ ; and  $t_{i,j}^n \neq 0$  implies  $a_i^n \sqsubseteq a_j^{n+1}$ .

- We claim that these requirements are satisfied by defining the transport numbers as follows. If  $b_{n+1} \sqcup a_i^n$  exists, then  $i = l_{j_i}$  for a unique  $j_i \in \{1, \dots, L_n\}$ , and we define

$$t_{i,j_i}^n := r_{j_i}^{n+1}, \quad t_{i,L_n+i}^n := r_{L_n+i}^{n+1}, \quad t_{i,j}^n := 0$$

for all  $j \notin \{j_i, L_n + i\}$ . If  $b_{n+1} \sqcup a_i^n$  does not exist, then we define

$$t_{i,L_n+i}^n := r_{L_n+i}^{n+1}, \quad t_{i,j}^n := 0$$

for all  $j \neq L_n + i$

## Some details of the proof: monotonicity (ii)

The intuition behind the above proof is as follows. In  $\nu_n$ , the weight of  $a_i^n$  is  $r_i^n$ , which in  $\nu_{n+1}$  is “distributed” in the weight of  $a_i^n$  and possibly the weight of  $b_{n+1} \sqcup a_i^n$ . If the supremum  $b_{n+1} \sqcup a_i^n$  does not exist, then the weight of  $a_i^n$  in  $\nu_{n+1}$  is the same as in  $\nu_n$  (because removing the set above  $b_{n+1}$  does not change the set); if  $b_{n+1} \sqcup a_i^n$  does exist, then  $\hat{r} a_i^n = (\hat{r} a_i^n \setminus \hat{r} (b_{n+1} \sqcup a_i^n)) \cup (\hat{r} (b_{n+1} \sqcup a_i^n))$ , which implies that the two weights in  $\nu_{n+1}$  sum to  $r_i^n$ .

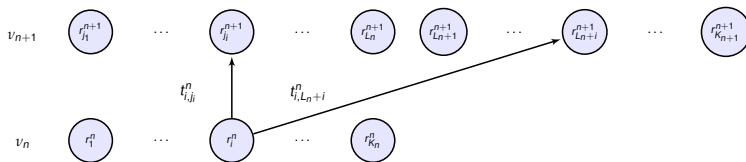


Figure: Transport numbers

## Some details of the proof: convergence (i)

[Eda97, Lemma 3.1] Let  $\nu_1$  and  $\nu_2$  be continuous valuations on a topological space  $X$ . Suppose  $B \subseteq \mathcal{O}(X)$ , where  $\mathcal{O}(X)$  is the topology of  $X$ , is a base which is closed under finite intersections. If  $\nu_1(\mathcal{O}) = \nu_2(\mathcal{O})$  for all  $\mathcal{O} \in B$ , then  $\nu_1 = \nu_2$ .

The countable basis  $B$  for our domain  $D$  gives rise to the topological base for its Scott topology, consisting of the sets  $\uparrow b_k$  for each  $b_k \in B$ ,  $k \in \mathbb{N}^*$ . Since  $B$  is closed under finite suprema, the topological base is closed under finite intersections. It suffices to ascertain that  $\bigsqcup_{n \in \mathbb{N}^*} \nu_n(\uparrow b_k) = \nu^*(\uparrow b_k)$  for each  $b_k \in B$ .

For each  $n \in \mathbb{N}^*$ ,

$$\nu_n(\uparrow b_k) = \sum_{i=1}^{K_n} \nu^* \left( \uparrow a_i^n \setminus \bigcup_{l=1}^{i-1} (\uparrow a_l^n) \right) \delta_{a_i^n}(\uparrow b_k) = \sum_{i: b_k \ll a_i^n} \nu^* \left( \uparrow a_i^n \setminus \bigcup_{l=1}^{i-1} (\uparrow a_l^n) \right)$$

$$\stackrel{\text{countable additivity}}{=} \nu^* \left( \bigcup_{i: b_k \ll a_i^n} \left[ \uparrow a_i^n \setminus \bigcup_{l=1}^{i-1} (\uparrow a_l^n) \right] \right) = \nu^*(B_n),$$

where  $B_n = \bigcup \{ \uparrow b_i \mid i \in \mathbb{N}^0, b_i = \bigsqcup_{j \in J} b_j \text{ for some } J \subseteq \{1, \dots, n\}, b_k \ll b_i \}$ , since, for  $n \in \mathbb{N}^*$ ,  $a_1^n, \dots, a_{K_n}^n$  are defined as the finite suprema of  $b_1, \dots, b_n$ .



## Some details of the proof: convergence (ii)

Then

$$\left( \bigsqcup_{n \in \mathbb{N}^*} \nu_n \right) (\uparrow b_k) = \lim_{n \rightarrow \infty} \nu^*(B_n) = \nu^* \left( \bigcup_{n=1}^{\infty} B_n \right),$$

the last equality following from the continuity of measures from below.

