Probability and induction.

'Probability' is an ambiguous word. In the history of ideas, it has been used with many different senses giving rise to different concepts of probability. Being associated with games of chance and gambling, death tolls and insurance policies, statistical inferences and the chancy world of modern physics, probabilities have been made susceptible to different interpretations. These interpretations may reflect on probabilities the objectivity of logic or the subjectivity of a person's belief and lack of knowledge, the frequencies of observed data or the real tendency of a system to yield an outcome. Commonly, but not always, are considered to be interpretations of the mathematical concept of probability which by itself and in itself has no empirical meaning.

The article attempts to present the different meanings of 'probability' and provide an introductory topography of the conceptual landscape. Without trying to provide a history of the idea, historical elements have been considered. Also, realizing that an exhaustive treatment would be difficult, we are focusing, mainly, on the discussion of induction and confirmation. The article is intended as a companion to another entry on IEP in which we discuss The Problem of Induction (Psillos and Stergiou, 2022); this explains why we do not deal here with Hans Reichenbach's major contribution to the interpretation of probability theory.

Table of Contents

1. Elements of Probability and its Interpretations
a. On Mathematical Probability

In the monograph *Foundations of the Theory of Probability,* first published in German in 1933, the Soviet mathematician A. N. Kolmogorov presented the definitive form of what is nowadays regarded an axiomatization of mathematical probability. The challenge of axiomatization has been set by D. Hilbert in the sixth of his famous twenty-three problems at the beginning of twentieth century (1902):

> …to treat in the same manner [as geometry], by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics.

Kolmogorov, addressing the problem, developed a theory of probability as a mathematical discipline "from axioms in exactly the same way as Geometry and Algebra" (1933:1). In his axiomatization, probability and the other primary concepts, devoid of any empirical meaning, are defined implicitly in terms of consistent and independent axioms in a set-theoretic setting. Thus, modern mathematical probability theory grew within the branch of mathematics called measure theory.

Kolmogorov called *elementary theory of probability* "that part of the theory in which we have to deal with probabilities of only a finite number of events." (*ibid*). A *random event* is an element of an *event space*; the latter being formalized by the set-theoretic concept of *field,* introduced by Hausdorff in *Set Theory* (1927). A field is a non-empty collection of subsets $\mathcal{S}$ of a given non-empty set Γ that has the following properties:

(a) for every pair of elements, A, B of $\mathcal{S}$, their union, $A \cup B$, belongs in $\mathcal{S}$;

(b) for every element A of $\mathcal{S}$ its complement with respect to Γ, $A^c = \Gamma \backslash A$, is in $\mathcal{S}$.

In probability theory the set Γ is called *sample space.*

To understand the above formalization, consider the simple example of tossing a die. Let Γ be the set of the six possible outcomes:

$$E_1, E_2, E_3, E_4, E_5, E_6.$$

The collection $\mathcal{S}$ of all $2^6 = 64$ subsets of Γ,

$$\emptyset, \{E_1\}, \{E_2\}, \ldots, \{E_6\}, \{E_1, E_2\}, \{E_1, E_3\}\ldots, \{E_5, E_6\}, \{E_1, E_2, E_3\}, \ldots, \{E_4, E_5, E_6\},$$
$$\{E_1, E_2, E_3, E_4\}, \ldots, \{E_3, E_4, E_5, E_6\}, \{E_1, E_2, E_3, E_4, E_5\}, \ldots, \{E_2, E_3, E_4, E_5, E_6\},$$
$$\{E_1, E_2, E_3, E_4, E_5, E_6\},$$

satisfies conditions (a) and (b); $\mathcal{S}$ is a field. The subsets of Γ represent different possibilities that can be realized in tossing a single die: the empty set, $\emptyset$, is a random event that represents an impossible happening. The singletons, $\{E_1\}, \{E_2\}, \ldots, \{E_6\}$, are the *elementary events*, since any other random event (except $\emptyset$) is a disjunction of these events, expressed by taking the set-theoretic union of the respective singletons. Finally, $\Gamma = \{E_1, E_2, E_3, E_4, E_5, E_6\}$, is an event that represents the realization of any possibility.

A function from a field $\mathcal{S}$ to the set of real numbers, $\mathbb{R}$,

$$p: \mathcal{S} \to \mathbb{R},$$

is called a *probability function* on $\mathcal{S}$, if it satisfies the following axioms:

i. $p(A) \geq 0$, for $A \in \mathcal{S}$;

ii. $p(\Gamma) = 1$;

iii. $p(A \cup B) = p(A) + p(B)$, for $A \cap B = \emptyset$;

In the simple example of tossing a die, a probability function p would assign a non-zero real number $p(E)$ to each element E of $\mathcal{S}$, according to axiom (i). Axiom (ii) requires that the random event which describes any possible outcome has probability 1, $p(\Gamma) = 1$. Axiom (iii), commonly called *finite additivity* property, tells us how to calculate the probability value of any random event from the probability values of elementary events, for instance,

$$p(\{E_1, E_2, E_3, E_4\}) = p(\{E_1, E_2\}) + p(\{E_3, E_4\}) = p(\{E_1\}) + p(\{E_2\}) + p(\{E_3\}) + p(\{E_4\}).$$

Notice that there are infinitely many admissible probability functions on the event space of the tossing of a die and that only one of them corresponds to a *fair* die, the one with $p(\{E_i\}) = \frac{1}{6}$.

