

Applied Philosophy of Science and Data Ethics

Dr. Carlos Vega

2021: Last compiled 2021-08-26

Contents

Preface	5
1 Scientific Goals, Methods and Knowledge	7
1.1 Overview	7
1.2 What is Science?	7
1.2.1 Scientific Goals and Knowledge	8
1.2.1.1 Data, information and knowledge	10
1.2.2 What is Philosophy of Science?	11
1.3 The scientific method	11
1.4 Methodology	12
1.5 Historical Examples	13
1.5.1 The most famous “failed” experiment	13
1.5.2 The smoke debate	14
1.5.3 Kekulé’s dream	16
2 Scientific Inference	17
2.1 Overview	17
2.2 Types of inferences	17
2.2.1 Deduction and Induction	17
2.2.2 Modus ponens and Modus tollens	19
2.3 The problem(s) of induction	21
2.3.1 David Hume’s Problem of Induction	21
2.4 The Hypothetico-deductive Method	23
2.4.1 A good hypothesis	24
2.4.2 Falsification	24
2.4.3 Confirmation	27
2.4.4 Other types of inference	29
2.5 Explanation	30
2.6 Examples	30
2.6.1 The problem is in your hands!	30
2.6.1.1 How a hypothesis is tested	31
2.6.2 Wason selection task	33
2.6.3 U.S.A. Presidents	34

2.6.4	<i>Yersinia pestis</i>	35
2.6.5	Risks of induction and non-epistemic values in ML	36
3	Extra Material	39
3.1	History of science	39
3.2	Theory-relatedness of observations	39
3.3	Gettier problems	39
3.4	Realism and anti-realism	40

Preface

Course Note:

This book is still under construction. There are more chapters coming. The current published version is a test to check serving the contents through GitHub.

This book gathers the contents addressed in the course **Applied Philosophy of Science and Data Ethics** from the [Master of Data Science](#) at the University of Luxembourg. This course will introduce basic philosophical and scientific concepts supported by examples and discussion. The course expects pro-active participation from the students in the form of presentations and essays as well as open debates.

This course aims to provide the students with guidelines and methodologies to identify epistemic and ethical issues present in data science. We expect the students to develop a critical eye that helps them mitigate such problems in their daily work as data scientists.

During this course, students will learn by example different layers of the scientific method and how they relate to data science and data ethics. In particular, they will learn how the mechanisms behind the data affect the data analysis, and how the different types of scientific inference condition method choice and affect the conclusions drawn from the analysis. In this sense, examples of statistical abuse, misconduct and bad visualization will be shown together with their, sometimes catastrophic, collateral consequences.

Disclaimer

Although the impact and extension of the topics addressed in this course are broad and diverse, its duration is limited. Hence the scope and depth of the contents are restricted. Consequently, several topics on Philosophy of Science are tackled superficially while some others are completely ignored. Such philosophical questions are handled from a practical data science point of view. Similarly, Data Ethics is a relatively new matter in continuous evolution. Therefore we will try to cope with the main issues in the most practical way.

Learning outcomes:

In line with the European Quality Framework, Bachelor degrees require a critical understanding of theories and principles, while Master degrees involves higher specialised knowledge and critical awareness of knowledge issues in a field. In this case, the field at issue is data science and the contents will tackle philosophical and ethical issues concerning data science. The aim is to provide students with a better understanding of method justification, which increases their knowledge about such methods, their scope, purpose and relation to other methods.

- Get familiar with the scientific goals and methods.
- Learn the most common data science and visualization misconduct problems.
- Critically evaluate ethical issues and method choice.

About this course

A great part of the first chapters of this course is inspired by the book from Prof. Dr Lars-Göran Johansson ([Johansson et al., 2016](#)) and the works of Prof. Dr Till Grüne-Yanoff ([Grüne-Yanoff, 2014](#)) as well as its great course at EDX on “Philosophy of Science for Engineers and Scientists”. Regarding the second part of the course covering data ethics. I would like to thank the University of Michigan for its online courses (from which I was already a fan during my PhD, especially Applied Data Science) and especially Prof. Dr H. V. Jagadish for its course on Data Science Ethics which inspired me on the contents and examples of this course. Moreover, this last part of the course would not have been possible without many relevant books on the topics tackled in this course (see References). Including The Book of Why ([Pearl and Mackenzie, 2018](#)); Ethics and Data Science ([Loukides et al., 2018](#)); Philosophy of Natural Science ([Hempel, 1966](#)); How charts lie ([Cairo, 2019](#)); Automating inequality ([Eubanks, 2018](#)). I hope any resemblance or imitation is seen as an act of flattery.

About this book

This book was made thanks to the great tutorial available on the book “[Open tools for writing open interactive textbooks \(and more\)](#)”.

License CC BY-SA 4.0 license

The book is released under a creative commons [CC BY-SA 4.0](#) license. This means that this book can be reused, remixed, retained, revised and redistributed (including commercially) as long as appropriate credit is given to the authors. If you remix, or modify the original version of this open textbook, you must redistribute all versions of this open textbook under the same license - CC BY-SA 4.0.

Chapter 1

Scientific Goals, Methods and Knowledge

1.1 Overview

1. A
2. B
3. C
4. D

1.2 What is Science?

This question attempts to answer what common features share subjects such as physics or biology to be called sciences, i.e. what it is that which *makes* something a science. Among other things, science aims to understand, explain and predict the world we live in. But also religions, astrology or alchemy attempt to understand, explain or predict our world. What makes them different from science?

Four historical elements are essential for the development of a scientific approach. Namely: to seek explanations of natural phenomena; to argue; to investigate the rules of argumentation and logical validity; to build them into a logically consistent system. ([Johansson et al., 2016](#))

Rather than finding a proper definition of science, which many have struggled with, we will focus on what makes science different and why its methods are called scientific.

1.2.1 Scientific Goals and Knowledge

The main goals of science include prediction, explanation, understanding, and design. Through observation we can draw explanations and achieve understanding of nature phenomena. Once we understand we can aim to make predictions. Thanks to our understanding we can as well design experiments, instruments and solutions that help us further explaining, understanding and predicting our world. Predicting X means knowing that at time t , X will happen. Explaining X means to know the cause(s) that produced X . Designing X requires knowing that artifact X will satisfy certain functions F . All these goals share a common ingredient, scientific knowledge. Scientists arrive to such knowledge by applying the scientific method (see § 1.3). The goals of science are achieved through a series of activities that constitute the scientific method which include systematic observation and experimentation, inductive and deductive reasoning, and the formation and testing of hypotheses and theories.

Knowledge is justified true belief — Plato (428 - 348 BC)

The most popular definition of knowledge was given by philosopher Plato in the above's quote. This definition specifies that a statement must meet three criteria to be considered knowledge. This definition of knowledge is sufficiently good for this course. However, the definition of knowledge is an ongoing debate among epistemologists. Although these criteria are necessary conditions, they are not sufficient as there are situations that satisfy all these conditions and yet don't constitute knowledge (see [Gettier cases](#)) but such cases are rather philosophical and will not be discussed during this course.

- **True** because statements must refer to an actual state of the world.
 - A wet sidewalk does not necessarily imply it rained even if you believe so.
 - Even if we are justified to believe that something is true, it might not be true.
- **Justified** because you need proper proof, evidence or reasons to defend our statement.
 - Even if it actually rained, a wet sidewalk caused by a sprinkler is not good justification for you to believe it rained.
- **Belief** because even under justified reasons about true facts, people can choose not to believe such knowledge. We define belief as to the state of mind of a person that thinks something is the case. This state of mind is of course *tied to the individual* and *comes in degrees*. We act based on our beliefs and values, and new knowledge can affect these.

Certainty of belief and truth are different. Is possible to have certain beliefs about false claims. Similarly, we can have uncertain beliefs about true claims. From tossing a coin, we can expect a fair probability in which head and tails have the same probability. But we cannot know for sure if the coin is biased or no until it lands. Similarly, is possible that even our best theories are wrong or partially wrong. Even after many successful experiments, they might be proved

wrong (see ??). In fact, scientific hypotheses can rarely if ever be proved right, they can, however, be proven wrong.

