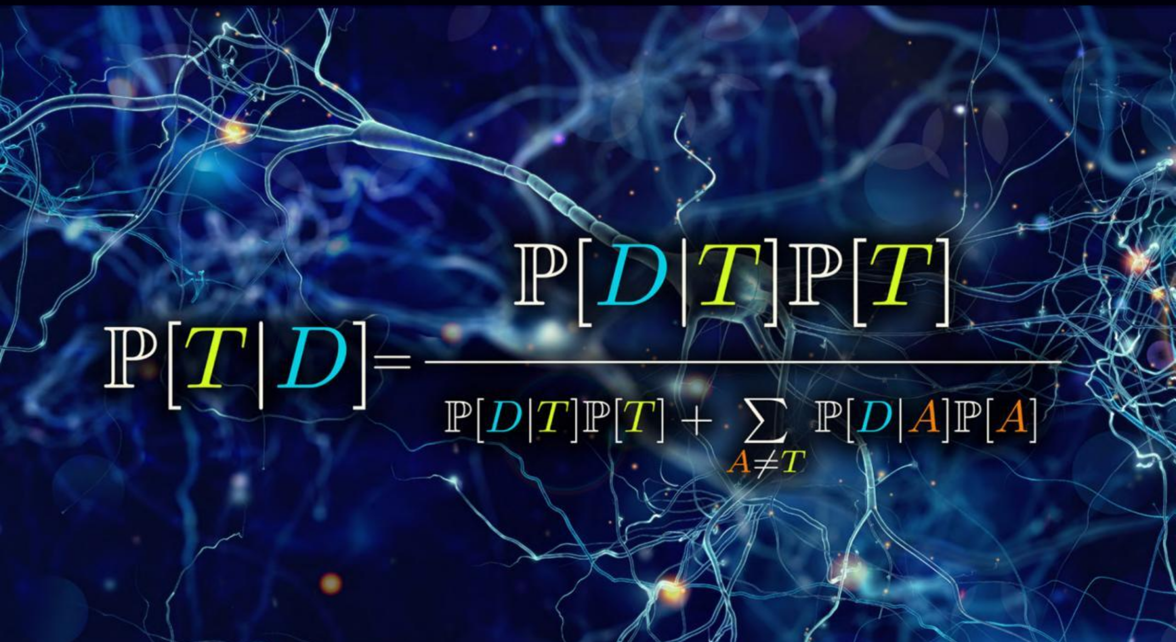
The Equation of Knowledge

From Bayes' Rule to a Unified Philosophy of Science

Lê Nguyên Hoang



$$\mathbb{P}[T|D] = \frac{\mathbb{P}[D|T]\mathbb{P}[T]}{\mathbb{P}[D|T]\mathbb{P}[T] + \sum_{A \neq T} \mathbb{P}[D|A]\mathbb{P}[A]}$$

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

First edition published 2020
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Library of Congress Cataloging-in-Publication Data
Names: Hoang, Lê Nguyên, author.
Title: The equation of knowledge : from Bayes' rule to a unified philosophy of science / Lê Nguyên Hoang.
Other titles: Formule du savoir. English.
Description: Boca Raton : C&H/CRC Press, 2020. | Translation of: La formule du savoir : une philosophie unifiée du savoir fondée sur le théorème de Bayes | Includes bibliographical references and index.
Identifiers: LCCN 2019060130 | ISBN 9780367428150 (hardback) | ISBN 9780367855307 (ebook)
Subjects: LCSH: Bayesian statistical decision theory. | Knowledge, Theory of. | Mathematics--Philosophy. | Science--Methodology.
Classification: LCC QA279.5 .H63 2020 | DDC 519.5/42--dc23
LC record available at https://lccn.loc.gov/2019060130

ISBN: 978-0-367-42815-0 (hbk)
ISBN: 978-0-367-85530-7 (ebk)

Typeset in LMRoman
by Nova Techset Private Limited, Bengaluru & Chennai, India

Contents

Foreword

Arriving in a small town, with a heavy suitcase, you head for the taxi station, where only one car is parked. Unfortunately, by the time you get close to it, a faster traveller has already taken it and disappeared before your eyes. What conclusion can you draw from this misadventure? That there seem to be taxis in this city - given its size, it was far from certain? And that thus, if you wait patiently, another taxi will eventually show up? Or should you conclude that one of the few taxis in the city has just driven away from you and that, given the size of the city, such a chance will not come back soon? These two interpretations are correct, but both depend on what you knew - or believed - before you got off the train.

The traveller who arrived in an unknown city, made hypotheses about the number of taxis and revised his hypotheses according to his observations is not very different from a baby who arrives in an unknown world, or from a researcher who, surprised by what others have been taking for granted, wonders why the sun rises every morning. Both explore the world, make assumptions, and revise them based on their observations.

What can we learn from our experiences? What can we know about the world? These are the questions that the magnificent book of Lê Nguyên Hoang invites us to examine.

On these questions, one point has been crystallizing the controversies for more than a century: Is it possible to associate a numerical value to a hypothesis that measures its likelihood? For some, such as Hans Reichenbach, this is the very purpose of probability theory. In particular, any observation that confirms a hypothesis increases its probability of being true: each observation of a black raven increases the probability that the hypothesis that all ravens are black is true. For others, such as Karl Popper, the assignment of a numerical value to such a hypothesis is an illusion. By observing a black raven, we can only conclude that our hypothesis that all ravens are black remains consistent with our observations.

At the heart of this controversy is a disconcertingly simple formula, Bayes' rule, “the equation of knowledge”, which gives its title to this book, and which allows computing the probability that we must attribute to a hypothesis after having made an observation - and thus makes Reichenbach right - but only on condition that we knew how to attribute a probability to this hypothesis, before making this observation - and thus makes Popper right.