Problems concerning a countably infinite number of random events require an additional axiom and the formalization of the event space as a *σ-field*. A field $\mathcal{S}$ is a σ-field if and only if it satisfies the following condition:

    (c) for every infinite sequence of elements of $\mathcal{S}$, $\{A_n\}_{n \in \mathbb{N}}$, the countably infinite union of these sets, $\bigcup_{n=1}^{\infty} A_n$ belong in $\mathcal{S}$.

Every field $\mathcal{S}$ of finite cardinality is a σ-field since any infinite sequence in $\mathcal{S}$ consists of a finite number of different subsets of Γ and their union is always in $\mathcal{S}$, according to (a). Yet this may not be the case if the field is constructed from a countably infinite set Γ. Imagine, for instance, a die of infinite faces, where the set Γ of possible outcomes is:

$$E_1, E_2, E_3, \ldots$$

Let the collection $\mathcal{S}$ consist of subsets A of Γ which are either of finite cardinality or their complement, $A^c = \Gamma \backslash A$, is of finite cardinality:

$$\mathcal{S} = \{A \subset \Gamma : A \text{ is finite or } A^c \text{ is finite }\}.$$

It's easy to show that $\mathcal{S}$ is a field. Yet it is not a σ-field, since the set

$$\bigcup_{n \in \mathbb{N}} \{E_{2n}\}$$

which is the infinite union of $\{E_{2n}\}$, $n \in \mathbb{N}$ does not belong to $\mathcal{S}$.

A probability function on a σ-field $\mathcal{S}$,

$$p : \mathcal{S} \to \mathbb{R},$$

satisfies the following axioms:

i'. $p(A) \geq 0$, for $A \in \mathcal{S}$;

ii'. $p(\Gamma) = 1$;

iii'. $p(\bigcup_{n=1}^{\infty} A_n) = p(A_1) + \cdots + p(A_N) + \cdots = \sum_{n=1}^{\infty} p(A_n)$, for $A_i \cap A_j = \emptyset$, for $i \neq j$.

It is evident that axiom (iii′), commonly called *countable additivity* property of the probability function, extends finite additivity to the case of a countably infinite family of events. Originally, Kolmogorov suggested a different axiom, equivalent to countable additivity, the *axiom of continuity* (1933: 14):

iii′′. For a monotone sequence of events $\{A_n\}_{n\in\mathbb{N}}$, with $A_n \supseteq A_{n+1}, n \geq 1$ such that $\bigcap_{n=1}^{\infty} A_n = \emptyset$, $p(A_n) \longrightarrow 0$ when $n \to \infty$.

In what follows we will see that many interpretations of mathematical probabilities are actually interpretations of elementary probability theory, and that they face serious problems when applied to mathematical probability theory formulated in σ-fields.

A special probability function $p(\bullet|A): \mathcal{S} \to \mathbb{R}$ can be defined on $\mathcal{S}$, if we are given a function p on $\mathcal{S}$ and a random event $A \in \mathcal{S}$ such that $p(A) \neq 0$:

$$p(B|A) = \frac{p(B \cap A)}{p(A)}, \text{ for } B \in \mathcal{S}$$

$p(\bullet|A)$ determines the *conditional probability $p(B|A)$ of some event $B \in \mathcal{S}$ given an event A,* while $p(B)$ is the *unconditional probability* of B.

The conditional probability given an event $A \in \mathcal{S}$ of any random event $B \in \mathcal{S}$, $p(B|A)$, can be understood as unconditional probability of an event D, $p_A(D)$, determined by a probability function p_A on a reduced event space $\mathcal{S}_A$ consisting of subsets of the event $A \in \mathcal{S}$ we conditionalize on; namely, $p_A: \mathcal{S}_A \to \mathbb{R}$, $p_A(D) = p(B|A)$, where $\mathcal{S}_A = \{D: D = B \cap A, \text{ for } B \in \mathcal{S}\}$.

In the tossing of a fair die example, the conditional probability of any outcome, event $B = \{E_i\}, i = 1, \dots 6$, given that it is an even number, event $A = \{E_2, E_4, E_6\}$, is provided by the conditional probability function $p(\bullet|A)$, defined on the σ-field $\mathcal{S}$. Since the die is fair, $p(\{E_i\}) = \frac{1}{6}$ for $i = 1, \dots 6$; also, $p(B \cap A) = \frac{1}{6}$ for $B = \{E_i\}, i = 2,4,6$, while $p(B \cap A) = 0$ otherwise; using the finite additivity axiom, $p(A) = p(\{E_2\}) + p(\{E_4\}) + p(\{E_6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$; so, $p(B|A) = \frac{1}{3}$, for $B = \{E_i\}, i = 2,4,6$, and $p(B|A) = 0$ otherwise. Now, consider the reduced event space $\mathcal{S}_A$ consisting of the subsets of $\{E_2, E_4, E_6\}$. Since the die is fair, $p_A(\{E_i\}) = \frac{1}{3}$ for $i = 2,4,6$ and, $p_A(B) = \frac{1}{3} = p(B|A)$ for $B = \{E_i\}, i = 2,4,6$, while $p_A(\emptyset) = 0 = p(B|A)$ otherwise.