By the interpolation property of continuous dcpos,

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{\substack{i=1, \\ b_k \ll b_i}}^{\infty} \uparrow b_i = \uparrow b_k,$$

so  $(\bigsqcup_{n \in \mathbb{N}^*} \nu_n) (\uparrow b_k) = \nu^*(\uparrow b_k)$ , and the result follows.

## Finitary approximation of a given stochastic process

We can think of the approximation of measures at the top (including the laws of stochastic processes) as a special case of this construction. Note that the bounded complete domain  $[T \rightarrow \mathbf{IS}]$ , with  $T = [0, 1]$ ,  $S = \mathbb{R}$ , has a countable basis closed under finite suprema. It is given by the step functions obtained from rational-valued intervals. We can therefore think about the valuations  $v_n$  as partial stochastic processes, which approximate and generate the law of the stochastic process,  $\mu$ , in the limit. Also, by choosing  $T$  to be a finite or countable set, we can treat the discrete time partial stochastic processes as a special case of the present construction.

## Computable stochastic processes

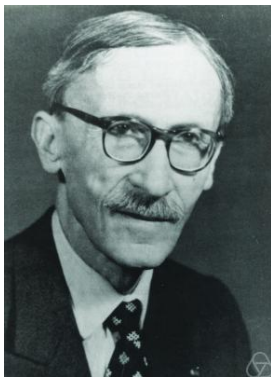
- An increasing chain of simple valuations  $\nu_0 \sqsubseteq \nu_1 \sqsubseteq \nu_2 \sqsubseteq \dots$ , where, for each  $i \in \mathbb{N}$ ,  $\nu_i = \sum_{j=1}^{n_i} r_{ij} \delta_{g_{ij}}$ , is **effective** if for each  $i$ ,  $n_i \in \mathbb{N}$  is recursively given,  $r_{i1}, \dots, r_{in_i}$  are computable, and  $g_{i1}, \dots, g_{in_i}$  are effectively given
- A stochastic process is **(domain-theoretically) computable** if there exists a total recursive function  $\phi : \mathbb{N} \rightarrow \mathbb{N}$  such that, for each  $i \in \mathbb{N}$ , gives  $N := \phi(i)$  in (1)

## Some closure properties of sets with computable measure

Given a measure  $\mu$ , let  $\mathcal{A}$  be a collection of  $\mu$ -measurable sets that is closed under finite intersections and such that the measure  $\mu(A)$  of each  $A \in \mathcal{A}$  is a computable real number. Then the following are also computable real numbers:

- $\mu(\bigcup_{i=1}^n A_i)$  for each  $n \in \mathbb{N}^*$ ,  $A_1, \dots, A_n \in \mathcal{A}$
- $\mu(A_1 \setminus A_2)$  for  $A_1, A_2 \in \mathcal{A}$
- $\mu(A \setminus (\bigcup_{i=1}^n A_i))$  for each  $n \in \mathbb{N}^*$ ,  $A_1, \dots, A_n \in \mathcal{A}$

## History: Paul Pierre Lévy (1886–1971)



Remarkably, **Paul Pierre Lévy** (who is known for, among many other things, one of the constructions of the Brownian motion) has contributed to domain theory — back in 1965 [Lév65] — even though he wasn't aware of its existence!

## Paul Lévy's formula (i)

- Let  $T = [0, 1]$ ,  $t \in T$ ,  $m_t := \min_{0 \leq s \leq t} W_s$ ,  $M_t := \max_{0 \leq s \leq t} W_s$
- The joint distribution of the processes  $W_t$ ,  $m_t$ ,  $M_t$  is given by

$$\mathbb{P}[a < m_t \leq M_t < b \text{ and } W_t \in A] = \int_A k(y) dy$$

- Here  $A \subseteq \mathbb{R}$  is a measurable set,

$$k(y) := \sum_{n=-\infty}^{\infty} p_t(2n(b-a), y) - p_t(2a, 2n(b-a) + y), \quad (3)$$

and

$$p_t(x, y) := \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/(2t)}$$

## Paul Lévy's formula (ii)

- It is convenient to regard this equation as a special case of the following function of two variables,  $x \in (a, b)$  and  $y \in (a - x, b - x) \subseteq (a - b, b - a)$ :

$$k(x, y) := \sum_{n=-\infty}^{\infty} p_t(2n(b - a), y) - p_t(2(a - x), 2n(b - a) + y)$$

- In (3),  $x$  is 0
- By introducing  $x$  we are effectively allowing the Brownian motion an intercept from the origin
- To make the dependence on  $a$ ,  $b$ , and  $t$  explicit, we shall write  $k(x, y; a, b; t)$