If this question seemed clear-cut - in Popper's favour - in the 20th century, the evolution of data collection techniques is renewing it today. When we believed, in the 20th century, that there were white crows, we would interpret the fact that three observed crows are black as a coincidence. If we observe, today, a thousand, a million, or a billion crows, and if they are all black, it takes a certain courage - even a certain obstinacy - to claim that no, not all crows are black, and the agreement of our observations is only coincidence. At least we are forced to concede that there must be a large proportion of black crows among all crows, and probably even that white crows are the exception. This objection to Reichenbach's thesis, which constituted the problem of the hypotheses a priori highlighted by Bayes' rule, is now put into perspective by the flood of data. Other problems, however, are emerging: How were these data collected? Doesn't the collection method introduce bias or even discrimination against white crows? Once again, we see how technological developments, particularly in scientific instrumentation techniques, are changing the way questions are asked in the philosophy of science.

This is what makes Lê Nguyên Hoang's book so exciting. It has been written at the time of an upheaval, when technological developments are changing the way we look at Bayes' rule and its place in the edifice of knowledge.

It has also have been written at a time when communication techniques are changing the way we talk about science. Trained in the hard school of online videos, Lê Nguyên Hoang has found a new tone to talk about science, a tone that is both rigorous and narrative, where examples illuminate the most abstract questions.

Gilles Dowek
Research at Inria
Professor of the École Normale Supérieure de Paris Saclay

Acknowledgment

Writing this book was a formidable journey, with many ups and downs. I would not have been able to achieve it without the help, kindness, and wit of the numerous people I am highly indebted to.

I would like to thank Thibaut Giraud, Julien Fageot, Maxime Maillot, David Loureiro, and Marie Maury, whose detailed reviewing and feedback were extremely precious. More generally, I would like to thank all those who accompanied my Bayesian reflections or proof-read my writing, in particular Peva Blanchard, Rachid Guerraoui, El Mahdi El Mhamdi, Alexandre Maurer, Julien Stainer, Hadrien Hendrikx, Sébastien Rouault, Matej Pavlovic, Clément Hongler, Christophe Michel, Sébastien Carassou, Mithuna Yoganathan, and Zachary Stargensky, among many others. I am also greatly thankful to the École Polytechnique Fédérale of Lausanne, the publisher CRC Press, and Gilles Dowek for being kind enough to take the time to write the foreword of this book. I would also like to thank all of those who follow me or whom I follow on social media, who all greatly contributed to my everyday desire to explore the topic of this book. I am particularly thankful towards fellow science communicator, especially on YouTube.

But above all, I would like to thank you, dear reader. The thought of sharing my Bayesian adventures with you is a wonderful source of joy and motivation. Thank you.

Preface

Logic has long been regarded as the primary foundation of knowledge. It is often said that if logic proves some fact, then this fact necessarily holds. If logic says so, then we humans should believe so.

Yet resting knowledge upon the foundations of logic is arguably flawed. Indeed, logic only handles one kind of reasoning, called *deductive reasoning*. Deductive reasoning derives conclusions from hypotheses. But much of science is actually about figuring out the adequate hypotheses, given collected data. This is called *inductive reasoning*. Unfortunately, logic does not seem fit to address this equally fundamental type of reasoning. In particular, it is not clear how logic should exploit (messy) empirical data to infer knowledge.

There is another more fundamental flaw of logic. Logic restricts itself to true and false. While this dichotomy allowed brilliant advancements in mathematics and some fields like computer science and fundamental physics, it seems very limited to make predictions in more complex settings, such as those of biology, social sciences, and everyday decision-making.

To fix these flaws, a small but growing number of mathematicians, philosophers, and computer scientists have proposed to replace logic by some other foundation of knowledge. Namely, they proposed to rest knowledge upon the laws of probability. This dramatic epistemological revolution was eventually named *Bayesianism*, after Thomas Bayes, one of the key figures in the history of probability theory.

Amazingly, Bayesianism seems to generalize all the desirable features of logic, while avoiding the pitfalls caused by its dichotomic view on knowledge and proposing a compelling framework for reasoning. To this end, it argues that knowledge should be phrased in *probabilistic* terms. In Bayesianism, nothing is known for sure. Instead, everything is a matter of *credences*, that is, confidence levels measured by probabilities with values between 0 and 1.

As data pours in, Bayesianism imposes us to *update* our credences, depending on whether the empirical data fit our theories or contradict them. Crucially, this Bayesian update is rigorously determined by a fundamental equation known as *Bayes' rule*. It is this fascinating equation that this book is about. It is this equation that we shall refer to as the *equation of knowledge*.

Disturbingly, Bayesianism is far from being consensual within the science community. In fact, most scientists probably ignore the existence of this philosophy of knowledge. Worse, some even argue against Bayesianism. Their arguments, mostly based on the *subjectivity* of Bayesian approaches, have long seemed very compelling.

In fact, for much of the 20th century, Bayesianism was very much frowned upon. It was even considered unscientific by many leading statisticians. Nevertheless, over the last few decades, an impressive variety of fields, from social sciences to biology, meteorology, and astrophysics, have been relying more and more on so-called *Bayesian methods*, to construct more precise and more predictive models.

Perhaps more impressively, the rise of artificial intelligence through machine learning and massive data has led to a formidable gain of interest in Bayesian computations. Bayes' rule can be found at the heart of numerous state-of-the-art algorithms, such as the one that reconstructed the first image of a black hole[1]. According to Stanford's philosophy encyclopedia, the empirical successes and compelling theoretical foundations of Bayesianism have recently even made it consensual among philosophers as the right philosophy of confirmation[2].

Yet Bayesianism is not perfect. In fact, there is however, one extremely compelling argument against Bayesianism. Namely, Bayesianism requires unreasonable amounts of computation. In fact, computer scientist Ray Solomonoff proved that, in arguably its purest form, Bayesianism is in fact *incomputable*. While Solomonoff also showed that any computable philosophy of knowledge was necessarily flawed (or, more precisely, *incomplete*), the incomputability of Bayesianism seems like a definite reason to give up on pure Bayesianism[3].