Kolmogorov's axiomatic account, the standard mathematical textbook account of probability theory, explicates the concepts of random event and event space in terms of set theory. Yet, Boole proposed

> … another form under which all questions in the theory of probabilities may be viewed; and this form consists in substituting for events the propositions which assert that those events have occurred, or will occur; and viewing the element of numerical probability as having reference to the truth of those propositions, not to the occurrence of the events concerning which they make assertion. (1853:190)

This formulation of probability theory is very common in philosophical contexts, especially when discussing inductive inference. It typically concerns *elementary* probability theory, presented in the language of sentential logic. Elements of this account can be found in Appendix 6.a and the reader may also consult (Howson and Urbach 2006: Ch.2). Here, we present just a few propositions of elementary probability theory as formulated in this setting that will be found useful in what follows:

- Probability 1 is assigned to tautologies and probability 0 to contradictions. All other sentences have probability values between 0 and 1.
- The probability of the negation of sentence a is $1 - p(a)$.
- The probability of the disjunction of two inconsistent sentences a, b is the sum of probabilities of the sentences:

$$p(a \vee b) = p(a) + p(b).$$

- The conditional probability of a sentence a given the truth of a sentence b is:
$$p(a|b) = \frac{p(a \wedge b)}{p(b)}, p(b) \neq 0.$$

- Bayes's Theorem. The *posterior probability* of a hypothesis h – i.e., the probability of h conditional on evidence e – is:
$$p(h|e) = \frac{p(e|h)p(h)}{p(e)}, \text{where } p(h), p(e) > 0,$$
where $p(e|h)$ is called *likelihood* of the hypothesis and expresses the probability of the evidence conditional on the hypothesis; $p(h)$ is called *prior probability* of the hypothesis; and $p(e)$ is the probability of the evidence.

We conclude this brief introduction to mathematical probability with the following instructive application of Bayes's theorem. A factory uses three engines A_1, A_2, A_3 to produce a product. The first engine, A_1, produces 1000 items, the second, A_2, 2000 items and the third, A_3, 3000 items, per day. Of these items, 4%, 2% and 4%, respectively, are faulty. What is the probability of a faulty product having been produced by a given engine in a day? Let h_i be the hypothesis: "A product has been produced by engine A_i in a day", for $i = 1,2,3$, and e: "A faulty product has been produced in a day". Then the prior probabilities of h_i are, $p(h_1) = \frac{1}{6}$; $p(h_2) = \frac{1}{3}$; $p(h_3) = \frac{1}{2}$ and the likelihoods are $p(e|h_1) = 0.04, p(e|h_2) = 0.02$; $p(e|h_3) = 0.04$, respectively. Using the theorem of total probability (see, Appendix 6a), we can calculate $p(e) = p(h_1)p(e|h_1) + p(h_2)p(e|h_2) + p(h_3)p(e|h_3) = \frac{1}{6} \cdot 0.04 + \frac{1}{3} \cdot 0.02 + \frac{1}{2} \cdot 0.04 = \frac{1}{30}$. By applying Bayes's theorem we obtain the posterior probability for each hypothesis, $p(h_1|e) = 0.20$; $p(h_2|e) = 0.20$; $p(h_3|e) = 0.60$, that is, the probability of a faulty product to have been produced by a given engine in a day.

b. Interpretations of probabilities

As any other part of mathematics, probability theory does not have on its own any empirical meaning and cannot be applied to games of chance, to the study of physical or biological systems, to risk evaluation or insurance policies and, in general, to empirical science and practical issues, unless we provide an interpretation of its axioms and theorems. This is what Wesley Salmon (1966: 63) dubbed *the philosophical problem of probability*:

> It is the problem of finding one or more interpretations of the probability calculus that yield a concept of probability, or several concepts of probability, which do justice to the important applications of probability in empirical science and in practical affairs. Such interpretations whether one or several would provide an explication of the familiar notion of probability.

Salmon suggested three criteria that an interpretation of probability is desirable to satisfy. The first one is called *admissibility,* and it requires that the probability concepts satisfy the mathematical relations of the calculus of probability, i.e., the axioms of Kolmogorov. This is a minimal requirement for the concept of probability to be an interpretation of mathematical probability but not a trivial one, since countable additivity may be a problem for some interpretations of probability (see, 2.a.i and 2.b), while in others, Kolmogorov's axioms are supposed to follow naturally

from the practice of gambling (see, 5.a and 5.b). The second criterion is *ascertainability*. It requires that there should be a method by which, in principle at least, we can ascertain values of probabilities. If it is impossible to find out what the values of probability are, then the concept of probability is useless. Again, not all suggested interpretations satisfy this requirement. According to Salmon, Reichenbach's frequency interpretation fails to meet this requirement (1966: 89ff.). Finally, *applicability* is the third criterion: a concept of probability should be applicable, i.e., it should have a practical predictive significance. The force of this criterion is manifested in everyday life, in science as well as in the logical structure of science. The concept of scientific confirmation provides a venerable example of application of probability theory.