## Domain-theoretic approximation of Wiener measure (i)

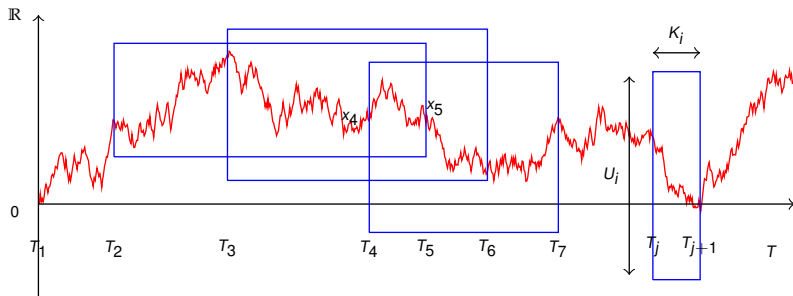
- Let  $V := V(K_1, \dots, K_n; U_1, \dots, U_n)$ ,  $n \in \mathbb{N}^*$  be a basic open set
- In our context, where  $X$  will be a nonempty compact interval,  $X \subseteq \mathbb{R}$ , the basic open set  $V \subseteq C(X, Y)$  induces a partition of  $X$ :

$$\mathcal{T}(V) := \{\min X, \max X\} \cup \bigcup_{i=1}^n \{\min K_i, \max K_i\}$$

- Regard it as a naturally ordered (in ascending order) tuple containing  $|\mathcal{T}(V)| \leq 2(n+1)$  (distinct) elements and refer to its elements as  $T_1, \dots, T_{|\mathcal{T}|}$ , where the dependence on  $V$  is implicit



## Illustration of a basic open set



## Domain-theoretic approximation of Wiener measure (ii)

- For  $i = 1, \dots, |\mathcal{T}| - 1$ , define

$$f_i(x, y) := \begin{cases} k(x, y; L_i, R_i; \Delta t_i) & \text{if } [T_i, T_{i+1}] \subseteq \bigcup_{j=1}^n K_j, \\ \frac{1}{\sqrt{\Delta t_i}} \phi\left(\frac{y-x}{\sqrt{\Delta t_i}}\right) & \text{otherwise,} \end{cases}$$

where  $\phi$  is the standard normal density function,  $\Delta t_i = T_{i+1} - T_i$ ,  
 $[L_i, R_i] := \bigcap_{j=1}^n \{U_j \mid [T_i, T_{i+1}] \subseteq K_j\}$ .

- Using the properties of conditional probability,

$$\begin{aligned} \mu_W(V) &= \int_{A_1} \int_{A_2} \dots \int_{A_{|\mathcal{T}|-1}} f_1(x_0, x_1) f_2(x_1, x_2) \dots \\ &\quad f_{|\mathcal{T}|-1}(x_{|\mathcal{T}|-2}, x_{|\mathcal{T}|-1}) dx_1 dx_2 \dots dx_{|\mathcal{T}|-1} \end{aligned}$$

where  $x_0 = 0$ , and, for  $i = 1, \dots, |\mathcal{T}| - 1$ ,

$$A_i := \bigcap_{j=1}^n \{U_j \mid T_{i+1} \in K_j\} - x_{i-1}$$

## Càdlàg processes

- While many (most?) processes studied in mathematical finance and other applied fields are continuous, some aren't. Of particular interest are **càdlàg processes** (“continue à droite, limite à gauche”), which admit jumps [CT03]. The behaviour of the markets on Monday (“Mad market Monday” according to Reuters) is a good example!
- A function  $f : [0, 1] \rightarrow \mathbb{R}$  is called a **càdlàg function** if, for every  $t \in [0, 1]$ , the left limit  $f(t-) := \lim_{s \uparrow t} f(s)$  exists; and the right limit  $f(t+) := \lim_{s \downarrow t} f(s)$  exists and equals  $f(t)$
- Anatoliy Volodymyrovych Skorokhod** (1930–2011) introduced a topology — the **Skorokhod topology** — on the space,  $D([0, 1], \mathbb{R})$ , of càdlàg functions to study the convergence in distribution of stochastic processes with jumps as an alternative to the compact-open topology. It is induced [Bil99] by the following metric, which makes  $D([0, 1], \mathbb{R})$  a complete separable metric space. Let  $\Lambda$  be the class of strictly increasing continuous mappings of  $[0, 1]$  onto itself. For  $\lambda \in \Lambda$  one defines

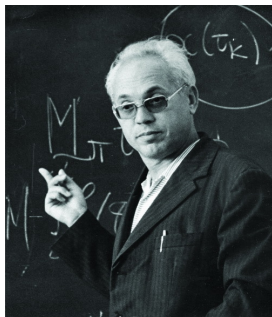
$$\|\lambda\| = \sup_{s \neq t} \left| \ln \frac{\lambda(t) - \lambda(s)}{t - s} \right|.$$

We can then define the metric as

$$d(f, g) = \inf_{\lambda \in \Lambda} \left\{ \sup |f(t) - g(\lambda(t))| + \|\lambda\| \right\}.$$

- How can one relate this to the Scott topology or other domain-theoretic topologies?