[1] *First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole.* The Event Horizon Telescope Collaboration (2019).

[2] The article on "abduction" says: "In the past decade, Bayesian confirmation theory has firmly established itself as the dominant view on confirmation; currently one cannot very well discuss a confirmation-theoretic issue without making clear whether, and if so why, one's position on that issue deviates from standard Bayesian thinking."

[3] In fact, as Turing showed in 1936 through the famous *halting problem*, much of mathematical knowledge is also out of our reach because of incomputability.

This is in fact what much of this book is about. After explaining the building blocks of pure Bayesianism and defending the epistemological superiority of pure Bayesianism through numerous theoretical and historical arguments, I shall argue that today's big challenge in epistemology is the design and implementation of what may be called *pragmatic Bayesianism*.

This is the quest of tractable methods to allow both computers and human brains to perform good approximations of Bayes' rule. As we shall see, the key to do so is to combine adequately computer science with probability theory. This book will discuss numerous promising approaches to do so. It will also provide an extremely wide variety of examples from very diverse fields of knowledge, to train and test our Bayesian thinking.

But perhaps most importantly, in this book, I shall stress the elegance of the properties of *pure Bayesianism*, as well as the excitement of the quest for *pragmatic Bayesianism*. The search for the most reliable paths to knowledge is arguably one of the greatest joys of being human. It is a privilege for me to share this thrilling journey with you, dear reader.

I

Pure Bayesianism

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

CHAPTER 1

On A Transformative Journey

The theory of probabilities is basically just common sense reduced to calculus; it makes one appreciate with exactness that which accurate minds feel with a sort of instinct, often without being able to account for it. If we consider the analytical methods that the theory gave rise to, the truth of the principles it relies on, the subtle logic that demands its application to solving problems, the public utility goods that is built upon it, and the extensions it has received and can still receive, given its application to the most important questions of natural philosophy and political economics; if we then observe that even in things that cannot be reduced to computation, probability theory allows the most reliable insights to guide us in our judgment, and that it teaches us to steer away from the illusions that often mislead us; we shall see that there is no science more worthy of our meditations, and whose results are more useful.

Pierre-Simon Laplace (1749-1827)

1.1 STUMPED BY A STUDENT

At the end of a lecture in probability and statistics I was giving at the École Polytechnique of Montreal, a trolling student came to test me with a simple-looking puzzle. A man has two kids. At least one is a boy. What is the probability that the other is a boy too?

After a few seconds of thoughts, I successfully gave the right answer, which, as we shall see, is *not* $1/2$. The student acquiesced, and moved on to the next puzzle. Suppose you now learn that at least one of the kids is a boy born on a Tuesday. What is the probability that the other kid is a boy too?

This time, though, my answer was wrong. The student had stumped me.

The usual reflex is certainly to regard these two puzzles as mere mathematical games. Sure, there is a right answer. But that answer is only valid in a rigid and restricted mathematical setting. Solving these puzzles is useful in exercises or exams at school. But it's *only* mathematics.

Yet, the puzzle of the troll student is just an ultra-simplified version of many questions that we face in our daily lives. Should I believe a medical diagnosis? Is the presumption of innocence justified? Do judges racially discriminate? Is terrorism worrying? Can one generalize from one example? From a thousand? A million? Is the argument of authority worth anything? Are financial markets trustworthy? Are GMOs harmful? Why would science be *more* right than pseudosciences? Are robots about to conquer the world? Is capitalism wrong? Does God exist? What's good and what's bad?

For most, such questions have absolutely nothing to do with mathematics. And indeed, math alone is insufficient to address such questions. World hunger will not be solved by only proving theorems. Nevertheless, math likely has a lot to offer. It can help better structure our thinking, identify key challenges, and provide unexpected solutions. This is why many endeavours are more and more mathematized - including humanitary aid[1].

Despite the flourishing of mathematical models, it seems that most of us still want to distinguish the "real world" from academic courses that schools force us to take. In particular, the real world, it is often said, far transcends the framework of mathematics. As a result, mathematical theorems never seem to *really* apply to reality. How stupid must one be to think that mathematics has anything to say about the equality of rights[2]?

Sadly, rejecting the usefulness of mathematics is not merely a bad-student reflex. Even years after failing the troll student puzzle, I had not yet realized that my mathematical mishap revealed my inability to correctly reason about the real world. I had not understood that a better understanding of the puzzle would be key to better analyze my traveling friends' advice to plan my next trip - we'll get there.

[1] *A Set-Partitioning Formulation for Community Healthcare Network Design in Underserved Areas.* M Cherkesly, ME Rancourt & K Smilowitz (2017).

[2] *Measuring unfairness feeling in allocation problems.* Omega. LN Hoang, F Soumis & G Zaccour (2016).

1.2 MY PATH TOWARDS BAYESIANISM

Granted, I did solve the troll student puzzle later that day, after some obscure and mysterious computations. But it was only three years later, in early 2016, when I investigated the frequentist-Bayesian debate, that I really took the time to meditate about the puzzle. Most importantly, at last, I finally took it out of its confined mathematical setting.

In particular, for the three years that followed, nearly once a day, I kept thinking about the magical equation that solves this puzzle. To my greatest pleasure, this mysterious equation started to reveal its secrets to me. Slowly but surely, this brilliant equation was seducing me. I began to see it everywhere. Months after months, my mind got flooded with the sublime elegance of this untameable equation. It was too much. I *had* to write about it. And I *had* to do this well. This is how, towards the end of 2016, I began the writing of the book you have just started.

The untameable equation I am talking about is what I like to pompously call the *equation of knowledge*. But mathematicians, statisticians, and computer scientists better know it as Bayes' rule.