Interpretations of probability theory may be classified under two general families: *inductive* and *physical probability*. The classical, the logical and the subjective interpretations of probability are deemed inductive, while the frequency and the propensity interpretations yield physical probabilities. To illustrate the difference between inductive and physical probability, an example may be instructive (Maher, 2006). Think of a coin that you know is either two-headed or two-tailed, but you have no information about what it is. What is the probability that it would land heads, if tossed? One possible answer would be that the probability is $\frac{1}{2}$, since there are two possibilities, and we have no evidence which one is going to be realized. Another answer would say that the probability is either 0, if the coin is two-tailed, or 1, if two-headed, but we do not know which. Maher suggests that if '$\frac{1}{2}$' occurs as a natural answer, then we understand 'probability' in the sense of inductive probability while the sense in which '0 or 1' occurs as a natural answer is *physical probability*. What is the difference between the two meanings? Inductive probability is relative to available evidence, and it does not depend on how the unknown part of the world is, i.e., on unknown facts of the matter. Thus, if in this example we come to know that the coin tossed has a head on one side, we should revise the probability estimate in the light of new evidence and claim that now the inductive probability is 1. On the other hand, physical probability is not relative to evidence, and it depends on facts that may be unknown. This is why the further piece of information we entertained does not alter the physical probability (it is still '0 or 1').

2. What is Probability?
a. The Classical Interpretation

Pierre Simon Laplace proposed what has come to be known as the *classical interpretation* in his work, *The Analytical Theory of Probabilities* (1812), and in the much shorter, *A Philosophical Essay on Probabilities* (1814); a book based on a lecture on probabilities he delivered in the *Ecole Normale*, in 1795. His deterministic view of the universe, Laplacian determinism, is legendary. Not only did he believe that every aspect of the world, any event that takes place in the universe is governed by the *principle of sufficient reason "…the evident principle that a thing cannot occur without a cause which produces it"* (1814: 3) but also that "[w]e ought … to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow." (1814: 4). Moreover, he claimed that the universe is knowable, in principle, and that a supreme intelligence that:

> could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula

the movements of the greatest bodies of the universe and those of the lightest atom. (*ibid*)

However, human intelligence is weak. It cannot provide an adequate unified picture of the world and subsume the macroscopic and microscopic realm under the province of a single formula. Nor can it give the causes of all events that occur and render them predictable. Thus, ignorance emerges as an expression of human limitation. Laplace stressed that:

> [t] he curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance. (1814: 6)

Due to ignorance of the true causes, he claimed, people believe in final causation, or they make chance ('hazard' in Laplacian terminology) an objective feature of the world. "[B]ut these imaginary causes" explains Laplace, "have gradually receded with the widening bounds of knowledge and disappear entirely before sound philosophy, which sees in them only the expression of our ignorance of the true causes." (1814: 3)

i. Probability as a Measure of Ignorance

In this context, Laplace interpreted probability as a measure of our ignorance making it dependent on evidence one is aware of, or, on lack of such evidence:

> Probability is relative, in part to this ignorance, in part to our knowledge. We know that of three or a greater number of events a single one ought to occur; but nothing induces us to believe that one of them will occur rather than the others. In this state of indecision, it is impossible for us to announce their occurrence with certainty. It is, however, probable that one of these events, chosen at will, will not occur because we see several cases equally possible which exclude its occurrence, while only a single one favors it. (1814: 6)

The measure of probability of an event is determined by considering *equally probable* cases that either favor or exclude its occurrence and the concept of probability is reduced to the notion of equally probable events:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible. (1814: 6-7)

Laplace claims that the probability of an event is the ratio of the number of favorable cases to that of all possible cases. And this *principle* of the calculus of probability has for Laplace the status of a definition:

> First Principle.—The first of these principles is the definition itself of probability, which, as has been seen, is the ratio of the number of favorable cases to that of all the cases possible. (1814: 11)

In the jargon of the mathematical theory of probability, one may consider a *partition* $\{A_k\}_{k=1...n}$ of the event space $\mathcal{S}$, i.e. a family of mutually exclusive subsets exhaustive of the sample space, $A_i \cap A_j = \emptyset$ and $\cup_{k=1}^{n} A_k = \Gamma$ – and assume equal probability for all random events A_k, $p(A_i) = p(A_j)$, for every $1 \le i, j \le n$.

Now, for every event E that is decomposable into any sub-family $\left\{A_{k_l}\right\}_{l=1...m} \subseteq \{A_k\}_{k=1...n}$,

$$E = \bigcup_{l=1}^{m} A_{k_l},$$

the probability of E is,

$$p(E) = \frac{m}{n} = \frac{number\ of\ favorable\ cases\ for\ E}{number\ of\ possible\ cases}.$$

We can easily show that a function defined in this way satisfies the axioms of elementary probability theory: $p(A) \ge 0$, for $A \in \mathcal{S}$; $p(\Gamma) = 1$; $p(A \cup B) = p(A) + p(B)$, for $A \cap B = \emptyset$. Hence, Laplace's first principle suggests an admissible, in Salmon's sense, interpretation of the *elementary* theory.

Countable additivity (axiom iii´), on the other hand, is not satisfied for an event space of countably infinite cardinality. To show this, consider an infinite partition $\{A_k\}_{k=1...\infty}$ and assign equal probability to all A_ks, $p(A_k) \ge 0$. Then by employing axioms i´ and ii´ along with the equal probability condition and countable additivity (axiom iii´), we are led to the following absurdity:

$$1 = p(\Gamma) = p(\bigcup_{k=1}^{\infty} A_k) =^{Count.Addit.} \sum_{k=1}^{\infty} p(A_k) = or \begin{matrix} 0 \\ \infty \end{matrix}$$

Hence, classical interpretation is not an admissible interpretation of the mathematical theory of probability in general. It singles out only certain models of probability theory (elementary theory) in which the cardinality of the event space is finite.