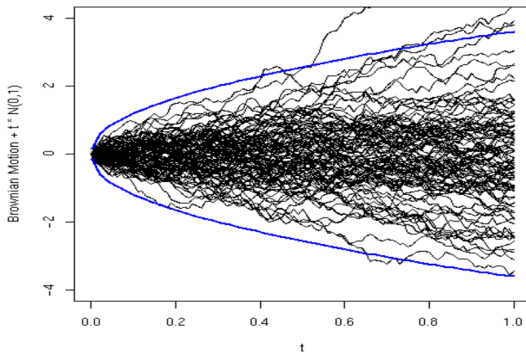
## History: Anatoliy Volodymyrovich Skorokhod (1930–2011)



Among **Anatoliy Volodymyrovich Skorokhod**'s contributions to the theory of stochastic and Markov processes, his topologies have been instrumental in the study of jump behaviour

## Computational considerations

What is the best order of enumeration of  $B := (b_1, b_2, \dots)$  to obtain a good rate of convergence?



## Stochastic integration

- A generalisation of the **Riemann-Stieltjes integral**. The integrands and the integrators are stochastic processes, as are the integrals themselves:

$$Y_t = \int_0^t H_s dX_s,$$

where  $H$  is a locally square-integrable process adapted to the filtration generated by the semimartingale  $X$

- More often than not,  $X$  is  $W$
- **Ito** integral:

$$\int_0^t H_s dX_s = \lim_{n \rightarrow \infty} \sum_{k=0}^{nt} H_{k/n} (X_{(k+1)/n} - X_{k/n})$$

- **Stratonovich** integral:

$$\int_0^t H_s dX_s = \lim_{n \rightarrow \infty} \sum_{k=0}^{nt} H_{(k+1)/n} (X_{(k+1)/n} - X_{k/n})$$

## The Wiener integral

- The **Wiener integral** is a Lebesgue integral over sets in an infinite-dimensional function space, such as  $C := C(T, \mathbb{R})$ , of functionals defined on these sets. Let  $F$  be a functional defined on  $C$  that is measurable with respect to the Wiener measure,  $\mu_W$ . Then the Wiener integral is the Lebesgue integral

$$\int_C F(x) d\mu_W(x)$$

- Let  $x = x(t) \in C$ ,  $n \in \mathbb{N}^*$ , and  $t_1, \dots, t_n \in T$ . Denote by  $x^{(n)}$  the broken line with vertices at the points  $(t_1, x(t_1)), \dots, (t_n, x(t_n))$ . Let  $F$  be a functional on  $C$ . For  $n \rightarrow \infty$ ,  $F(x^{(n)}) \rightarrow F(x)$  in the sense of strong convergence
- [Kov63] If  $F$  is a **continuous bounded functional**,

$$\int_C F(x) d\mu_W(x) = \lim_{n \rightarrow \infty} \frac{1}{\pi^n t_1 (t_2 - t_1) \dots (t_n - t_{n-1})} \times$$

$$\int_{\mathbb{R}^n} F_n(x_1, \dots, x_n) \exp \left[ -\frac{x_1^2}{t_1} - \sum_{j=1}^{n-1} \frac{(x_{j+1} - x_j)^2}{t_{j+1} - t_j} \right] dx_1 \dots dx_n,$$

where  $F_n(x_1, \dots, x_n) := F(x^{(n)})$

## The Wiener integral and the Feynman path integrals

- **Analytical continuation:** Consider the Wiener measure with covariance  $\lambda \in \mathbb{R}^+$ , a functional  $F$  on  $C([0, 1], \mathbb{R})$ . The following holds:

$$\int_{C([0,1],\mathbb{R})} F(\omega) dW_\lambda(\omega) = \int_{C([0,1],\mathbb{R})} F(\sqrt{\lambda}\omega) dW(\omega).$$

- What if  $\lambda$  is complex? The left-hand side is meaningless, whereas the right-hand side is OK if  $F$  is suitably analytical and measurable. When  $\lambda = i$ , we get the **analytically-continued Wiener integral**.
- In particular, we can apply this to the Feynman path integral representation of the **Schrödinger equation**. Consider the heat equation with potential  $V$

$$-\frac{\partial}{\partial t} u(t, x) = -\frac{1}{2} \Delta_x u(t, x) + V(x)u(t, x), \quad x \in \mathbb{R}^d.$$

The solution in terms of a Wiener integral is given by the **Feynman-Kac formula**:

$$u(t, x) = \int_{C([0,1],\mathbb{R})} e^{-\int_0^t V(\omega(s)+x) ds} u(0, \omega(t) + x) dW(\omega)$$