“We never are definitely right, we can only be sure we are wrong” —
Richard Feynman

Below you can find a [clip from the last lecture](#) of a series of 7 special Messenger Lectures given by the renowned American theoretical physicist Richard Phillips Feynman. [The transcription can also be found online.](#)

While it is relatively easy to determine cases of failed justification, it is much harder to identify what suffices to justify a belief. Few claims can be conclusively proven so that no doubt remains. An ideal justification of a belief would consider all relevant reasons for and against believing a statement. This is why science is a human enterprise where justifications, hypotheses and experiments are made public for review, replication or rebuttal.

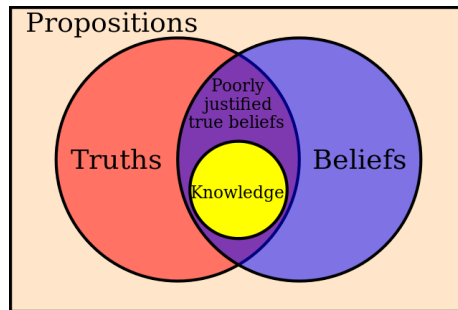


Figure 1.1: A Venn diagram illustrating the classical theory of knowledge.

Definitions should not be accepted without reason, and instead, we should attend to the arguments that support such definitions. Certain definitions may have widespread popularity but that doesn't make them any more true. For example, a dolphin is a mammal even if many people consider it a fish. In the same way, tomatoes and cucumbers are fruits for botanists even if we daily sort them as vegetables. And sometimes, even the [EU](#) and [the Supreme Court of the United States of America](#) need to act to set certain market debates.

As another example, the first astronomers who lacked the telescope believed in the geocentric model because their observations did not suffice to reject it. These first astronomers had **false justified beliefs**. After the advent of the telescope in 1609, the geocentric model was rejected. But how do we know we are not in a similar situation that of the pre-telescope astronomers? Events as recent as the Michelson-Morley (see [1.5.1](#)) experiment, the expeditions of Sir Arthur S. Eddington (see ??) or the predictions of Urbain Le Verrier (see ??) have changed our conceptions of the universe and physics forcing scientists to re-formulate models and theories.

In the next chapter we will see how knowledge is obtained.

1.2.1.1 Data, information and knowledge

Nowadays, technology allows us to collect data into datasets, transform datasets into information and arrive at new knowledge. Such processes have always been crucial in science but computer science comes to question concepts such as data, information and knowledge. (Johansson et al., 2016)

By **knowledge**, we can understand three different things. First, knowledge of truths, e.g. we know that the sun rises on the east. Such knowledge can be obtained by reading a book or listening to the radio. The second category of knowledge consists of skills, such as riding a bike or speaking a foreign language. However, this knowledge requires more than language to be communicated. It requires practice. Finally, the third category is the knowledge of objects, what Bertrand Russell called knowledge by acquaintance (Russell, 1912). This knowledge is obtained through experience.

In common English, we can't distinguish between knowledge of truths and objects. However, languages such as German, French or Spanish make clear this distinction by using different verbs, Russell proposed to use the word “acquaintance”.

- German: wissen vs kennen.
- French: savoir vs connaître.
- Spanish: saber vs conocer.

Example from William James (1890) — “I am acquainted with many people and things, which I know very little about, except their presence in the places where I have met them. I know the colour blue when I see it, and the flavour of a pear when I taste it; [...]”

What is the difference between data and information?

The following excerpt from (Johansson et al., 2016) may clarify this question:

But why call the input ‘information’? The reason seems to be that we can describe the input as being about something, often the state of the environment. It has content. Or rather, when we humans describe the input and the workings of the system we find it natural to talk as if the information-containing system consciously sent messages to us humans; we say that the systems obtain information, transmit information or store information about something, as if it were like a human mind. The core feature of this use of the word ‘information’ is thus its aboutness, its intentionality.

Finally data. It is common in computer science to say that information is data with meaning. This is ok as far as it goes, but what is ‘meaning’? And how do data acquire meaning? It seems that minimally it means that meaningful data becomes information when we have been able to formulate declarative sentences expressing the information that is obtained from a data set. Almost anything can

be data. In order to obtain data from e.g. a story, from light from distant stars, or from the result of an experiment, we need to divide the stream of sounds, lights, or states of detectors into distinct items. When using written text as data source one must divide the string of linguistic signs into distinct items, such as words or longer or shorter expressions. [...] In short, in order to obtain a data set, we need to define a principle for dividing up something into distinct pieces. Hence from a conceptual point of view, discerning data and collecting a data set presupposes that we have a prior principle of making distinctions within a phenomenon. [...] Sometimes we have lot of background knowledge from start.

In short, a piece of knowledge is a piece of information for which the knower can provide good reasons.

1.2.2 What is Philosophy of Science?

One of the tasks of philosophy of science is to question assumptions that scientists take for granted. For example, suppose a scientist conducts an experiment that yields a particular result. The scientist then repeats the experiment a couple of times more obtaining the same result. The scientist then stops repeating the experiment, convinced that repeating it under the same conditions will produce the same result. But *why* does the scientist assume that future repetitions will provide the same outcome? How are we sure this is true?

Therefore, one of the main objectives of philosophy of science is to study the methods and methodologies of enquiry used in the sciences, understanding how techniques like experimentation, observation and theory building enable scientists to reveal new knowledge. Philosophy of science asks questions such as: What is knowledge? What is a scientifically acceptable observation? What makes an explanation scientific? What is a scientific theory?

Finally, the philosophy of science tackles a wide range of topics that would require its own master. Moreover, not all topics are directly related to the aims of this course and the scope of the master. For this reason, a brief summary of the topics left behind is included in § 3. Of course, the curriculum is subject to change in the future and the list might change too.

1.3 The scientific method

The scientific method is the main pillar of science. All science begins with *observation*, as this is the **first step** of the scientific method. Moreover, such observation must be *repeatable*, either actually or potentially. Once an observation has been made, the **second step** involves the definition of a *problem*, or in other words, asking a question about the observation. However, such a question needs to be valuable scientifically, it must be *relevant* and must be *testable*. Questions need to be reformulated until they become testable. The **third step**

may seem a rather unscientific procedure as it involves guessing what the answer to the question might be by postulating a *hypothesis*. The **fourth step** will tell the scientist if the *hypothesis* is correct through *experimentation*, which tests the validity of a scientific guess. Notwithstanding, experiments do not guarantee a scientific conclusion. Experiment results represent *evidence*, i.e. the hypothesis in answer to the question is confirmed as correct or invalidated. Given the latter, a new hypothesis with new experiments might be needed. Finally, experimental evidence is key for the **fifth step** of the scientific method, the formulation of a *theory*. A good theory has a *predictive* value, usually predicting that something is *likely* to happen with a certain degree of probability (Nidditch, 1968).

1.4 Methodology

A method is a particular tool to reach a particular goal (e.g. statistical test). Methodology is the systematic assessment and justification of method choice. Scientists often need to choose between alternative methods in order to reach a particular scientific goal. But specifying a goal does not directly determine what method to choose. We need to consider the reasons why some method is better than another for a particular goal. This process could require a better definition of the initial goal or learning more about the context and domain where the methods will be applied. Methodology must be distinguished from describing methods, which usually concerns the design and implementation of particular research approaches and focus on the technical aspects (e.g. how to program simulations or set up instruments).

For example, a laboratory experiment can be advantageous because the test conditions can be controlled but laboratory experiments might not be realistic enough for certain tests. On the other hand, a field experiment provides more realistic test conditions but is difficult to control all variables.

Similar considerations may be necessary for other seemingly trivial questions such as model choice or data visualization. Should we use a significance test or a Bayesian approach? Should we present our results using a bar chart or a violin plot? Should we use a structural model or a quantum model? Methodology asks questions such as: What methods are available to reach a particular goal? What reasons speak for or against the alternatives? How should we weight the reasons to form a final decision?

How do we decide between alternative methods? Is there a way to determine what is rational to choose? Traditionally there are three ways to choose between alternative methods.

By convention, The methods are chosen because you have been taught to, or because is an established convention between your peers. Conventionalisms create long-term when methods become dominant in a field. A good example is the use of p-value in hypothesis significance testing. Similarly, accuracy and precision metrics in Machine Learning can be considered conventionalism. More

problems arise when different disciplines have different conventions, hindering inter-disciplinary work.