Another criticism raised against the classical interpretation (Hajek, 2019) is related to its applicability. The classical interpretation of probability allows only rational-valued probability functions, defined in terms of a ratio of integers. However, in many branches of science, theories (for instance, quantum mechanics) assign to events irrational probability values. In these cases, one cannot interpret probability value in terms of the ratio of the number of favorable, over the total number of cases.

As we have already discussed, in the definition of probability, Laplace presupposes that all cases are equally probable. This fact gives rise to a well-known criticism, namely, that of circularity of the definition of probability: if the relation of equiprobability of two events depends conceptually on what probability is, then the definition of probability is circular. To avoid this criticism, the soviet mathematician and student of Kolmogorov Boris Gnedenko, considered the notion of equal probability a primitive notion "which is …basic and is not subject to a formal definition." (1978: 23)

Laplace, in several places, wrote about "equally possible" cases as if 'possibility' and 'probability' were terms that could be used interchangeably. To assume that is to commit a category mistake, as Hayek has pointed out, since possibilities do not come in degrees. Nevertheless, as we shall see in section 3.a.1, the connection between possibility and probability can be established in terms of Keynes's *principle of*

indifference. In the same section we will discuss the paradoxes of indifference that also undermine Laplace's idea of probability.

b. Probabilities as Frequencies

The frequency interpretation of probability can be traced back to the work of R. L. Ellis and John Venn, in the middle of nineteenth century and it has been described as "a 'British Empiricist' reaction to the 'Continental rationalism' of Laplace" (Gillies 2000: 88). In Ellis's article "On the Foundations of the Theory of Probability" (1842) we identify the rudiments of this interpretation:

> If the probability of a given event be correctly determined, the event will, on a long run of trials, tend to recur with frequency proportional to this probability.

Venn presented his own account, a few years later, in 1866, in *The Logic of Chance:*

> we may define the probability or chance … of the event happening in that particular way as the numerical fraction which represents between the two different classes in the long run. (the quote is from 3rd edition, 1888: 163)

The real boost, however, for the frequency interpretation has been given in the early twentieth century, with the advent of Logical Empiricism, by Richard von Mises, in Vienna, and Hans Reichenbach, in Berlin. The first, in his work *Probability, Statistics and Truth*, published in German in 1928, provides a thorough mathematical and operationalist account of probability theory as empirical science, alike empirical geometry and the science of mechanics. The account has been presented more rigorously in von Mises' posthumously published work, entitled *Mathematical Theory of Probability and Statistics* (1964). Reichenbach presented his mature views on probability in the work *The Theory of Probability: an inquiry into the logical and mathematical foundations of the calculus of probability* originally published in Turkey, in 1935. In this work, Reichenbach attempted to establish a probability logic, based on the relation of probability implication, which is governed by four axioms. Relative frequencies of sub-series of events in a larger series are interpreted as probabilities and they are shown to satisfy the axioms of probability logic. However, Reichenbach's milestone contribution concerns the connection between probability theory and the problem of induction. In this section, we will focus, mainly, on the frequency interpretation of probability as suggested by von Mises while for Reichenbach's views the reader may consult our IEP entry on The Problem of Induction (Psillos and Stergiou, 2022).

Von Mises claimed that the subject matter of probability theory are the repetitive events – "same event that repeats itself again and again" - and the mass phenomena – "a great number of uniform elements … [occurring] at the same time" (1928: 11). Probability, according to von Mises, is defined in terms of a *collective*, a concept which "denotes a sequence of uniform events or processes which differ by certain observable attributes, say colors, numbers or anything else" (1928: 12). For example, take a plant coming from a given seed as a single instance of a collective which consists of a large number of plants coming from the given type of seed. All members of the collective differ from each other with respect to some attribute, say the color of the flower or the height of the plant. Respectively, in the case of tossing a die the collective consists of the long series of tosses and the attribute which distinguishes the instances is the number that appears on the face of the die. The mathematical

representation of such finite empirical collectives is given in terms of their idealized counterpart, the infinite ordered sequences of events, which exhibit attributes that are subsets of the *attribute space* of the collective (which is no different from what we have called *sample space*).

Yet, to be an empirical collective, a sequence of events should satisfy two empirically well-confirmed laws that dictate the mathematical axioms of probability theory in the ideal case of the infinite sequences. The first law, dubbed by Keynes (1921: 336), *Law of Stability of Statistical Frequencies,* requires that:

> the relative frequencies of certain attributes become more and more stable as the number of observations is increased. (von Mises 1928: 12)

Thus, if Ω is the attribute space, $A \subseteq \Omega$ is an attribute and $m(A)$ is the number of manifestations of A in the first n members of the collective, the relative frequency, $\frac{m(A)}{n}$, tends to a fixed number as the number n of observations increases. According to von Mises, the Law of Stability of Statistical Frequencies is confirmed by observations in all games of chance (dice, roulette, lotteries, etc.), in data from insurance companies, in biological statistics, and so on (von Mises 1928: 16-21). This empirical law gives rise to the axiom of convergence for infinite sequences of events: for an arbitrary attribute A of a collective C, $\lim_{n\to\infty} \frac{m(A)}{n}$ exists.