Bayes' rule is a mathematical theorem of remarkable simplicity. It's a compact equation, which is often taught in high school. It has a one-life proof, and only relies on multiplication, division, and the notion of probability. In particular, it seems vastly easier to learn than many other concepts in mathematics that high school and university students are asked to master.

And yet I'd claim that even the best mathematicians do not understand Bayes' rule - and there is even some mathematics that explains our inability to grasp this equation! More modestly, there is absolutely no doubt that *I* still do not understand Bayes' rule. Indeed, if I did, I would have immediately seen how the fact that at least one kid is a boy born on a Tuesday affects the likely gender of his sibling. I would have instantly given some relevant answer to the *troll* student. He would not have stumped me.

Over the last two years, I have been torturing my mind so as to never fail like this again. I want to know, understand, and feel Bayes' rule. I have already learned a lot, and I am still learning so much! I meditate on Bayes' rule almost every day, as if it were some sort of God I had to devote parts of my days to. And what a pleasure this is! Far from being a repetitive strain, these meditations have continuously fed my curiosity, as they have been discreetly whispering many of the unexpected implications of Bayes' rule. One after the other.

After long months of thinking, I ended up concluding that few ideas were as deep as Bayes' rule. I fell in love with Bayes' rule to the point where I now gladly claim that "rationality" essentially boils down to applying Bayes' rule - in which case no one is rational! This is the foundation of what might be call Bayesian philosophy, or *Bayesianism*.

1.3 A UNIFIED PHILOSOPHY OF KNOWLEDGE

Since I have not yet had the time to present Bayes' rule, for now, I will be intentionally vague about what Bayesianism is. But basically, if I had to sum it up in three clumsy phrases, I would give the following definition. Bayesianism supposes that any model, theory, or conception of "reality" is mere belief, fiction, or poetry; in particular, "all models are wrong". Empirical data must then force us to adjust the importance, or *credence*, that we assign to the different models. Crucially, the way credences are adjusted must obey Bayes' rule as rigorously as possible.

I have long rejected the relevancy of this philosophy of knowledge. It seems to discredit any concept of reality or truth, that many scientists cherish. Yet, it seems to perfectly fit what physics Nobel laureate Richard Feynman once said[3]: "I can live with doubt and uncertainty, and without knowing. I think it's much more interesting to live not knowing than to have answers that may be wrong. I have approximate answers, I have possible beliefs and different degrees of certainty about different things. But I am not absolutely sure of anything. And there are many things I don't know anything about. But I don't *have* to know an answer. I don't feel frightened by not knowing things."

You might fancy this viewpoint. Or you might want to reject altogether this approach to knowledge. Yet, before rejecting or adhering to Bayesianism, I can only encourage you to first take the time to meditate Bayes' rule and its consequences.

In this book, sadly, the main guide that I'll be has a very incomplete understanding of Bayes' rule. To help us in our thoughts, I will invoke a (female) fictitious character, the *pure Bayesian*, and we will try to imagine how this *pure Bayesian* behaves in different contexts. More than myself, it's this *pure Bayesian* that we shall put to the test. This is what we shall do again and again in this book. We shall repeat thought experiments which will be challenges that the *pure Bayesian* will have to face. And we shall carefully scrutinize, judge, and criticize the behaviour of the *pure Bayesian* - although these criticisms will often quickly turn into that of our intuition and of our relentless overconfidence.

[3] *The Feynman Series - Beauty*. Reid Gower (2011).

Now, the first Bayesian in history worthy of this name, the great Pierre-Simon Laplace, only had a partial description of the *pure Bayesian*. But over half a century ago, all computations, thoughts, and predictions of the *pure Bayesian* were rigorously described by the brilliant Ray Solomonoff. Unfortunately, as we shall discuss it in length, the *pure Bayesian* that Solomonoff described seems to necessarily violate the laws of physics (in particular the Church-Turing thesis[4]).

This forces us to restrict ourselves to some approximate Bayesianism, which I shall call *pragmatic*. *Pragmatic Bayesianism*, which differs from *pure Bayesianism* by its need of (fast) computability, will be incarnated by another fictitious (male) character, which I shall call the *pragmatic Bayesian*. Unfortunately (or not!), my description of the *pragmatic Bayesian* will be very incomplete, as pragmatic Bayesianism is still a huge and very open field of research - and it's not clear whether it can one day be fully closed.

As you are probably starting to guess, understanding the *pure Bayesian* and the *pragmatic Bayesian* is no easy task. To do so, we will have to discuss numerous fundamental concepts of mathematics, logic, statistics, computer science, artificial intelligence, and even notions of physics, biology, neuroscience, psychology, and economy. We will have to explain logarithms, contraposition, p-values, Solomonoff complexity, and neural networks, as well as entropy, Darwinian evolution, false memory, cognitive biases, and financial bubbles. What's more, we shall also invoke several cases from the history of science to test our two fictitious heroes.

I know. This is a lot to take in to understand Bayes' rule.

The good news is that I love explaining modern science - I have my own (French) YouTube channel called Science4All! Thus, rather than reading this book as a treatise in philosophy, I invite you to (also) read it as a science and mathematics popularisation book. In fact, on our way to Bayesianism, I will not hesitate to take some detours through the world of science, with the secret goal to tease you and make you want to find out more about scientific theories!

But let's get back to philosophy for now. As you can guess, I have surrendered to the appeals of Bayesianism. After long months of meditation, Bayesianism seduced me to the point where I felt the need to write about it. I kept being marvelled by the intelligence of the *pure Bayesian*. And I now aspire to resemble her more and more. Even long after the beginning of the writing of this book, I have kept discovering, again and

[4] *The Universal Turing Machine*. ZettaBytes, EPFL. R Guerraoui (2016).

again, the uncountable breathtaking wonders of what has since become my favorite mathematical equations of all.