Outcome-oriented. While choosing the method that yields the best results may seem well-intended and appropriate, this certainly sounds very vague too. The intention is to find a method that serves some purpose best, but this purpose is sometimes not sufficiently clear. Science frequently involves long-term projects where the final material outcome is uncertain or unknown. For example, the International Space Station or the Large Hadron Collider. This methodology raises the question of how to measure the outcome. For instance, is speed the best way to assess which car is best? Should we focus on fuel autonomy or pollution instead? What about combining all of them?

Reason-based. Choosing the method based on the overall best reasons seems the best option, particularly when the reasons include considerations that justify choosing a method over others for a given scientific goal (e.g. prediction). But sometimes there are methods that despite providing more valuable results could be unethical and/or illegal. For example, randomized control trials (RCT) are often employed to test the effectiveness of a new drug. Participants are divided *at random* into two groups (treatment and control), eliminating the effect of confounding factors on the outcome of interest. However, RCTs are not always feasible, for either practical or ethical reasons. For instance, it won't be ethical to assign people to smoke for decades in order to study if cigarette smoking causes cancer. These other aspects need to be weighted together with the scientific reasons during method choice.

See 1.5.2 for an example of how reason-based methods are not always easy to implement while at the same time, outcome-oriented methods led the mainstream of an important debate.

1.5 Historical Examples

1.5.1 The most famous “failed” experiment

The Michelson-Morley experiment (1887) was designed to detect the motion of the Earth through the luminiferous aether. XIX century physicists used aether to explain how light could be transmitted through empty space between the Sun and the Earth. The result of this experiment is considered to be the first strong evidence against the then-prevalent aether theory, and the beginning of a new line of research that eventually led to special relativity, which rules out a stationary aether.

To the ancients, the concept of a void universe was impossible. Aristotle arrived at the hypothesis of the aether to explain the cosmos and several natural phenomena such as the movement of the planets. By the XIX century, the aether became more than a philosophical need. Whenever there is a wave, something must be waving. But what waves when light waves travel from the Sun? For

XIX physicists, the aether was the medium through which light waves from the Sun would propagate.

Michelson and Morley attempted to detect the absolute motion of Earth through space. For that, they set an experiment in which a beam of light was sent through a half-silvered mirror used to split the light beam into two beams travelling at right angles to one another. The beams were then reflected back to the half-silvered mirror by two respective mirrors and recombined into a single beam. The experiment can be seen as a race between two light beams. If the beams arrive in a tie, the result is a bright spot at the centre of the interference pattern, otherwise, a destructive interference would make the centre of the image dark. The hypothesis foretold that a tie was not possible since the two beams were racing on a moving track. It was assumed that the Earth was moving through the aether and therefore the beam should trace different paths with respect to the aether.

The extent to which the negative result of the Michelson–Morley experiment influenced Einstein is disputed. However, the null result helped the notion of the constancy of the speed of light gain acceptance in the physics community. This example shows the impact a well-designed experiment can have.

For a longer and deeper explanation of the experiment, its historical context and consequences, watch [episode 41 from The Mechanical Universe](#). The timeline of luminiferous aether can be found [at the Wikipedia](#).

1.5.2 The smoke debate

In the mid-1700s, James Lind discovered that citrus fruits prevent scurvy, while in the mid-1800s, John Snow figured out that water contaminated with faecal matter caused cholera. These two examples share a common fortunate one-to-one relation between cause and effect. Deficiency of vitamin C is necessary to produce scurvy. Similarly, cholera bacillus is the only cause of cholera.

However, during the late 1950s and early 1960s, whether or not smoking caused lung cancer was not clear. The subject of the debate wasn't tobacco or cancer but rather the word *caused* as one of the most important arguments against the smoking-cancer hypothesis was the possible existence of confounding factors that may cause lung cancer and nicotine dependency. Many smokers live long lives without getting lung cancer while others develop cancer without ever smoking. Plotting the rates of lung cancer and tobacco consumption makes the connection impossible to miss (See Figure 1.2). However, time-series data are poor evidence for causality. Researchers already knew about RCT though its use was unethical in this case.

Austin B. Hill proposed to compare patients already diagnosed with cancer to a control group of healthy volunteers. The results showed that all but two of the 649 lung cancer patients had been smokers. This type of study is today called a case-control study because it compares cases to controls. However, this

method has some drawbacks too. First, the study is retrospective, meaning that we study people known to have cancer and look back to understand why. Second, the probability logic is backwards, as the data tell us the probability that a cancer patient is a smoker instead of the probability that a smoker will get cancer.

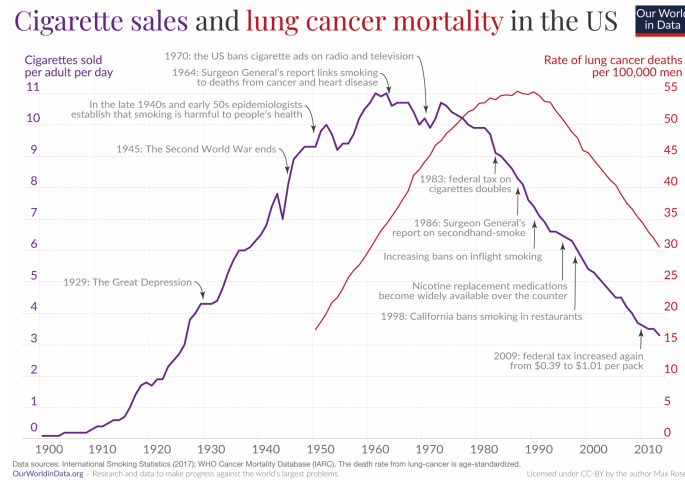


Figure 1.2: Source: [Our World in Data](#).

However, this method has some drawbacks too. First, the study is retrospective, meaning that participants known to have cancer are considered and researchers look back to understand why. Second, the probability logic is backwards, as the data tell us the probability that a cancer patient is a smoker instead of the probability that a smoker will get cancer. Moreover, case-control studies admit several possible sources of bias such as recall bias or selection bias. Hospitalised cancer patients were not a representative sample of the population, not even from the smoke population. Researchers were careful to call their results an “association”. Later on, the study was replicated with similar results. Deniers such as R. A. Fischer were right to point out that repeating a biased study doesn’t make it any better as is still biased.

We won’t focus on how this story ends here but is important to notice how methods chosen based on scientific reasons are sometimes tough to implement and often need to fight against outcome-oriented studies such as those sponsored by leading tobacco companies.

In the end, many subsequent studies settled the smoking-cancer debate. We will come back to this example in upcoming sections of the course. If you can’t wait, read Chapter 5 from the Book of Why, by Judea Pearl and Dana Mackenzie ([Pearl and Mackenzie, 2018](#)).

1.5.3 Kekulé's dream

The third step of the scientific method (see § 1.3) requires guessing an answer (or hypothesis) to a previously determined question. There is no clear method to arrive at a hypothesis. Experience, historical context, and previously failed hypothesis condition how a hypothesis is conceived. But sometimes hypotheses can be reached in the most unlikely and unconventional of ways. It makes no difference as long as the hypothesis is then scientifically tested before its acceptance. One of the most famous examples is the structural model of the benzene molecule. In 1865 the chemist August Kekulé hit on the hypothesis of the structure after dreaming of a snake trying to bite its tail (See Figure 1.3).

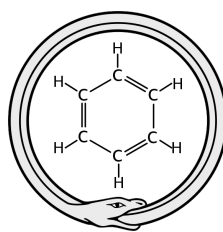


Figure 1.3: Source and credits to: Haltopub, from [Wikimedia](#).

Chapter 2

Scientific Inference

2.1 Overview

But our sun is only one of a billion-trillion stars within the observable universe. And those countless suns all obey natural laws some of which are already known to us. How did we discover that there are such laws? If we lived on a planet where nothing ever changed, there wouldn't be much to do, there'd be nothing to figure out. There'd be no impetus for science. And if we lived in an unpredictable world where things changed in random or very complex ways, we wouldn't be able to figure things out. And again, there'd be no such thing as science. But we live in an in between universe where things change alright, but according to patterns, rules, or as we call them, laws of nature. If I throw a stick up in the air, it always falls down. If the sun sets in the west, it always rises again the next morning in the east. And so it's possible to figure things out. We can do science. And with it we can improve our lives. — Carl Sagan

2.2 Types of inferences

Most scientific conclusions are uncanny at first glance and difficult to believe without more information and proper explanations about them (e.g. expansion of the universe, electromagnetism, etc.). How do scientists reach such unlikely conclusions? An inference is the act of reaching a conclusion from known facts but there are multiple types as we will see below.