- This formula works for many  $V$  of interest



## Edalat integration

- **Edalat integration** was introduced in [Eda95a] for bounded real-valued functions on compact metric spaces embedded into continuous domains (i.e. spaces of maximal points), and bounded Borel measures on those compact metric spaces
- Extended to locally compact spaces by **Edalat** and **Sara Negri** [EN98]
- Extended to bounded real-valued functions on Hausdorff spaces embedded into continuous domains by **John D. Howroyd** [How00]. This extension is applicable in our setting, as  $C$  is a Hausdorff space

## Howroyd's extension of Edalat integration (i)

- Let  $X \leftrightarrow \text{Max}(D) \hookrightarrow D$  be a dense embedding of  $X$  into the maximal points of a continuous domain  $D$  equipped with the Scott topology
- Let  $f : X \rightarrow \mathbb{R}$  be a bounded function
- Let  $\mu$  be a Borel probability measure on  $X$  such that  $\bar{\mu}(U) := \mu(U \cap X)$  defines a continuous valuation on the Scott open sets of  $D$

## Howroyd's extension of Edalat integration (ii)

- [LL03] Let  $\nu = \sum_{b \in |\nu|} r_b \mu_b \in \mathbf{P}^1(D)$  be a simple valuation where  $|\nu|$  is the support of  $\nu$  and  $\mu_b$  is a point valuation for  $b \in D$
- Then the lower sum and upper sum of  $f$  w.r.t.  $\nu$  are defined as

$$S^l(f, \nu) = \sum_{b \in |\nu|} r_b \inf f(\uparrow b \cap X),$$

and

$$S^u(f, \nu) = \sum_{b \in |\nu|} r_b \sup f(\uparrow b \cap X),$$

respectively

## Howroyd's extension of Edalat integration (iii)

- The lower E-integral and upper E-integral of  $f$  w.r.t.  $\bar{\mu}$  are defined as

$$E- \int_* f d\mu = \sup\{S^l(f, \nu) : \nu \ll \bar{\mu}, \nu \text{ simple}\},$$

and

$$E- \int^* f d\mu = \inf\{S^u(f, \nu) : \nu \ll \bar{\mu}, \nu \text{ simple}\},$$

respectively

- The bounded function  $f : X \rightarrow \mathbb{R}$  is said to be **E-integrable** w.r.t.  $\mu$  if

$$E- \int^* f d\mu = E- \int_* f d\mu$$

- If  $f$  is E-integrable, the **E-integral** of  $f$  is denoted by  $E- \int f d\mu$  and is defined to be the value of the lower or upper integral:

$$E- \int f d\mu = E- \int^* f d\mu = E- \int_* f d\mu$$

## The case of the (partial) Wiener measure

- In our case:
  - $X = C = C(T, \mathbb{R})$
  - $D = [T \rightarrow \mathbb{R}]$
  - $\mu = \mu_W$
  - $f = F$ , a bounded functional on  $C$
- [How00, Theorem 13] If a function is E-integrable then it is Lebesgue integrable, and the values of the integrals agree

## A word to Ray Solomonoff



Ray Solomonoff (1926–2009)

A very *conventional* scientist understands his science using a *single 'current paradigm'* — the way of understanding that is most in vogue at the present time. A more *creative* scientist understands his science in *very many ways*, and can more easily create new theories, new ways of understanding, when the 'current paradigm' no longer fits the current data. [Sol09]



Samson Abramsky and Achim Jung.

Domain theory.

In *Handbook of logic in computer science (vol. 3)*, pages 641–761. Oxford University Press, Inc., 1994.



Mauricio Alvarez-Manilla, Abbas Edalat, and Nasser Saheb-Djahromi.

An extension result for continuous valuations.

*Journal of the London Mathematical Society*, 61(2):629–640, April 2000.



Alexander Bain.

*Mind and Body. The Theories of Their Relation.*

D. Appleton and Company, 1873.



G.M. Birtwistle, Ole-Johan Dahl, Bjørn Myhrhaug, and Kristen Nygaard.

*Simula Begin.*

Van Nostrand Reinhold, 1973.



Paul Bilokon and Abbas Edalat.

A domain-theoretic approach to Brownian motion and general continuous stochastic processes.

In *CSL-LICS '14 Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, 2014.



Paul Bilokon and Abbas Edalat.

A domain-theoretic approach to Brownian motion and general continuous stochastic processes.

*Theoretical Computer Science*, 691:10–26, 2017.



Patrick Billingsley.

*Convergence of Probability Measures*.

Wiley Series in Probability and Statistics. Wiley-Interscience P, 2 edition, 1999.



Christopher M. Bishop.

*Pattern Recognition and Machine Learning*.

Information Science and Statistics. Springer, 2007.



Jens Blanck.

Domain representations of topological spaces.

*Theoretical Computer Science*, 247:229–255, 2000.



Alonzo Church.

*The Calculi of Lambda-Conversion*.