When I started this book, I was an enthusiastic Bayesian. By now, I have become a convinced Bayesian. I would even call myself an extremist Bayesian, especially compared to others that call themselves Bayesians as well. But more importantly, I would like to become a *competent* Bayesian some day. I dream about the day I'll be able to apply Bayes' rule, as I have become convinced that this is the only way to finally be a rational being!

Ironically, the emotional momentum that Bayes' rule has given me sounds like irrational delirium. I cannot deny it. You may justifiably frown at me. You should be frowning at me. Indeed, I'm even pretty sure I am suffering from a huge cognitive bias caused by a sacralization of Bayes' rule. After all, it's impossible for me to be indifferent to the many secrets of Bayes' rule that I have managed to uncover myself - even though many others uncovered these secrets half a century before me.

Having said that, conscious of this bias, I promise I have fought - and I still do - against the *pure Bayesian*. I have kept trying to prove her wrong; I have kept trying to win a debate against her. In vain.

1.4 AN ALTERNATIVE TO THE SCIENTIFIC METHOD

In mathematics, when a conjecture seems to hold, we usually try to prove it to make it a theorem. Well, this is almost the case of Bayesianism!

As we shall see, for instance, the Jaynes-Cox theorem proves that Bayesianism is the only generalization of Aristotelian logic able to deal with plausibility in a coherent manner. Solomonoff's completeness theorem proves that the *pure Bayesian* will eventually identify all patterns in a data set. What's more, the theorem of expected gains given additional information shows that the *pure Bayesian* never loses by acquiring more data. Finally, statistical decision theory shows that Bayesian inferences are essentially the only admissible learning rules, in the sense that a learning rule is not dominated by any other, if and only if it boils down to applying Bayes' rule[5].

Many additional theorems supporting Bayesianism are unfortunately not discussed in this book. For instance, The Teller[6]-Skyrms[7] theorem

[5] These theorems are explained in chapters 3, 7, 9 and 12.

[6] *Conditionalisation and observation.* Synthese. P Teller (1973).

[7] *Dynamic Coherence and Probability Kinematics.* Philosophy of Science. B Skyrms (1987).

asserts that only a Bayesian is never extorted by a "Dutch book" scheme. Joyce's theorem[8] proves that we gain by making our beliefs follow the laws of probability, as prescribed by Bayesianism. Many of these theorems are nicely illustrated by the famous two envelope paradox[9].

Unfortunately though, I have had to state these theorems in a rough manner, as they rely on definitions and theorems that are hard to explain briefly. This is a major problem. In fact, any purist who wants to reject Bayesianism will be able to question and reject the hypotheses of the theorems. I do not claim that the theorems *prove* Bayesianism.

More generally, it seems impossible to *rationally* convince oneself that Bayesianism is the *right* philosophy of knowledge, the right theory of theories, or the right definition of rationality. After all, to convince ourselves of the relevancy of a concept, we need to first have in hand a philosophy of knowledge that measures the relevancy of concepts. To theorise theories, we need a theory that judges and discriminates theories of theories. To discuss rationality rationally, we need a rational definition of rationality.... We have a snake-biting-its-tail problem.

This difficulty is absolutely not specific to Bayes' rule. Any philosophy of knowledge seems doomed to suffer from self-reference. Besides, mathematicians have struggled for centuries to avoid self-referencing theories. Without much success (thanks a lot, Gödel!).

For instance, a supporter of Popper's philosophy, which is sometimes regarded as a description of the *scientific method*, will want to found knowledge on *falsifiability*. Yet, the very requirement of falsifiability does not seem falsifiable. Popper's philosophy seems inconsistent. Or, at least, it does not seem possible to accept Popper's philosophy according to Popper's criteria. This has led many to draw a line between philosophy and science, or between science or theology. This is called the *demarcation problem*. Yet, if you really think about it, this imaginary line is a pure (undesirable?) artefact of Popper's philosophy[10].

When it comes to self-consistency, the *pure Bayesian* performs better. Indeed, while she cannot prove the validity of her philosophy outside her framework, the *pure Bayesian* - for whom, as we shall see, all is belief - seems able to discuss Bayesianism without contradicting itself. Even better, I have applied Bayes' rule to my credences on Bayesianism.

[8] *A Nonpragmatic Vindication of Probabilism*. Philosophy of Science. J Joyce (1998).

[9] *Solve the Two Envelopes Fallacy*. Looking Glass Universe. M Yoganathan (2017).

[10] *Beyond Falsifiability: Normal Science in a Multiverse. S Carroll (2018).*

My heuristic computations have only increased my belief in Bayesian philosophy[11].

But there are two other more convincing arguments that have led me to favour Bayesianism over any other philosophy of knowledge. The first is the universality of Bayesianism. As opposed to Popper's philosophy which restricted the range of (scientific) knowledge, for instance by insisting on the reproducibility of scientific experiments[12], Bayesianism has no restriction on its range of applicability. Any phenomenon, whether it belongs to sociology, history, or theology, can be analyzed through the prism of Bayesianism. Bayesianism is a *universal* philosophy of knowledge.

The second argument consists of the rigor, the concision, and the clarity of Bayesianism. Indeed, Bayesianism defines inference rules[13] so precise that applying (even approximately) these rules seems to be sufficient to learn "well enough" about the world. This is a computer scientist's dream. The computer scientist would then only have to push the start button to enable the machine to reach its goal by simply following instructions. Of course, this is above all a description of artificial intelligence! And it's definitely not an accident if, over the last three decades, Bayes' rule has been at the heart of many research breakthroughs in this domain.

This rigidity of Bayesianism heavily contrasts with the *malleability* of most common versions of the scientific method. Indeed, many approaches often consider a sort of statistical toolbox, from which statistical tools may be (cherry-)picked for data analysis. Unfortunately, it has been argued that this allowed scientists to bias their conclusions[14], especially under publishing, financial, or ideological incentives. This has been argued to have led to a blow-up of misleading scientific publications.