2.2.1 Deduction and Induction

A good argument is one whose conclusions follow from its premises. But how do we tell if the conclusion is a consequence of its premises? Is often assumed that

as long as the premises are valid, the conclusions will be valid too. This does not imply that the conclusion is also true. The premises might not be true, but if they are true, then the conclusion will also be true. However, is the truth of the premises always *necessarily sufficient* for the truth of the conclusions? Logicians distinguish between deductive and inductive inference. (Douven, 2021)

Below there is an example of a deductive inference with two premises followed by a conclusion.

```
All Frenchmen like cheese
Loubin is a Frenchman
-----
Therefore, Loubin likes cheese
```

```
All As are Bs
a is an A
-----
Therefore, a is a B
```

We call an inference *deductive* whenever the conclusion *necessarily* follows from the premises. **The truth of the premises *guarantees* the truth of the conclusion.** Or in other words, what is inferred is *necessarily* true if the premises from which it is inferred are true. We call this type of inferences *explicative*.

Not all inferences are deductive. For example:

```
The first five eggs in the box were good.
All the eggs have the same best-before date stamped on them.
-----
Therefore, the next egg will be good too.
```

In this case, the premises do not entail the conclusion. Even if the previous eggs were good, it is possible that the next egg will be rotten. In this case, is logically possible for the premises to be true and yet the conclusion false. We call this type of inferences *inductive*. Contrary to deduction, where the truth of the premises guarantees the truth of the conclusion, **inductive inferences are *ampliative* — since whose conclusions go beyond what is contained in their premises** — and their conclusions could be totally wrong even if infinitely many examples confirm them. (Bergadano, 1991)

In these regards, deduction seems safer than induction. Whenever we reason deductively we can be sure that given true premises we will reach true conclusions. On the other hand, **inductive reasoning can take us from true premises to a false conclusion.** Notwithstanding, we rely on inductive reasoning every day. For instance, every day we turn on our computers and we are confident they will not explode in our faces. (Okasha, 2016) But why? Simply because we do it every morning and it has never exploded up to now.

We are sure that the sun will rise tomorrow, and if we are asked why we believe so, we will naturally answer “Because it always does”. We believe that it will rise

in the future because it has risen in the past. Of course, when we are challenged to answer what *justifies* our belief we can refer to the laws of motion and nature. But will the laws of motion remain the same tomorrow? ([Russell, 1912](#))

2.2.2 Modus ponens and Modus tollens

Course Note:

The following content relates to deduction and is usually taught in high school philosophy courses as part of propositional calculus. It will help getting a better understanding of the deductive inference rules. If this is already clear to you, feel free to jump to the problem(s) of induction [2.3](#).

There are two rules of inference in deductive reasoning. We call deduction top-down logic because we reach particular conclusions from general premises. Whereas in bottom-up logic the conclusion is reached by generalizing from specific cases.

- Modus ponens: $P \text{ implies } Q$. P is true. Therefore Q must also be true.
- Modus tollens: If P , then Q . Not Q . Therefore, not P .

The form of a **modus ponens** argument looks like a syllogism consisting of two premises and a conclusion. The first premise is a conditional if-then claim (e.g. $P \text{ implies } Q$). The second premise is an assertion that P (the antecedent of the first premise) is indeed true. From these two premises, it can be concluded that Q , (the consequent of the first premise) must be true as well.

If P , then Q .

P .

Therefore, Q .

The next example fits the form of *modus ponens*.

If today rains, John will take the umbrella.

Today is raining.

Therefore, John will take the umbrella.

The argument is valid but it doesn't matter if the statements in the argument are actually true. An argument can be valid but nonetheless unsound if their premises are false. *Modus ponens* rule can be written as $P \rightarrow Q, P \vdash Q$. In logic, an argument is sound if it is both valid in form and its premises are true.

On the other hand, the form of a **modus tollens** argument also consists of two premises and a conclusion. The first premise is a conditional if-then claim (e.g. $P \text{ implies } Q$). The second premise is an assertion that Q (the consequent of the conditional claim) is not the case. From these two premises, it can be

concluded that P is also not the case. *Modus tollens* rule can be written as $P \rightarrow Q, \neg Q \vdash \neg P$.

If P , then Q .

Not Q .

Therefore, not P .

Modus tollens is specially important in falsification (see 2.4.2). For instance, we take our hypothesis H to test and assume that is true. If H is true, then consequent C is true. We make an observation and see that C is false. Therefore, we conclude that H is false.

If H , then C .

C is false.

Therefore, H is false.

There are other forms of arguments that are apparently **similar but invalid forms**.

Affirming the consequent. This formal fallacy consists of taking a true conditional statement $P \rightarrow Q$ and invalidly inferring its converse $Q \rightarrow P$. For example, the statement “if the light is broken, the room would be dark” does not justify inferring the converse “the room is dark, therefore the lamp is broken”. This situations may arise when a consequent has more than one possible antecedent.

Denying the antecedent. This fallacy is committed by reasoning in the form: If P , then Q . Therefore, if not P , then not Q . This kind of arguments can seem valid at first glance. Consider this famous example from Alan Turing:

If each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines. — Alan Turing

Men could still be machines that do not follow a definite set of rules.

Another trivial example of this second fallacy.

If you are a bus driver, then you have a job.

You are not a bus driver.

Therefore, you have no job.

2.3 The problem(s) of induction

Do scientists use induction? Pretty much all the time. Whenever scientists move from limited data to general conclusions scientists reason inductively. **In inductively valid arguments, the (joint) truth of the premises is very likely (but not necessarily) sufficient for the truth of the conclusion.** For instance, a newspaper may run the headline “scientists find experimental proof that transgenic maize is safe to eat”. This means scientists tested transgenic maize on a large number of people without finding any issues. Does this *strictly prove* that transgenic maize is safe? Is this prove as strong as the proof of the Pythagoras’ theorem? Going from “the transgenic maize didn’t harm any of the people on whom it was tested” to “the transgenic maize will not harm anyone” is an inductive inference, not deductive.

Writing Note:

Suppose the following inductive inference I : If the probability of observing R , given that H is true, is smaller than a significance level of 0.05, then reject H . Is important to distinguish between the two following things:

- Justification *with* an inference rule: Justifying the conclusion by pointing to the premise and the inference rule. Inference rules justify conclusions.
- Justification *of* an inference rule: What makes I a good inductive inference? Why not choosing other parameters? The choice of a particular inference rule must be justified.

2.3.1 David Hume’s Problem of Induction

We use induction to justify our statements but how do we justify induction itself? How would you convince someone else that induction is a good inference method? The Scottish philosopher David Hume (1711-76) argued that the use of induction cannot be rationally justified at all.

In 1739, still under the shadow of the bubonic plague in Europe, David Hume publishes *A Treatise of Human Nature*, presumably without knowing that his work would not only continue to be debated more than 200 years later, but also still remarkably relevant in the technological advances of our time. In *the problem of induction* Hume argues that we cannot make a causal inference just by *a priori* means, and poses the question of how we can conclude from the observed to the unobserved.

Hume admitted that we use induction all the time in everyday life and science but insisted that this is just a matter of brute animal habit. What does he mean by that? Bertrand Russell (1872-1970) gives us a good example on this. He argues that the inductive association is also present in animals.

“And this kind of association is not confined to men; in animals also it is very strong. A horse which has been often driven along a

certain road resists the attempt to drive him in a different direction. Domestic animals expect food when they see the person who usually feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken. [...] The mere fact that something has happened a certain number of times causes animals and men to expect that it will happen again. Thus our instincts certainly cause us to believe that the sun will rise to-morrow, but we may be in no better a position than the chicken which unexpectedly has its neck wrung.” — (Russell, 1912)

Hume arrived to this conclusion by noting that whenever we make inductive inferences we presuppose the *uniformity of nature*. Remember the eggs box example in § 2.2.1 ? Our reasoning depends on the assumption that objects that we have not examined yet will resemble those objects that we have already examined. Then, Hume argues that we cannot prove the truth of the uniformity assumption. Basically, from the mere act of being able to imagine a world where nature is not uniform but changes at random it follows that we cannot prove that the uniformity assumption is true. Also, if we try to argue for the uniformity assumption on empirical grounds, we end up reasoning in a circle.