Annals of Mathematical Studies. Princeton University Press, 1941.



Cornell Aeronautical Laboratory, Inc.

Mark I Perceptron operator's manual (project PARA).

Report VG-1196-G-5, Cornell Aeronautical Laboratory, Inc., February 1960.



Daniel Crevier.

*AI: The Tumultuous Search for Artificial Intelligence*.

BasicBooks, 1993.





Rama Cont and Peter Tankov.  
*Financial Modelling With Jump Processes.*  
Financial Mathematics Series. Chapman & Hall, 2003.



Hubert L. Dreyfus and Stuart E. Dreyfus.  
Making a mind versus modeling the brain: AI at a crossroads.  
*Daedalus*, 117, 1988.



Henning Dekant.  
Out of the AI winter and into the cold.  
[https://wavewatching.net/2013/08/12/  
out-of-the-ai-winter-and-into-the-cold/](https://wavewatching.net/2013/08/12/out-of-the-ai-winter-and-into-the-cold/), August 2013.



George Davie and Willem L. Fouché.  
On the computability of a construction of Brownian motion.  
*Mathematical Structures in Computer Science*, 23(6):1257–1265, December 2013.



Armen Der Kiureghian and Ove Ditlevsen.  
Aleatory or epistemic? does it matter?  
In *Special Workshop on Risk Acceptance and Risk Communication*. Stanford  
University, March 2007.



Ole-Johan Dahl, Bjørn Myrhaug, and Kristen Nygaard.  
*Common base language.*  
Norwegian Computing Centre, 1970.



Abbas Edalat.

Domain theory and integration.

*Theoretical Computer Science*, 151(1):163–193, November 1995.



Abbas Edalat.

Dynamical systems, measures, and fractals via domain theory.

*Information and Computation*, 120(1):32–48, July 1995.



Abbas Edalat.

When Scott is weak on the top.

*Mathematical Structures in Computer Science*, 7(5):401–417, October 1997.



Thomas Erker, Martín Escardó, and Klaus Keimel.

The way-below relation of function spaces over semantic domains.

*Topology and its Applications*, 89(1–2):61–74, November 1998.



Abbas Edalat and Sara Negri.

The generalized Riemann integral on locally compact spaces.

*Topology and Its Applications*, 89(1–2):121–150, November 1998.



Abbas Edalat and Philipp Sünderhauf.

A domain-theoretic approach to computability on the real line.

*Theoretical Computer Science*, 210(1):73–98, January 1998.



Belmont G. Farley and Wesley A. Clark.

Simulation of self-organizing systems by digital computer.

*IRE Transactions on Information Theory*, 4(4):76–84, 1954.



Scott Ferson and Lev R. Ginzburg.

Different methods are needed to propagate ignorance and variability.  
*Reliability*, 54:133–144, 1996.



Willem L. Fouché.

Arithmetical representations of Brownian motion I.  
*The Journal of Symbolic Logic*, 65(1):421–442, March 2000.



Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

*Deep Learning*.  
Adaptive Computation and Machine Learning. MIT Press, 2017.



Gerhard Gierz, Karl Heinrich Hofmann, Klaus Keimel, Jimmie D. Lawson, Michael Mislove, and Dana Stewart Scott.

*Continuous Lattices and Domains*.  
Number 93 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2003.



Jean Goubault-Larrecq.

Extensions of valuations.  
*Mathematical Structures in Computer Science*, 15(2):271–297, 2005.



Jean Goubault-Larrecq.

*Non-Hausdorff Topology and Domain Theory: Selected Topics in Point-Set Topology*,  
volume 22 of *New Mathematical Monographs*.  
Cambridge University Press, 2013.



Stephen Grossberg.

Adaptive pattern classification and universal recording, i: Parallel development and coding of neural feature detectors.

*Biological Cybernetics*, 23:121–134, 1976.



Stephen Grossberg.

Adaptive pattern classification and universal recording, ii: Feedback, expectation, olfaction, and illusions.

*Biological Cybernetics*, 23:187–202, 1976.



Donald Olding Hebb.

*The Organization of Behavior: A Neuropsychological Theory*.

Wiley and Sons, 1949.



Geoffrey E. Hinton.

Deterministic Boltzmann learning performs steepest descent in weight-space.

*Neural computation*, 1(1):143–150, 1989.



Geoffrey E. Hinton.

Learning multiple layers of representation.

*Trends in Cognitive Sciences*, 11(10):428–434, October 2007.



John J. Hopfield.

Neural networks and physical systems with emergent collective computational abilities.

*Proceedings of the National Academy of Sciences of the United States of America: Biophysics*, 79:2554–2558, April 1982.



Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh.

A fast learning algorithm for deep belief nets.

*Neural Computation*, 18(7):1527–1554, 2006.



J. D. Howroyd.

A domain-theoretic approach to integration in Hausdorff spaces.