More recently, through the work of researchers like Josh Tenenbaum, Karl Friston, and Stanislas Dehaene, Bayesianism has also become an essential theoretical framework to understand how our own intelligence works. In particular, in 2012, Dehaene gave a series of lectures

[11] This is related to so-called *hierarchical* Bayesian models, which we shall discuss in chapter 19. Note, however, that the *pure Bayesian* cannot technically discuss pure Bayesianism, at least in Solomonoff's setting, since she should only consider *computable* theories. But as she shall discuss in length in chapter 7, *pure Bayesianism* is *incomputable*.

[12] Reproducibility can be seen as a condition imposed by frequentism.

[13] We shall soon see what this means.

[14] *Medical Nihilism*. Oxford University Press. J Stegenga (2018).

at the prestigious Collège de France entitled *The statistician brain: The Bayesian revolution in cognitive sciences*. "Many biologists are skeptical with the idea that, in neuroscience, there may be general theories", he said. "[But] it really seems that [Bayesianism] yields a theoretical framework which can be applied in an extremely general manner [...] The very existence of general patterns in the architecture of the brain seems to be explained by the hypothesis [according to which] the brain is organized to compute statistical Bayesian inferences."

(Pragmatic) Bayesianism seems to be Nature's solution to natural intelligence[15].

1.5 THE OBJECTIVITY MYTH

Mysteriously enough, though, Bayesianism has long been rejected by several generations of first rank scientists. Why is that? Were the great scientists irrational? What caused the rejection of Bayesianism? And if this rejection is unjustified, what was the fallacy of these great scientists?

It turns out that the two centuries of epistemological war that this book hopes to put an end to boil down to the concept of *objectivity*. Better, the opposition between *subjective Bayesians* and *objective frequentists* can be summed up by the following questions: *What is a probability?*

I have a personal connection to this fascinating question. It was given to me in an oral exam for the entrance to the École Normale Supérieure (ENS) in Paris. This exam was supposed to be the presentation of a year-long project. I was quite proud of mine. I had modelled soccer games, estimated the levels of teams, and simulated different competitions[16]. In particular, based on two years of sports results, my simulations concluded that Portugal, France, and Italy were the three main favorites of the 2006 World Cup. Their probabilities of winning were 20%, 15%, and 10%. Not bad, given that the three teams would end up, respectively, 4th, 2nd, and 1st in the competition!

Examiners of the École Centrale and the École des Mines really enjoyed my work. They gave me 19 points out of 20. However, my simulations did not get the ENS examiners excited. They quickly stopped me. What they wanted to know is whether I could define what probabilities are.

[15] *Les algorithmes du vivant*. TEDxSaclay. LN Hoang (2018).
[16] *A model of football games*. Science4All. LN Hoang (2013).

My answer was frequentist. I claimed that the probability of an event was its limit frequency, when an experiment is repeated an infinite number of times. In particular, any empirical frequency would thus be an approximation of some fundamental and objective probability. Frequentists or not, the purists at ENS did not appreciate my efforts. They expected me to rediscover a mathematical definition of probabilities, for instance as unitary measures of sigma-algebras. I got 6 out of 20.

But let's not mourn my fate. Let's focus on what the *pure Bayesian* would call a naive mistake.

I was born a frequentist. I grew up searching for truths, whether mathematical or scientific. I accepted the existence and superiority of *objective* results. Even in 2013, when the troll student challenged me, the major part of the course I was teaching was essentially frequentist - and I thought that these were the *right* statistics to teach! Besides, my own model of soccer games was a classical example of a frequentist approach which, like Stein paradox[17], would have gained by acquiring some Bayesian flavour.

But crucially, the very nature of the probabilities I was manipulating could not be frequencies! The frequency with which France wins the 2006 World Cup is not 15%. This frequency is 0. There has been and there will have been only one 2006 World Cup. And France lost it.

But if the 15% predicted by the model was clearly not a frequency, what is it? Can we still say that it is a probability?

Yes, says the *pure Bayesian*. It's the probability that France wins the World Cup *according to the mathematical model*. In particular, this probability is *subjective*; it's the opinion of the model. But crucially, all probabilities are like this. According to the *pure Bayesian*, no probability is *objective*; and whoever disagrees confuses his subjective desires with a reality to force upon others. Probabilities are *model-dependent*.

Think about it. Any method to search for and organize knowledge is doomed to be biased by the mere choice of this method rather than another - especially if one starts invoking the imprecise Ockham razor, already "established" scientific knowledge, or the very problematic *p-values*. Worse, the way we look at, manage, and select our data inevitably biases the conclusions derived from the data analysis. As we shall discuss in length, facts are often incredibly misleading[18].

[17] We shall discuss Stein paradox in chapter 13.
[18] *How statistics can be misleading.* TED-Ed. M Liddell (2016).

What's more, the explicit mention of the method that was followed is insufficient. As data scientists using *machine learning* to extract useful information from *Big Data* quickly learn, the absence of human intervention is absolutely no guarantee of objectivity. Humans or machines, it seems that we always *have* to reason within our models of the world. This shows, the *pure Bayesian* claims that knowledge is necessarily *subjective*. It depends on the algorithm used to compute that knowledge.

This should make you feel uneasy. Bayesianism seems to lead to relativism. If all knowledge is subjective, does this mean that all opinions equally matter? Of course not. We may each see our own red; this does not mean that all opinions about the presence of red in the US flag are equally reliable.

In particular, those who apply Bayes' rule to the same data will end up giving their credences to the same models, especially if there are lots of data. What's more, even with relatively few data, the models that will win the credences of Bayesians will be way more relevant and *useful* than the favorite models of those who, exposed to the same data, have not applied Bayes' rule.