Note on uniformity of nature:

Notice how Machine Learning (ML) models can be regarded as inductive machines performing inductive inferences based on previous observations. For the ML model to perform well on novel data, is often assumed that novel data will resemble past data.

Hume refers to this assumption as the Principle of Uniformity of Nature: “*If reason determined us, it would proceed upon that principle, that instances, of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same.*”

And it continues: “*Our foregoing method of reasoning will easily convince us, that there can be no demonstrative arguments to prove, that those instances, of which we have had no experience, resemble those, of which we have had experience. We can at least conceive a change in the course of nature; which sufficiently proves, that such a change is not absolutely impossible. To form a clear idea of any thing, is an undeniable argument for its possibility, and is alone a refutation of any pretended demonstration against it.*”

(Hume, 1739) T. 1.3.6.4

The conclusion then is that our tendency to project past regularities into the future is not underpinned by reason. The problem of induction is to find a way to avoid this conclusion, despite Hume’s argument (Henderson, 2020). **Hume’s problem of induction is still an active area of research for philosophers.** There are many different ways to respond to Hume’s argument, yet none is

fully convincing. Peter Strawson (1950s) used the following analogy: justifying induction is like asking whether the law is itself legal. This is rather odd, since the law is the standard against which the legality of other things is judged. Others, like Karl Popper (1902-1994) argued that science is not in fact based on inductive inferences at all and presented a deductivist view of science. We will study this in detail in § 2.4.2.

As scientists, Hume's problem of induction may leave a huge void in our heart. An empty feeling that science is indeed fallible and the sudden realisation of the impossibility of establishing the truth or falsity of scientific laws (Rosenberg and McIntyre, 2019). But perhaps there is a way to fill such gap, and perhaps big part of the effort of science is put on filling this void with as much certainty as possible.

Hume's argument concerns specific inductive inferences such as `All observed instances of A have been B` and `The next instance of A will be B`. Hume's argument proceeds as follows:

- Every inference is either inductive or deductive.
- To justify an inductive inference rule I , this rule must be inferred from some premises.
- Is not possible to infer the rule I deductively, because there are no necessary connection between past and future inferences.
- Therefore, the rule I must be inferred inductively.
- When inferring I inductively, we must invoke another inductive inference rule J to justify this induction. But then, how do we justify J ? ... [*infinite regress*]

Just because an inference rule has yield true conclusions in the past does not necessarily imply that it will do so in the future. Consequently, Hume concludes that no inductive inference rule can be justified. But, does this mean all scientific inductive inferences are not justified?

Note for data scientists!

If we visualise the data as points in a plane; every set of finite points belongs to infinite functions or curves. The problem of induction, in this case, consists in establishing criteria that allow us to say that the finite series of data confirms only one of the functions, or less dramatically but just as problematic, that one is more confirmed than the others (Díez and Moulines, 1997). (See the problem of underdetermination in § 2.4.3).

2.4 The Hypothetico-deductive Method

In the section about the **scientific method**, we learnt how scientists begin proposing (or guessing) unproven hypotheses. After an initial consideration of the problem and collection of data a conjecture or hypothesis to explain a particular

phenomena is formulated. Afterwards, deduction is used to derive consequences or observable implications $\{C_i\}$ from such hypotheses H . These consequences should be relevant for H and observable directly or with the help of instruments (e.g. microscope, MRI, etc.). Next, hypotheses are put to test and either based on the results scientists decrease or increase the confidence over the hypotheses.

- 1. Propose a hypothesis H .
- 2. Deduce observable consequences $\{C_i\}$ from H .
- 3. Test. Look for evidence that conflicts with the predicted consequences $\{C_i\}$ in order to disprove H .
- 4. If $\{C_i\}$ is false, infer that H is false, reformulate H . (See § 2.4.2)
- 5. If $\{C_i\}$ is true, increase confidence in H . (See § 2.4.3)

For relevant examples, check 2.6.1 and 2.6.2.

2.4.1 A good hypothesis

There are though some criteria for a [good hypothesis](#). Apart from other criteria such as parsimony, scope, fruitfulness and conservatism, these are other criteria to recall.

- It should be an statement that can be either true or false (e.g. “Boiling point of a liquid increases with increase in pressure”). In other words, it should be **testable and falsifiable**. We must be able to test the hypothesis using the methods of science and according to Popper’s falsifiability criterion, it must be possible to gather evidence that will reject the hypothesis if it is indeed false.
- A hypothesis must not be a tautology (i.e. claims that are necessarily true or false; e.g. “Either it will rain tomorrow or it will not rain.” or “all bachelors are unmarried”).
- Hypotheses should be informed by previous theories or observations and logical reasoning.
- Finally, the hypothesis should be positive. That is, the hypothesis should make a positive statement about the existence of a relationship or effect, rather than a statement that a relationship or effect does not exist.
- Finally, it should have some generality (e.g. “things of certain type...” or be about some non-directly observable property of a particular.

https://en.wikipedia.org/wiki/Hypothetico-deductive_model

2.4.2 Falsification

According to the Hypothetico-deductive method (H-D), a hypothesis is formulated, then relevant consequences are deduced, and finally we observe whether

these consequences are false or true. Depending on these observations the hypothesis will be either falsified or confirmed.

Is important to note a key difference between confirmation and falsification. In step 4 of the **H-D method** we can infer the falsity of the hypothesis from the falsity of even a single one of the expected consequences. In contrast, in step 5 confirmation of the hypothesis is not inferred from the truth of even a large set of the consequences. Instead, we only increase our confidence on the hypothesis after finding that many consequences of the hypothesis are true. This difference is referred as the **asymmetry between confirmation and falsification**. Although a scientific theory can never be proved true by a finite amount of data, it can be proved false, or refuted by a single experiment.

“No amount of experimentation can ever prove me right; a single experiment can prove me wrong.” — Albert Einstein

This asymmetry forms the basics of Karl Popper’s (1902-1994) falsificationism.

- Conjecture falsifiable hypotheses.
- Seek to falsify these hypotheses with observable evidence.
- Reject any falsified hypothesis as false.
- Never accept an hypothesis as true - only keep non-falsified hypotheses as so far not-rejected.

Popper is quite radical in this last step. Confirmation places no role at all. One never infer the truth of hypotheses from the observations regarding their implications. Not even increase the confidence in the truth of the hypothesis. Popper hoped to avoid Hume’s problem of induction by not employing induction in science. Popper thought that science was and should be deductive, and therefore that the lack of justification for inductive inferences was not as damaging for science. Below example is illustrative.

Suppose a scientist is testing the hypothesis that all pieces of metal conduct electricity. Even if every piece of metal they examine conducts electricity, this doesn’t prove that the hypothesis is true, for reasons that we’ve seen. But if the scientist finds even one piece of metal that fails to conduct electricity, this conclusively refutes the theory. For the inference from ‘this piece of metal does not conduct electricity’ to ‘it is false that all pieces of metal conduct electricity’ is a deductive inference—the premise entails the conclusion. So if a scientist were trying to refute their theory, rather than establish its truth, their goal could be accomplished without the use of induction.
— (Okasha, 2016)

However, this view of science process could be rather limiting with respect to the actual scientific practice. First, it does not allow to distinguish between non-falsified hypotheses. Popper argues that obtaining evidence in favour of a given theory is generally easy, and holds that such *corroboration* should count scientifically only if it is the positive result of a genuinely *risky* prediction, which

might conceivably have been false.

It is logically impossible to verify a universal proposition by reference to experience (as Hume saw clearly), but a single genuine counter-instance falsifies the corresponding universal law. In a word, an exception, far from “proving” a rule, conclusively refutes it. — (Thornton, 2021)

Second, in scientific practice hypotheses rarely have immediate observable consequences, they often require measurements or experiments to do so. For instance, the hypothesis “this liquid contains 3 consequences” does not entail any direct observable consequence. We might use distillation or chromatography to test such hypothesis but this requires relying on **auxiliary hypothesis** (e.g. the distillation machine works properly). This consideration quite changes the **H-D method** steps. Moreover, we never test a single hypothesis alone, but only in conjunction with various auxiliary hypotheses (Duhem-Quine Thesis). One relevant example is the work of Galileo Galilei and his reports of mountains on the moon and Jupiter satellites. Philosophers such as Cesare Cremonini refused to look through the telescope, arguing that the instrument itself might have introduced artefacts, producing a visual illusion. Therefore, Duhem-Quine thesis states that in order to falsify a hypothesis we must be confident that the responsible for falsity of the consequence are not the auxiliary hypotheses but the main hypothesis.