*London Mathematical Society Journal of Computation and Mathematics*, 3:229–273, August 2000.



Tony Hey and Gyuri Pápay.

*The computing universe: a journey through a revolution.*

Cambridge University Press, 2015.



Geoffrey E. Hinton and Ruslan R. Salakhutdinov.

Reducing the dimensionality of data with neural networks.

*Science*, 313:504–507, 2006.



Trevor Hastie, Robert Tibshirani, and Jerome Friedman.

*The Elements of Statistical Learning: Data Mining, Inference, and Prediction.*

Springer Series in Statistics. Springer, 2 edition, 2011.



William James.

*The Principles of Psychology.*

Henry Holt and Company, 1890.



Frédéric De Jaeger, Martín Escardó, and Gabriele Santini.

On the computational content of the Lawson topology.

*Theoretical Computer Science*, 357(1):230–240, July 2006.



Claire Jones.

*Probabilistic Non-determinism*.

PhD thesis, University of Edinburgh, July 1990.

Supervised by Gordon D. Plotkin.



Claire Jones and Gordon D. Plotkin.

A probabilistic powerdomain of evaluations.

*In Proceedings of the Fourth Annual Symposium on Logic in Computer Science (LICS)*, June 1989.



Stephen Cole Kleene.

Representation of events in nerve nets and finite automata.

*Annals of Mathematics Studies*, 34:3–41, 1956.



Teuvo Kohonen.

Correlation matrix memories.

*IEEE Transactions on Computers*, 21(4):353–359, April 1972.



Teuvo Kohonen.

Self-organization of topologically correct feature maps.

*Biological Cybernetics*, 43:59–69, 1982.



Teuvo Kohonen.

*Self-Organization and Associative Memory*.

Springer-Verlag, 2 edition, 1988.



Andrey Nikolayevich Kolmogorov.

On the interpretation of intuitionistic logic.

*Mathematische Zeitschrift*, 35:58–65, 1932.



Andrey Nikolaevich Kolmogorov.

Grundbegriffe der Wahrscheinlichkeitsrechnung.

*Ergebnisse der Mathematik und ihrer Grenzgebiete*, 2(3):1–62, 1933.



Andrey Nikolaevich Kolmogorov.

O ponyatii algoritma.

*Uspehi matematicheskikh nauk*, 5(4):175–176, 1953.



Andrey Nikolaevich Kolmogorov.

*Foundations of the Theory of Probability*.

Chelsea Publishing Company, 1956.



I.M. Kovalchik.

The Wiener integral.

*Russian Mathematical Surveys: Uspekhi Mat. Nauk*, 18(1):97–134, 1963.



Ioannis Karatzas and Steven E. Shreve.

*Brownian Motion and Stochastic Calculus*.

Graduate Texts in Mathematics. Springer, 2 edition, 1991.



Jimmie D. Lawson.

Computation on metric spaces via domain theory.

*Topology and its Applications*, 85:247–263, 1998.



Yann LeCun.

A theoretical framework for backpropagation.

In D. Touretzky, G. Hinton, and T. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, San Mateo, CA, June 1988. Morgan Kaufmann.



Paul Lévy.

*Processus stochastiques et mouvement brownien*.

Paris, Gauthier-Villars, 2 edition, 1965.



Jimmie D. Lawson and Bin Lu.

Riemann and Edalat integration on domains.

*Theoretical Computer Science*, 305:259–275, 2003.



John McCarthy.

Programs with common sense.

In *Symposium on the Machanization of Thought Processes*, Teddington, England, November 1958. National Physical Laboratory.



John McCarthy.

Recursive functions of symbolic expressions and their computation by machine, part I.

*Communications of the ACM*, pages 184–195, 1960.



David A. Medler.

A brief history of connectionism.

*Neural Computing Surveys*, 1:61–101, 1998.





[Marvin Minsky.](#)

*The psychology of computer vision*, chapter A framework for representing knowledge, pages 211–277.

McGraw-Hill, 1975.



[Graem A. Ringwood Matthew M. Huntbach.](#)

*Agent-Oriented Programming: From Prolog to Guarded Definite Clauses*, volume 1630 of *Lecture Notes in Artificial Intelligence*.

Springer, 1998.



[Warren McCulloch and Walter Pitts.](#)

A logical calculus of ideas immanent in nervous activity.

*Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.



[Marvin Minsky and Seymour Papert.](#)

*Perceptrons: An introduction to computational geometry*.

MIT Press, 1969.



[Marvin Minsky and Seymour Papert.](#)

Progress report on artificial intelligence.

Technical report, Massachusetts Institute of Technology, December 1971.



[Allen Newell and J. Cliff Shaw.](#)

Programming the logic theory machine.

In *Western Joint Computer Conference*, February 1957.



[Brent Øksendal.](#)

*Stochastic Differential Equations: An Introduction with Applications.*  
Universitext. Springer, 6 edition, 2010.