Note that Bayesianism (especially its pragmatic version) is not a substitute to modelling; it is rather a meta-model whose purpose is to discern useful models. The foundation of Bayesianism is in fact very well summed up by Bayesian George Box's holy quote: "All models are wrong, some are useful". I will often be using it! Whether this quote is "true" or not, I've found it incredibly useful to shortcut endless debates which seemed doomed to go nowhere - and thus to greatly bore me.

The *pure Bayesian* much prefers to judge the *usefulness* of models. Especially their *predictive* usefulness. Not their truth. Yet, according to her, judging adequately the usefulness of models can only be done through Bayes' rule.

1.6 THE GOALS OF THE BOOK

While I do intend to share and explain my enthusiasm for Bayesianism, and while I have the secret hope that this will make mathematicians, philosophers, and scientists question what they thought they knew about their disciplines, the goal of this book is actually not to convert you to Bayesianism. What I would like above all is to share with you some of the marvels that have made me fall in love with Bayesianism. And I am willing to bet - a typical Bayesian reflex - that you will be surprised and, I hope, seduced, by the many astounding consequences of Bayes' rule, as well as by its ubiquity in applied mathematics, in our own intuitive thinking, and in the organization of our societies.

Bayesianism explains why the scientific community is far more reliable than any of its members and why our brains are constantly victims of the anchor effect. It explains why it's more desirable to combine incompatible models and why Ockham's razor is an essential tool. It could even be the key to understanding the working of our memory and the usefulness of our dreams. Just as biologist Theodosius Dobzhansky once asserted that "nothing in biology makes sense except in the light of evolution", I would claim that a spectacularly wide range of mechanisms can only be understood through Bayesian lenses.

My discovery of Bayesianism has also been for me the chance to finally measure the extent of my ignorance. This is thanks to the language of probability theory that allows to quantify uncertainty. But most importantly, my inability to apply Bayes' rule, even in the simplest cases like the troll student puzzle, has forced me to acknowledge how bad I am at thinking. I have often had an irrational and unjustified confidence in my intuition, sometimes accompanied with a mysterious distrust of Bayes' rule. But after losing so many debates against the *pure Bayesian*, my Bayesian journey has forced me to acknowledge my unwavering overconfidence. In fact, this will be a major objective of this book. We will fight our overconfidence and measure the extent of our ignorance.

The rest of the book is roughly divided in four parts. In the first part, from chapters 2 to 7, we will straightforwardly tackle Bayes' rule and *pure Bayesianism*. Next, chapters 8 to 13 will reveal the hidden and unexpected presence of Bayesian principles in many phenomena. Then, chapters 14 to 19 will study *pragmatic Bayesianism* and its essential tools. Finally, the last three chapters will be a bit different. Chapter 20 will analyze the antirealist consequences of Bayesianism. Chapter 21 will track down the origins of my personal beliefs to better question our widespread overconfidence. Last but not least, chapter 22 will study consequences of Bayesianism on moral philosophy.

Unfortunately, this book, like any finite book, is absolutely not exhaustive. I apologize for its uncountable deficiencies. In particular, I will not take the time to compare in details Bayesianism to alternative philosophies of knowledge. My goal is more modest: I would like to help you understand the key aspects of Bayesianism. Or, at least, of what I understood. Indeed, like this book, my brain is finite too. Please forgive the extent of my ignorance. I will try to mention all the Bayesian reasonings I find worth mentioning, but I will necessarily omit what I do not know and what I have mistakenly considered unimportant.

In addition to my cognitive limitations, the depth of the book will also be limited by my desire to make it accessible to a wide audience. No prerequisite is assumed. As a result, I will not be nearly as rigorous as the *pure Bayesian* would want me to be - although I will do my best so as not to lead you to any misinterpretation. This is a popularization book.

Nevertheless, there is a good chance that you will not understand everything. Since I really wanted to present some of the most convincing arguments in favor of Bayesianism, I have chosen to provide sections of high mathematical level. These sections are "starred". Be warned. Even doctors in mathematics will struggle to understand all the notions of this book.

Do not rush the reading of the book. Take the time to ponder it. But do not give up either. The book is *not* increasingly difficult. You should be able to find pleasure in any chapter without having read the previous ones - even though it's probably better to read the chapters in the right order. This is not the textbook of a course. There will be no exam. You do not have to understand it all. I would even advise you to skip difficult paragraphs and carry on the reading. My goal, after all, is absolutely not to transform you into experts of Bayesianism.

What I would like above all is for you to search for beauty and pleasure in Bayesian thinking, as well as in the sciences useful to understand and illustrate Bayesianism. I would like you to behave like an explorer who has just arrived in unchartered territory and has plenty of intriguing wildlife, landscapes, and cultures to discover; and who will not necessarily spend much time learning all the subtleties of the language of local people. I would like you to enjoy *your* journey.

If you roughly follow my footsteps, I hope to fill you with enthusiasm, fascination, and questioning. This is the main goal of the book.

FURTHER READING

Probability Theory: The Logic of Science. Washington University. ET Jaynes (1996).

Rationality: From AI to Zombies. Machine Intelligence Research Institute. E Yudkowsky (2015).

Thinking in Bets: Making Smarter Decisions When You Don't Have All the Facts. Portfolio. A Duke (2018).

Medical Nihilism. Oxford University Press. J Stegenga (2018).

Conditionalisation and Observation. Synthese. P Teller (1973).

Dynamic Coherence and Probability Kinematics. Philosophy of Science. B Skyrms (1987).

A Nonpragmatic Vindication of Probabilism. Philosophy of Science. J Joyce (1998).

Beyond Falsifiability: Normal Science in a Multiverse. S Carroll (2018).

A Set-Partitioning Formulation for Community Healthcare Network Design in Underserved Areas. M Cherkesly, ME. Rancourt & K Smilowitz (2017).

Measuring Unfairness Feeling in Allocation Problems. Omega. LN Hoang, F Soumis & G Zaccour (2016).

The Feynman Series - Beauty. Reid Gower (2011).

Bayes: How one Equation Changed the Way I Think. J Galef (2013).