- 1. Propose a hypothesis H .
- 2. Deduce observable consequences $\{C_i\}$ from H in conjunction with auxiliary hypotheses AH_j
- 3. Test. Look for evidence that conflicts with the predicted consequences $\{C_i\}$ in order to disprove H .
- 4. If $\{C_i\}$ is false, infer that $H \& \{AH_j\}$ is false, reformulate H .
- 5. If $\{C_i\}$ is true, increase confidence in $H \& \{AH_j\}$.

Semantic Note:

Note the difference between *falsifiable* and *falsified*.

Falsifiability is a quality of a hypothesis or a theory. Is the quality of a conjecture or hypothesis to be proven wrong. Some theories have no empirical implications. Popper claimed that astrology and Freud's psychoanalysis were not falsifiable. He argued that *falsifiability* demarcates whether a theory is scientific or not (see the [demarcation problem](#) ([Hansson, 2021](#))). Similarly, some hypotheses might be more falsifiable than others because they have more empirically testable implications. For example, Newton's law of gravitation is falsifiable (e.g. it is falsified by "The brick fell upwards when released").

Falsification is the observation that an implication of a hypothesis is not true which implies (by *modus tollens*) the falsity of the hypothesis. Hypothesis can only be falsified if they are falsifiable.

Falsification uses the valid inference modus tollens: if from a statement P we logically deduce Q , but what is observed is $\neg Q$, we infer that P is false. For example, given the statement "all swans are white" and the initial condition "there is a swan here", we can deduce "the swan here is white", but if what is observed is "the swan here is not white" (say black), then "all swans are white" is false, or it was not a swan.

The take-away message from falsification is that despite proposing an unrealistically restrictive practice of science, it might be a useful inference method for scientists. However, they should be aware of its limitations and for instance, bear in mind the pitfalls of *ad-hoc* modifications. (See negative weight in phlogiston theory ([Grünbaum, 1976](#))). An *ad-hoc* hypothesis is added to a theory to save it from being falsified. A modification is considered *ad-hoc* if it reduces the falsifiability of the hypothesis in question.

2.4.3 Confirmation

Confirmation is the act of using evidence to justify increasing the confidence in the hypothesis. Confirmation is not based on deductively valid inferences. For instance, in the **H-D method** we identify some C that is an implication of H . H implies C , then if H is true we conclude (by *modus ponens*) that C is also true. Moreover, if we observe that C is false, then we conclude (by *modus tollens*) that H is false as well. While these two inference rules are deductively valid, they do not tell us what to conclude if the implication C is true. There is no valid deductive rule that can be used for the case where H implies C and C is true. We cannot deduce anything from that.

Modus ponens	Modus tollens	Induction
H, then C	H, then C	H, then C
H	not C	C
-----	-----	=====
C	not H	H

Instead, any rule used here must amplify the information contained in the premises to infer the conclusions. Therefore we must make use of inductive inferences. Inductive inferences are fallible (inductions that fail are common e.g. predicting the weather, stock investing). But fallibility comes in degrees and this degree is affected by the kind and quality of the evidence as well as the inference rule employed. Scientists have attempted to quantify confidence, most prominently by using probabilities. For instance, if an observation O confirms hypothesis H , therefore we say that $P(H|O)$ is greater than $P(H|\neg O)$ where $P(H|O)$ means “the probability of H given O ”.

There is certain debate on this last point. Not everybody agrees that it makes sense to assign probabilities to hypotheses because they differ on the interpretation of the concept of probability. **Frequentists** interpret the probabilities as the frequencies of repeatable observable events. Therefore probabilities cannot be assigned to hypotheses since these are not events, nor observable or repeatable. Another problem is that probabilities are already used to express a property different from confidence. For instance, we may say that the probability of tails when throwing a coin is 50%. But then someone may ask us how confidence we are about our claim. Even if we can also answer that second question with a probability, is clear that these two numbers express different things.

Note for data scientists!

Is important to note the relevance of frequentist and Bayesian approaches in artificial intelligence. Both frequentist and Bayesian are statistical approaches to learning from data. But there is a broad distinction between the frequentist and Bayesian. The frequentist learning only depends on the given data, while the Bayesian learning is performed by the prior belief as well as the given data (Jun, 2016).

The frequentist computes the probability of result or data D given hypothesis H is true, i.e. $P(D|H)$. In comparison, the Bayesian approach focus on the probability of hypothesis H when the result or data D occurs, i.e. $P(H|D)$ (Orloff and Bloom, 2014).

Understanding that confirmation comes in degrees may help clarify the last step of **H-D method**. Observing C to be true, increases our degree of confidence that H is true. But why is this? A naïve answer to this question is that observing C confirms H because H is compatible with C . But this seems rather weak justification. Indefinitely irrelevant implications could be inferred from a hypothesis. For instance:

```
I have pancreas cancer, I have a pancreas
I have a pancreas
-----
I have pancreas cancer
```

A clear deductive consequence from this example is that indeed I have a pancreas.

However, observing that I do have a pancreas should not confirm the claim that I have pancreas cancer. To solve this issue we should introduce a criteria for relevance to make sure that the chosen implications are relevant to the question. This is key part of the scientific process as this often depends on the domain knowledge we have about the matter we are investigating.

An additional problem to the compatibility issue, is that very many hypotheses are compatible with any given observation. This is called the problem of underdetermination.

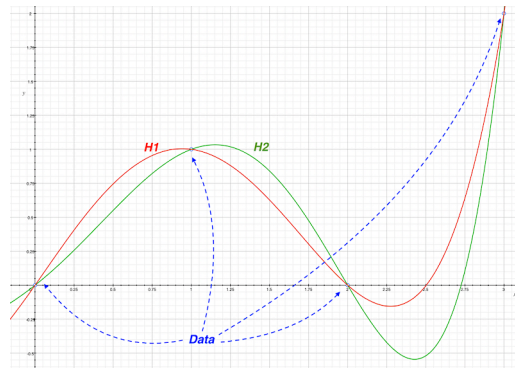


Figure 2.1: The problem of underdetermination illustrated with a chart.

Note for data scientists!

According to *anti-realists*, there will always be multiple *competing* theories about unobservable entities (e.g. atoms) which can account for the data equally well. In other words, such theories are *undetermined* by the empirical data. But then, how do scientists justify choosing one theory over another? *Realists* often reply that aforementioned scenario is only possible in the trivial sense. In fact, scientists often struggle to find even *one* theory that fits the data properly.

But why is this important for data scientists? Often you will find many ML models or solutions that fit your available data or fulfil your requirements, and yet you will have to decide which model/solution is best. If possible, validation with external data and other assessments must be conducted, but sometimes solutions are also chosen based on *non-epistemic* values, such as making society more just or making money.

Is also important to notice how the problem of underdetermination relates to the popular *No Free Lunch Theorem* which is very relevant in the Machine Learning community (Dotan, 2020).

2.4.4 Other types of inference

[complete?]

2.5 Explanation

[continue] [Consider moving this to chapter 3]

At this point, many of you probably have already related data science to two of the goals of science: explanation and prediction. But how do they relate to one another? and what is a scientific explanation? Either to satisfy our natural curiosity or for a further purpose, science has always attempted to understand how the world works. The German philosopher Carl Hempel attempted to answer this question in the 1950s with what is known as the *covering law* model of explanation. He stated that a scientific explanation is an answer given in response to *explanation-seeking why-questions* (e.g. why salt dissolves in water).

According to Hempel, explanations are structured like an argument, i.e. a set of premises followed by a conclusion. Therefore, the conclusion of such an argument states that certain phenomenon occurs, e.g. “salt dissolves in water”. On the other hand, the premises indicate why the conclusion is true. Then, the challenge lays in the relationship that should follow between such premises and the conclusion. For Hempel, the premises should all be true and entail the conclusion, i.e. the argument should be deductive and *sound*. Additionally, the premises should contain at least one general law (e.g. all metals conduct electricity), a.k.a. *laws of nature*. Newton explained the elliptical orbits of planets alluding a general rule (his law of universal gravitation) together with some minor assumptions. This example fits Hempel’s model very well, but not all scientific explanations do.