D. B. Parker.

Learning-logic: Casting the cortex of the human brain in silicon.  
Technical Report TR-47, Center for Computational Research in Economics and  
Management Science, MIT, Cambridge, MA, April 1985.



Judea Pearl.

*Probabilistic reasoning in intelligent systems: networks of plausible inference.*  
Morgan Kaufmann, 1988.



Kevin L. Priddy and Paul E. Keller.

*Artificial Neural Networks: An Introduction.*  
The International Society for Optical Engineering (SPIE) Press, 2005.



Jordan B. Pollack.

No harm intended: A review of the perceptrons expanded edition.  
*Journal of Mathematical Psychology*, 33(3):358–365, 1989.



Nathaniel Rochester, John H. Holland, L. H. Haibt, and W. L. Duda.

Tests on a cell assembly theory of the action of the brain, using a large digital  
computer.  
*IRE Transactions on Information Theory*, 2(3):80–93, 1956.



David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams.

Learning representations by back-propagating errors.

*Nature*, 323:533–536, October 1986.



**Richard H. Richens.**

Preprogramming for mechanical translation.

*Mechanical Translation*, 3(1):20–25, 1956.



**Frank Rosenblatt.**

The perceptron — a perceiving and recognizing automaton.

Report 85-460-1, Cornell Aeronautical Laboratory, 1957.



**Frank Rosenblatt.**

On the convergence of reinforcement procedures in simple perceptrons.

Report VG-1196-G-4, Cornell Aeronautical Laboratory, Buffalo, New York, February 1960.



**David E. Rumelhart and David Zipser.**

Feature discovery by competitive learning.

*Cognitive Science, A Multidisciplinary Journal*, 9(1):75–112, January 1985.



**Dana Stewart Scott.**

Outline of a mathematical theory of computation.

*In Proceedings of the Fourth Annual Princeton Conference on Information Sciences and Systems*, 1970.



**Dana Stewart Scott.**

Outline of a mathematical theory of computation (technical monograph prg-2).

In *Oxford University Computing Laboratory Programming Research Group Technical Monographs*, 1970.



**Nasser Saheb-Djahromi.**

Cpo's of measures for nondeterminism.

*Theoretical Computer Science*, 12:19–37, 1980.



**John Rogers Searle.**

The intentionality of intention and action.

*Cognitive Science*, 4:47–70, 1980.



**Steven E. Shreve.**

*Stochastic Calculus for Finance. Volume II: Continuous-Time Models.*

Springer-Verlag, 2004.



**M. B. Smyth.**

Topology.

In *Handbook of logic in computer science (vol. 1): background: mathematical structures*, pages 641–761. Oxford University Press, Inc., 1992.



**Ray Solomonoff.**

*Information theory and statistical learning*, chapter Algorithmic probability, theory and applications, page 11.

Springer Science and Business Media, 2009.



**Paul John Werbos.**

*Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.*

PhD thesis, Harvard University, 1974.



**Paul John Werbos.**

Backpropagation through time: what it does and how to do it.

In 78, editor, *Proceedings of the IEEE*, volume 10, pages 1550–1560, October 1990.



**Paul John Werbos.**

*The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting.*

Wiley-Interscience, 1994.



**Bernard Widrow and Ted Hoff.**

Adaptive switching circuits.

In *IRE Western Electric Show and Convention Record*, volume 4, pages 96–104, August 1960.



**Bernard Widrow.**

*Self-Organizing Systems*, chapter Generalization and information storage in networks of adaline “neurons”, pages 435–461.

Spartan Books, Washington, DC, 1962.



**Norbert Wiener.**

*A Comparison Between the Treatment of Algebra of Relatives By Schroeder and That By Whitehead and Russell.*

PhD thesis, Harvard University, 1913.



Norbert Wiener.

Differential space.

*Journal of Mathematical Physics*, 2:131–174, 1923.



David Williams.

*Weighing the Odds: A Course in Probability and Statistics*.

Cambridge University Press, 2001.



Terry Winograd.

Procedures as a representation for data in a computer program for understanding natural language.

MIT AI Technical Report 235, Massachusetts Institute of Technology, February 1971.



Terry Winograd.

Procedures as a representation for data in a computer program for understanding natural language.

*Cognitive Psychology*, 3(1), 1972.



Bernard Widrow and Michael A. Lehr.

30 years of adaptive neural networks: Perceptron, madaline, and backpropagation.

*Proceedings of IEEE*, pages 1415–1442, September 1990.



Bernard Widrow and Michael A. Lehr.

Backpropagation and its applications.

In *Proceedings of the INNS Summer Workshop on Neural Network Computing for the Electric Power Industry*, pages 21–29, Stanford, August 1992.



Alfred North Whitehead and Bertrand Russell.

*Principia Mathematica*.

Cambridge University Press, 1910.