This law can be traced back to the views of von Mises's predecessors. For instance, Venn thought that probability is about "a large number or succession of objects, or, as shall term it, *series* of them" (1888: 5). This series should be 'indefinitely numerous' and it should "combine[s] individual irregularity with aggregate regularity" (1888: 4). All series, for Venn, initially exhibit irregularity, if one considers only their first elements, while, subsequently, a regularity may be attested. This regularity, however, can be unstable and it can be destroyed in the long run, in the "ultimate stage" of the series. According to Venn, a series is of the *fixed type* if it preserves the uniformity while it is of the *fluctuating type* if "the uniformity is found at last to fluctuate." (1888: 17). Probability is defined only for series of the fixed type; if a series is of the fluctuating type, it is not the subject of science (1888: 163). But what does it mean, in terms of relative frequencies, that a series is of the fixed type? "The one [fixed type] tends without any irregular variation towards a fixed numerical proportion in its uniformity". (*ibid*).
In more detail:

> [a]s we keep on taking more terms of the series we shall find the proportion still fluctuating a little, but its fluctuations will grow less. The proportion, in fact, will gradually approach towards some fixed numerical value, what mathematicians term its limit. (1888: 164)

The second presupposition for a sequence to be a collective is an original contribution of von Mises. Apart from the existence of limiting relative frequencies in infinite sequences, he demanded the sequence to be *random* in the sense that there is no rule-governed selection of a subsequence of the original sequence that would yield a different relative frequency of the attribute in question from the one obtained in the original sequence. In von Mises (1957: 29) own words:

> …these fixed limits are not affected by place selection. That is to say, if we calculate the relative frequency of some attribute not in the original sequence, but in a partial set, selected according to some fixed rule, then we require that the relative frequency so calculated should tend to the same limit as it does in the original set… The fulfilment of the condition…will be as the Principle of Randomness or the Principle of Impossibility of a Gambling System.

In a more detailed account of how the subsequence is obtained by place selection, von Mises (1964: 9) explained that in inspecting all elements of the original sequence, the decision to keep the nth element in or to reject it from the subsequence depends either on the ordinal number n of this element or on the attributes manifested in the $(n-1)$ preceding elements. This decision does not depend on the attribute exhibited by the nth or by any subsequent element.

Von Mises suggested that we should understand the Principle of Impossibility of a Gambling System by analogy to the Principle of Conservation of Energy. As the energy principle is well-confirmed by empirical data about physical systems, so the principle of randomness is well-confirmed for random sequences manifested in games of chance and in data from insurance companies. Moreover, as the principle of conservation of energy prohibits the construction of a perpetual motion machine, the principle of impossibility of a gambling system prohibits the realization of a rule-governed strategy in games of chance that would yield perpetual wealth to the gambler:

> We can characterize these two principles, as well as all far-reaching laws of nature, by saying that they are restrictions which we impose on the basis of our previous experience, upon our expectation of the further course of natural events. (1928: 26)

Having defined the concept of a collective that is appropriate for the theory of probability in terms of the two aforementioned laws, we may, now, define the 'probability of an attribute A within a given collective C' in terms of the limiting value of relative frequency of the given attribute in the collective:

$$p_C(A) = \lim_{n\to\infty} \frac{m(A)}{n}.$$

Thus defined, probabilities are always conditional to a given collective. Does, however, this definition provide an admissible concept of probability in compliance with Kolmogorov's axioms?

It is straightforward that axioms (i) and (ii) are satisfied. Namely, since for every $n \in \mathbb{N}, 0 \leq \frac{m(A)}{n} \leq 1$, it follows that $0 \leq p_C(A) \leq 1$. And if the attribute examined consists in the entire attribute space Ω then it will be satisfied by any member of the sequence, $\frac{m(\Omega)}{n} = \frac{n}{n} = 1$, so, taking limits, $p_C(\Omega) = 1$.

Regarding the axiom of finite additivity, (iii), we have that for any pair of mutually exclusive attributes, A, B, the number of times that either A or B occurs is the sum of the occurrences of A and B, since the two cannot occur together:

$$m(A \cup B) = m(A) + m(B) \Rightarrow \frac{m(A\cup B)}{n} = \frac{m(A)}{n} + \frac{m(B)}{n}.$$

By taking limits:

$$p_C(A \cup B) = p_C(A) + p_C(B).$$

However, von Mises' concept of probability does not satisfy the axiom of countable additivity (axiom iii′). To show that, consider the following infinite attribute space $\Omega = \{A_1, \dots, A_k, \dots\}$ and assume that each attribute A_k appears only once in the course of an infinite sequence of repetitions of the experiment, then $p_C(A_k) = 0$, for every $k \in \mathbb{N}$. If the countable additivity condition were true, then $p_C(\Omega) = p_C(A_1) + \dots + p_C(A_k) + \dots = 0$. However, this is absurd, since it violates the normalization condition $p_C(\Omega) = 1$. To provide a probability theory that satisfies all Kolmogorov axioms, von Mises restricted further the scope of a collective. In addition to the Law of Stability of Statistical Frequencies and the Principle of Randomness, in his *Mathematical Theory of Probability* he required a third, independent, condition that a collective should satisfy (von Mises 1964: 12). Namely, that for a denumerable attribute space $\Omega = \{A_1, \dots, A_k, \dots\}$:

$$\sum_{k=1}^{\infty} \lim_{n \to \infty} \frac{m(A_k)}{n} = 1.$$

To define conditional probability, we may begin with a given collective C and pick out all elements that exhibit some attribute B. Assuming that they form a new collective C_B, we calculate the limiting relative frequency $p_{C_B}(A) = \lim_{n \to \infty} \frac{m(A)}{n}$ in C_B.