General Law	(explanans)
Particular Facts	(explanans)

Phenomenon to be explained	(explanandum)

2.6 Examples

2.6.1 The problem is in your hands!

Ignaz Semmelweis, a Hungarian physician, was a member of the First Maternity Division at the Vienna General Hospital from 1844 to 1848. Semmelweis was distressed to find a big proportion of the women who delivered their babies contracted a serious and often fatal illness known as childbed fever. In 1844, 8.2% of mothers died from the disease, 6.8% in 1845 and 11.4% in 1846. However, in the adjacent Second Maternity Division which had as many women as the first, the death toll was much lower (2.3%, 2% and 2.7% respectively).

From this moment on, various explanations were considered, subjected to test and then rejected.

The first explanation attributed the issue to “epidemic influences” described as “atmospheric-cosmic-telluric changes” spreading over districts and causing

childbed fever. This hypothesis did not explain why the first division had more cases than the second. Neither explained the lack of cases in the city of Vienna. Epidemics such as cholera are not so selective. Finally, Semmelweis notes that women who had to give birth in the street on their way to the hospital had a lower death rate than the average for the first division.

On a different view, overcrowding of the first division was proposed as a cause but Semmelweis pointed out that the second division was much crowded. Moreover, there were no differences regarding diet or general care of the patients.

In 1846, a commission was appointed to investigate the issue, which attributed the prevalence to injuries in the first division resulting from rough examination by medical students. Semmelweis refuted this view since: a) the injuries of birth itself are more extensive than those from the examination. b) midwives' examinations from the second division were similar. c) as a consequence of the commission the number of students was halved and the examinations were reduced to a minimum. The mortality increased.

After considering peculiar conjectures (e.g. delivery position, priest visits), an accident gave Semmelweis the decisive clue. In 1847, a colleague of his received a puncture wound in the finger, from the scalpel of a student while performing an autopsy. His colleague died after an illness with similar symptoms to those observed in the victims of childbed fever. Note, that the role of micro-organisms had not yet been recognized at the time. Semmelweis ordered all medical students to wash their hands with a chlorinated lime solution before making examinations, especially after performing dissections in the autopsy room.

Mortality fell to 1.27% in the First Division compared to 1.33% in the second. In further support of his hypothesis, Semmelweis notes that midwives from the Second Division did not dissect cadavers. This also explained the "street births" low mortality since women were rarely examined as they already gave birth. Semmelweis concluded that the cause was infection by cadaveric material and putrid matter.

Example and discussion extracted from Chapter 2 of ([Hempel, 1966](#))

2.6.1.1 How a hypothesis is tested

Some conjectures (e.g. differences in diet, crowding or care) were trivial to test as their assumptions conflict with readily observable facts. Others were not as straightforward and required certain interventions. For example, changing the routine of the priest or the birth position. If the hypothesis H is true, then certain observable events I should occur (e.g. drop in mortality) under specified circumstances (e.g. lateral delivery position). Semmelweis experiment showed the test implication to be false, rejecting the hypothesis in consequence.

If H is true, then so is I .
But (as the evidence shows) I is not true.

H is not true.

This is a good example of *modus tollens* (see 2.2.2). However, let us consider now the case where observation or experiment confirms the test implication *I*. From the hypothesis that childbed fever is blood poisoning produced by cadaveric matter, Semmelweis infers that antiseptic measures will reduce mortality rates. Now, the experiment shows the test implication to be true. But this favourable outcome does not prove the hypothesis true.

If H is true, then so is I.
(as the evidence shows) I is true.

H is true.

This reasoning is deductively invalid and referred to as the *fallacy of affirming the consequent* (see 2.2.2). The conclusion may be false even if its premises are true. Thus, even if many implications of a hypothesis have been confirmed by tests, the hypothesis may still be false.

If H is true, then so are I1, I2, ...
(as the evidence shows) I1, I2, ... are all true.

H is true.

Above's argument still commits the fallacy. Note that although the many tests do not provide conclusive proof for a hypothesis, they provide at least some support or confirmation for it.

Chapter 4 of (Hempel, 1966) continues on this.

In the absence of unfavorable evidence, the confirmation of a hypothesis will normally be regarded as increasing with the number of favorable test findings. [...] **the increase in confirmation effected by one new favorable instance will generally become smaller as the number of previously established favorable instances grows.** If thousands of confirmatory cases are already available, **the addition of one more favorable finding will raise the confirmation but little.**

Note for data scientists!

Notice how Machine Learning (ML) models can also be affected by the previous statement. Many researchers blindly rely on the dogma *the more data, the merrier* but is not just the amount of data that matters but also its variety. The greater the variety, the stronger the resulting support for the trained model.

2.6.2 Wason selection task

Consider the following hypothetico-deductive reasoning problem created by Peter Cathcart Wason employing the logical rule of implication:

You are shown a set of four cards placed on a table, each of which has a number on one side and a colored patch on the other side. The visible faces of the cards show 3, 8, red and brown. Which card(s) must you turn over in order to test the truth of the proposition that if a card shows an even number on one face, then its opposite face is red?



Figure 2.2: Wason selection task or four-card problem.

Hypothesis H: “If a card shows an even number on one face, then its opposite face is red”

Test whether H is false. Which consequences of H do you need to consider - i.e. which cards do you need to turn over? Under what conditions would this statement be false?

These are the possible situations:

- If the 3 card is red (or brown), that doesn’t violate the rule. The rule makes no claims about odd numbers. (Denying the antecedent)
- If the 8 card is not red, it violates the rule. (Modus ponens)
- If the red card is odd (or even), that doesn’t violate the rule. The red color is not exclusive to even numbers. (Affirming the consequent)
- If the brown card is even, it violates the rule. (Modus tollens)

Table 2.1: Truth table for $p \rightarrow q$. (*) In instances of *modus ponens* we assume as premises that $p \rightarrow q$ is true and p is true. Only one line of the truth table — the first — satisfies these two conditions (p and $p \rightarrow q$). On this line, q is also true. Therefore, whenever $p \rightarrow q$ is true and p is true, q must also be true. (**) In instances of *modus tollens* we assume as premises that $p \rightarrow q$ is true and q is false. There is only one line of the truth table — the fourth line — which satisfies these two conditions. In this line, p is false. Therefore, in every instance in which $p \rightarrow q$ is true and q is false, p must also be false.

p	q	$p \rightarrow q$
T	T	*T
T	F	F

p	q	$p \rightarrow q$
F	T	T
F	F	**T

There are two ways to face the problem and reach the solution. First, we can choose the cards based on *modus ponens* and *modus tollens* as follows: From *modus ponens* we need to check the cards that are even. If even cards are not red, then the claim is false.

If even, then red. (claim)
even (obs)

Therefore, red. (conclusion)

From *modus tollens* we need to check the cards that are not red i.e. brown. If brown cards are even, then the claim is false.

If even, then red. (claim)
not red (obs)

Therefore, not even (conclusion)

Another approach is to take the truth table of $p \rightarrow q$ and take the case where $p \rightarrow q$ is false - second line - i.e. when p is true and q is false or for our case, when a card is even and its back not red (so brown). From this we need to take the cards that are p and $\neg q$ i.e. even cards and brown cards.

2.6.3 U.S.A. Presidents



Figure 2.3: Presidents of the United States of America as of 2021.

Suppose we aim to predict whether the next president of the United States of America will be a woman or not. If we rely solely on the gender of previous presidents, by induction we will predict a zero chance. But by understanding how a person becomes a presidential candidate, and how previously became a candidate for their party, we can take into account the network of people involved in the process and recalculate our forecast with higher precision. In this case the rules are clearly defined in the law. Pouring these bits of domain knowledge into our model will show that chances are increasing over time. Encoding the rules behind the data heavily increased the robustness and precision of our model. Thanks to these rules our inference became deductive rather than inductive, since the conclusion necessarily follows from the premises; and as long as the premises are true the conclusion will also be true.