Think Rationally via Bayes' Rule. Big Think. J Galef (2013).

Solve The Two Envelopes Fallacy. Looking Glass Universe. M Yoganathan (2017).

How Statistics Can Be Misleading. TED-Ed. M Liddell (2016).

The Universal Turing Machine. ZettaBytes, EPFL. R Guerraoui (2016).

Les Algorithmes du Vivant. TEDxSaclay. LN Hoang (2018).

A Model of Football Games. Science4All. LN Hoang (2013).

CHAPTER 2

Bayes' Theorem

One of the biggest paradigm shift in my thinking, and also for a lot of people I know, has been learning about Bayes' rule.

Julia Galef (1983-)

We are spontaneously irrational creatures unable to correctly revise our beliefs, and understanding Bayes' rule can really help us improve.

Thibaut Giraud (1986-)

2.1 THE TROLL STUDENT PUZZLE

Let's get back to the troll student puzzle. A man has two kids. At least one is a boy. What is the probability that the other is a boy too? I invite you to try to solve the problem by yourself. Even if you do not succeed, the intellectual effort you'll put in will likely be useful for the sequel.

I'll now discuss the solution. The simplest approach to solve the puzzle consists of listing all possible cases. Let's call Alex and Billie the two kids. There are four possibilities:

- Alex and Billie are boys.
- Alex is a boy. Billie is a girl.
- Alex is a girl. Billie is a boy.
- Alex and Billie are girls.

These four possibilities seem equally likely *a priori*, that is, before we learn that at least one is a boy. Actually, they aren't. Biologists would

If this is what you think, then you are making a mistake that numerous first-rank mathematicians did before you. No need to feel shame! The Monty Hall problem has confused a lot of very smart people. In 1990, when Marilyn vos Savant proposed a correct solution to this problem in the journal *Parade*, 10,000 readers, including 1,000 PhD graduates, wrote to the journal and asserted that vos Savant got it wrong.

Even the world-class mathematician Paul Erdös, the man who has published the most in the history of mathematics, refused to believe vos Savant's rigorous proof. It's only when faced with simulation results that, to his dismay, Erdös reckoned he was wrong. The great Erdös did not understand Bayes' rule. He was not alone.

I was 13 when I first discovered the Monty Hall problem. I did not know Bayes' rule. Nevertheless, there was a convincing reasoning which was accessible to me. Indeed, if you know ahead of time that you will not switch curtains, then all happens as if Monty Hall did not build suspense by revealing a goat curtain. The probability of finding the car would thus be the probability that the curtain you initially chose hid the car. This probability is $1/3$. Your chance of winning when not switching curtains is one third. Weirdly enough, though, while I was quite convinced by this reasoning, I was still unable to determine the probability of winning by switching curtains.

Yet, if you lose by keeping your curtain, it means that the other curtain hides the car. After all, this other curtain is the one that Monty Hall suspiciously left unrevealed. In fact, what happens is that 2 out of 3 times, there is a goat behind your initial curtain. Whenever this is the case, once there are two curtains left, the car will necessarily be behind the other curtain. You then win by switching curtains. Two times out of 3.

The math is indisputable. You double your chances of winning by switching curtains! The *pure Bayesian* would win twice as often as he who has not thought through the problem carefully and sticks to his initial choice.

If you are still not convinced by this reasoning, I invite you to do the experiment yourself, as Erdös did. In an excellent BBC documentary, the mathematician Marcus du Sautoy posed the Monty Hall problem to comedian Alan Davies. Alan Davies thought he had his chances in a repeated Monty Hall game with no curtain switching, as opposed to du Sautoy who always switched curtains. After 20 attempts, Davies only won twice. Du Sautoy won 16 times. Granted, these figures do not match the $1/3$ and $2/3$ of the Bayesian theory - this is an instance of the law of

$\mathbb{P}[😷|👎]$. We obtain the most important equation of the philosophy of knowledge that this book presents, also known as *Bayes' rule*. Please take the time to notice its calligraphic elegance and the pattern that the symbols follow.

$$\mathbb{P}\left[😷 \mid 👎\right] = \frac{\mathbb{P}[👎|😷]\ \mathbb{P}[😷]}{\mathbb{P}[👎]}.$$

In words, the probability of having Ebola given a bad-news test is derived by multiplying the probability of a bad-news test if sick (which requires some thinking!) by the prior probability of being sick, divided by the probability of a bad-news test.

As announced in the first chapter, all you need are multiplications and divisions! How simple is that?

Of course, what makes this equation hard is not the computations it requires, but rather its interpretation - at least in the simple examples of this chapter. It's extremely easy (and tempting!) to misinterpret one of the terms of the equation. I strongly encourage you to take the time to think them through.

2.6 THE COMPONENTS OF BAYES' RULE

In the right expression, the probability $\mathbb{P}[😷]$ is the *prior*, sometimes known as the *base rate*. It's what we could (or rather, should) believe before learning the result of the test. In our case, we estimated it by comparing the number of known Ebola victims to the population of sub-Saharan countries. But this is merely a rough estimate. Besides, we did not even account for the length of your stay, which surely is a major thing to take into consideration to estimate the prior. Just as important is the frequency of interactions with local people, as well as exposure to sick individuals. All these effects are incredibly hard to quantify. We'll work with our rough estimate here.

The other quantity in the numerator in the right expression is the probability $\mathbb{P}[👎|😷]$ of a bad-news test given that we are sick. This term requires a bit of imagination. It requires us to leave the real world to imagine an alternative one where we definitely got infected by Ebola. In this alternative world, would we get a bad-news result to the test? This is the question that $\mathbb{P}[👎|😷]$ answers.

Contrary to us, the *pure Bayesian* not only can think along the lines of others' worldview, it's what she does day in day out! This is the art of *thought experiments*. These are essential components of Bayesianism.

image
not
available

image not available

Index