The conditional probability of A given B in the collective C is then:

$$p_C(A|B) = p_{C_B}(A).$$

In case attribute B is manifested only a finite number of times in C, then C_B is a set of a finite cardinality; hence, it does not qualify as a collective and conditional probability is not defined. To avoid this ill-defined case, Gillies suggested that we require that $p_C(B) \neq 0$. Given this condition he shows all prerequisites for C_B to be a collective are satisfied and conditional probability can be defined (Gillies, 2000:112).

Von Mises's account of probability has been criticized as being too narrow with respect to the common use of the term 'probability': there are important situations in which we apply the term although we cannot define a collective. Take for instance, von Mises's question "Is there a probability of Germany being at some time in the future involved in a war with Liberia?" (1928: 9) Since we do not refer to repetitive or mass events, we cannot define a collective and, in the frequency interpretation, the question is meaningless, since 'probability' is meaningfully used only with reference to a collective. Hence, many common uses of 'probability' in ordinary language become illegitimate if we think in terms of the empirical science of probability as delineated by von Mises.

Some may think that this is not an objection at all: von Mises explicates probability in a way that legitimizes only *some* uses of the term as it occurs in ordinary language and in this way he deals with *the problem of single-case probabilities* that burdens the frequency interpretation: associating probability with (limiting) relative frequency

yields trivial certainty (probability equal to 1) for all unrepeated or unrepeatable events. The solution offered by von Mises is to exclude definitionally such events from the domain of application of the concept of probability.

Of course, there are alternative ways to understand probability, not as relative frequency, that render its use to unrepeated or unrepeatable events legitimate. Take for instance the subjectivist account (see section 5), which considers probability as a measure of the degree of belief. In this conception, the question acquires meaning requesting the degree of belief an agent would assign to that proposition. In addition, to be on the safe side and avoid paradoxes, one may request coherence from the agent, i.e., that their degrees of belief satisfy Kolmogorov's axioms of probability.

A criticism raised against von Mises's account by de Finetti underlines that the theory fails to deal with the role of probability in induction and confirmation:

> If an essential philosophical value is attributed to probability theory, it can only be by assigning to it the task of deepening, explaining or justifying the reasoning by induction. This is not done by von Mises... (De Finetti 1936)

In response to investigations on probability that aim to produce a theory of induction, von Mises claims that probability theory itself is an inductive science and it would be circular to try to justify inductive methodology by means of a science that applies it or to provide any degree of confirmation for any other branch or science:

> According to the basic viewpoint of this book, the theory of probability in its application to reality is itself an inductive science; its results and formulas cannot serve to found the inductive process as such, much less to provide numerical values for the plausibility of any other branch of inductive science, say the general theory of relativity. (1928: vii)

However, it's not that frequency interpretation, in general, does not contribute to the problem of induction. As we have examined elsewhere, [IEP entry on The Problem of Induction (Psillos and Stergiou, 2022)], Reichenbach thought that the frequency interpretation of probability theory provides a new context for understanding the problem of induction.

c. Are Propensities Probabilities?

The propensity interpretations are a family of accounts of physical probability. They aim to provide an account of objective chance in terms of probability theory. Originally, this interpretation has been developed by Karl Popper (1959) but later David Miller, James Fetzer, Donald Gillies and others developed their own accounts (see, Gillies 2000). Paul Humphreys (1985) describes propensities as:

> [I]ndeterministic dispositions possessed by systems in a particular environment, exemplified perhaps by such quite different phenomena as a radioactive atom's propensity to decay and my neighbor's propensity to shout at his wife on hot summer days.

The problems that guided Popper to abandon the frequency interpretation of probability and to develop this new account had to do, on the one hand, with the

interpretation of quantum theory, on the other, with the objective single-case probabilities.

To deal with the problem of single-case probabilities, Popper suggested that probabilities should be associated not with sequences of events but with the generating conditions of these sequences i.e., "the set of conditions whose repeated realisation produces the elements of the sequence" (1959). He claimed that "probability may … be said to be *a property of the generating conditions*" (*ibid*). This was not just an analysis of the meaning of the term 'probability'. Popper claimed to have proposed, "*a new physical hypothesis* (or perhaps a metaphysical hypothesis) analogous to the hypothesis of Newtonian forces. It is the hypothesis that every experimental arrangement (and therefore every state of the system) generates physical propensities which can be tested by frequencies." (*ibid*).

The propensity interpretation is supposed to avoid a number of problems faced by the frequency interpretation; for instance, it avoids the problem of inferring probabilities in the limit. But, especially in Popper's version, it faces the problem of specifying the conditions on the basis of which propensities are calculated – the ascertainability requirement fails. Given that an event can be part of widely different conditions, its propensity will vary according to the conditions. Does it then make sense to talk about the true objective singular probability of an event?