We can identify two issues in the first approach of our example: First, partial data can misrepresent the underlying phenomena that shapes the data, producing a model that does not resemble the real world. This is especially notable in the case of bias and confounders which are further aggravated by the lack of domain knowledge in designing the solutions. The second issue relates to induction. Contrary to deduction, where the truth of the premises guarantees the truth of the conclusion, inductive inferences are *ampliative* — since whose conclusions go beyond what is contained in their premises — and their conclusions could be totally wrong even if infinitely many examples confirm them (Bergadano, 1991). This *ampliative* factor has also an amplifying effect over the partial data from which we infer a conclusion. In this case, considering only the final results of the elections amplified the bias derived from a partial collection of the data, reducing the chances of women being predicted as president to zero.

From (Vega, 2021).

2.6.4 Yersinia pestis

This excerpt from Plague and Cholera is a great example of how laboratory conditions can act as an unintended auxiliary hypothesis that must be taken into account during research. It was 1894 in Hong Kong and all was set for an intellectual duel between Alexandre Yersin and Kitasato Shibasaburō that eventually unveiled the cause of the disease plague.

From the moment of his disembarkation in torrential rain, Yersin sees the bodies of plague victims lying in the street, in pools of standing water, in parks, aboard moored junks. British soldiers, acting on authority, remove the sick and empty their houses, pile everything up and set fire to it. [...]

‘I notice many dead rats lying on the ground.’ The first note scribbled by Yersin that evening concerns sewers spewing out decomposed bodies of rats. Since Camus, that has seemed obvious, but not then. [...] By telegram, and as a concession to diplomacy, British governor Sir William Robinson gives Yersin explicit authority to come and study plague in Hong Kong. However, bad faith on the British side is clear to see, and it is even worse with the Japanese team under Shibasaburo Kitasato, who intends to reserve all autopsies for himself. [...]

Never again, in the history of humanity, will there be such an opportunity to become the person who vanquished plague. A few more weeks of devastation will mean a few thousand more bodies to study. [...] Kitasato, though, has a handicap advantage. Not a single cadaver will be placed at Yersin’s disposal. [...]

For Yersin’s benefit he [Father Vigano] arranges, in just two days, to have a bamboo-framed, straw-covered hut erected near the Al-

ice Memorial Hospital. With the matter of his living quarters and laboratory settled, Yersin installs a camp bed, unlocks the cabin trunk, and sets out microscope and test tubes. Vigano then greases the palms of the British sailors in charge of the hospital mortuary, where the bodies are stacked prior to being cremated or buried, and buys several from them. Yersin proceeds to ply his scalpel. [...] ‘The bubo is quite distinct. In less than a minute I have it out and take it up to my laboratory. I make a quick preparation and place it under the microscope. One glance reveals a veritable mess of microbes, all similar. They are small stubby rods with rounded ends.’ [...] Yersin becomes the first human being to observe the plague bacillus, as Pasteur was the first to observe those of silkworm pebrine, ovine anthrax, chicken cholera and canine rabies.

What Kitasato describes, having sampled organs and blood and disregarded the bubo, is the pneumococcus of a collateral infection, which he mistakes for the plague bacillus. Without luck, without chance, genius is nothing. The agnostic Yersin is blessed by the gods. Subsequent studies will show that one reason for Kitasato’s failure is that he enjoyed the benefits of a proper hospital laboratory, including an incubator set at the temperature of the human body, a temperature at which pneumococcus proliferates, whereas the plague bacillus develops best at approximately twenty-eight degrees centigrade, the mean temperature in Hong Kong at that time of year and the temperature at which Yersin, with no incubator, conducts his observations.

From Plague and Cholera, by Patrick Deville. ([Deville, 2014](#))

I absolutely recommend this book about Alexandre Yersin life. A Swiss-French physician and bacteriologist, pupil of Louis Pasteur, that trying to run away from himself became an agronomist and an explorer of the highlands of Vietnam and Cambodia.

2.6.5 Risks of induction and non-epistemic values in ML

I recommend the following [blog post](#) from Simon Fischer. I copy a fragment here but the whole article is very interesting.

For example, when we think of the problem of *filter bubbles* we are less and less confronted with opposing world views. Moreover, the idea that the future resembles the past, gives us examples of how Amazon has developed an algorithm for recruiting new staff which only hired males (Dastin, 2018). Even though the model might be correct from an epistemological point of view, such as accuracy or simplicity, it questions non-epistemic values, such as fairness. [...]

Another problem arises with regard to Popper’s falsification ap-

proach. We cannot be sure what we have falsified: the hypothesis, the auxiliary assumptions, or even both? Consequently, under these considerations, it appears that the risks of drawing conclusions from machine learning outweigh the benefits. [...]

In the case of Amazon the false hypothesis and background assumptions were found rather quickly. But there could be more subtle biases around us which we are not yet aware. This again shows the twofold consequences in terms of inductive risk: The danger of scientists implementing these biases into the algorithms and the benefit of amplifying these biases, and thus making them visible to us. — (Fischer, 2020)

Chapter 3

Extra Material

This section includes related questions and topics not tackled in this course.

3.1 History of science

Although many historical examples are given to the students to illustrate the different chapters and topics addressed in this course, the history of science is not part of the course curriculum. However, we aim to include a rich but brief summary of the history of science as optional reading in further semesters of this course.

3.2 Theory-relatedness of observations

“A related topic is the theory-relatedness of observations; some have claimed that there are no such things as fully theory-independent observations. If true, it would undermine the possibility of objectivity of science and force us to accept strong relativism. I believe that this disastrous consequence can be avoided and that there really is a basis of theory-neutral data, also in the humanities.” — Philosophy of Science for Scientists.

3.3 Gettier problems

The definition of knowledge is an ongoing debate among epistemologists. Although the three criteria from Plato are necessary conditions, they are not sufficient as there are situations that satisfy all these conditions and yet don't constitute knowledge (see [Gettier cases](#)) but such cases are rather philosophical and will not be discussed during this course.

3.4 Realism and anti-realism

For now this will not be included as part of the course curriculum. For a short account of this topic, read Chapter 4 from ([Okasha, 2016](#)).

Bibliography

- Bergadano, F. (1991). The problem of induction and machine learning. In *IJCAI*.
- Cairo, A. (2019). *How charts lie: Getting smarter about visual information*. WW Norton & Company.
- Déville, P. (2014). *Plague and cholera*. Hachette UK.
- Díez, J. A. and Moulines, C. U. (1997). Fundamentos de filosofía de la ciencia.
- Dotan, R. (2020). Theory choice, non-epistemic values, and machine learning. *Synthese*, pages 1–21.
- Douven, I. (2021). Abduction. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press.
- Fischer, S. (2020). The necessity of non-epistemic values in machine learning modelling.
- Grüne-Yanoff, T. (2014). Teaching philosophy of science to scientists: why, what and how. *European Journal for Philosophy of Science*, 4(1):115–134.
- Grünbaum, A. (1976). Ad hoc auxiliary hypotheses and falsificationism. *The British Journal for the Philosophy of Science*, 27(4):329–362.
- Hansson, S. O. (2021). Science and Pseudo-Science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition.
- Hempel, C. G. (1966). *Philosophy of natural science*. Prentice-Hall Englewood Cliffs, N.J.
- Henderson, L. (2020). The Problem of Induction. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2020 edition.

- Hume, D. (1739). *A Treatise Upon Human Nature*. Oxford University Press, Oxford.
- Johansson, L.-G. et al. (2016). *Philosophy of science for scientists*. Springer.
- Jun, S. (2016). Frequentist and bayesian learning approaches to artificial intelligence. *International Journal of Fuzzy Logic and Intelligent Systems*, 16(2):111–118.
- Loukides, M., Mason, H., and Patil, D. (2018). *Ethics and data science*. O’Reilly Media.
- Nidditch, P. H. (1968). *The philosophy of science; edited by P. H. Nidditch*. Oxford U.P London.
- Okasha, S. (2016). *Philosophy of science: very short introduction*. Oxford University Press.
- Orloff, J. and Bloom, J. (2014). Comparison of frequentist and bayesian inference. class 20, 18.05, spring 2014.
- Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books, Inc., USA, 1st edition.
- Rosenberg, A. and McIntyre, L. (2019). *Philosophy of science: A contemporary introduction*. Routledge.
- Russell, B. (1912). *The problems of philosophy*.
- Thornton, S. (2021). Karl Popper. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition.
- Vega, C. (2021). From hume to wuhan: an epistemological journey on the problem of induction in covid-19 machine learning models and its impact upon medical research. *IEEE Access*, 9:97243–97250.