Even if this problem is not taken seriously (after all, the advocate of propensities may well claim that propensities are the sort of thing that varies with the conditions), it has been argued on other grounds that probabilities cannot be identified with propensities. Namely, the so-called *inverse probabilities*, although they are mathematically well-defined, remain uninterpreted since it does not make sense to talk about inverse propensities. Suppose, for instance, that a factory produces red socks and blue socks and uses two machines (Red and Blue) one for each color. Suppose also that some socks are faulty and that each machine has a definite probability to produce a faulty sock, say one out of ten socks produced by the Red machine are faulty. We can meaningfully say that the Red machine has an one tenth propensity to produce faulty socks. But we can also ask the question: given an arbitrary faulty sock, what is the probability that it has been produced by the Red machine? From a mathematical point of view, the question is well-posed and has a definite answer [for a detailed computation of probabilities in a similar example, see section 1a above]. But we cannot make sense of this answer under the propensity interpretation. We cannot meaningfully ask: what is the propensity of an arbitrary faulty sock to have been produced by the Red machine? Propensities, as dispositions, possess the asymmetry of the cause-and-effect relation that cannot be adequately expressed in terms of the symmetric conditional probabilities. Thus, there are well-defined mathematical probabilities that cannot be interpreted as propensities (see Humphreys 1985).

Is this really a problem for the propensity interpretation? We would say 'yes' if a probability interpretation aspires to conform with Kolmogorov's axioms (admissibility requirement) and, also, claims to provide a *complete* interpretation of probability calculus. But this condition is not universally accepted. One may suggest that probability interpretations are *partial* interpretations of the probability calculus or even take the more radical position to abandon the criterion of admissibility, as Humphreys suggested.

3. Probability as the Logic of Induction
a. Keynes and The Logical Concept of Probability

John Maynard Keynes presented his account of probability in the work titled *A Treatise on Probability* (1921). He attempted to provide a logical foundation for probability based on the concept of *partial entailment.* In deductive logic, entailment, considered semantically, expresses the validity of an inference and partial entailment is meant to be its extension to inductive logic. From a semantical point of view, partial entailment expresses a probability relation between the conclusion of an inference and its premises, i.e., that the conclusion is rendered *likely true* (or more likely to be true) given the truth of the premises. Here is how Keynes (1921: 52) understood this extension and its relation to probability:

> Inasmuch as it is always assumed that we can sometimes judge directly that a conclusion follows from a premiss, it is no great extension of this assumption to suppose that we can sometimes recognise that a conclusion partially follows from, or stands in a relation of probability to a premiss.

And:

> We are claiming, in fact, to cognise correctly a logical connection between one set of propositions which we call our evidence and which we suppose ourselves to know, and another set which we call our conclusions, and to which we attach more or less weight according to the grounds supplied by the first.... It is not straining the use of words to speak of this as the relation of probability. (Keynes 1921: 5–6)

Thus, partial entailment rests on an analogy with deductive (full) entailment and both concepts express logical relations, the former of deductive and the latter of inductive logic. Here is an example: the conjunction (p and q) entails deductively p; by analogy, it is said that, though proposition p does not (deductively) entail the conjunction (p and q), it entails it partially, since it entails one of its conjuncts (for instance, p). The difference between the two kinds of entailment stems from the fact that validity of an inference, expressed in deductive entailment, is a yes-or-no question, while the probability relation, expressed in partial entailment, comes in degrees. Keynes (1921: 4) considered *probability* to be the degree of rational belief that a future occurrence of an event under specified circumstances is partially entailed from past evidence for the occurrence of similar events under similar circumstances:

> Let our premises consist of any set of propositions h, and our conclusion consist of any set of propositions a, then, if a knowledge of h justifies a rational belief in a of degree α, we say that there is a *probability-relation* of degree α between a and h.

To say that the probability of a conclusion is high or low given a set of premises is not for Keynes a matter of subjective evaluation of the believer. It shares the objectivity of any other logical relation between propositions. That is why Keynes (1921: 4) talks about the degree of *rational* belief and *not* simply of a degree of belief:

> ... in the sense important to logic, probability is not subjective. It is not, that is to say, subject to human caprice. A proposition is not probable because we think it so. When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed

objectively, and is independent of our opinion. The Theory of Probability is logical, therefore, because it is concerned with the degree of belief which it is *rational* to entertain in given conditions, and not merely with the actual beliefs of particular individuals, which may or may not be rational.

It should be noted that Keynes based his defense of the logical character of the probability relations on what he called "logical intuition", viz., a certain capacity possessed by agents in virtue of which they can simply "see" the logical relation between the evidence and the hypothesis. It is in virtue of this shared intuition that different agents can have the same rational degree of belief in a certain hypothesis in light of certain evidence. This view was immediately challenged by Frank Ramsey, who, referring to Keynes's "logical relations" between statements, noted: "I do not perceive them and if I am to be persuaded that they exist it must be by argument" (1926, 63).

It should be clear that for Keynes probability is not always quantitative. He believed that qualitative probabilities are meaningful as well and that the totality of probabilities, or of degrees of rational belief, may include both numbers and non-numerical elements. In the usual numerical probabilities, all probabilities lie within the unit interval and they are all comparable in terms of the relation 'being greater than or equal to' as defined in real numbers. This relation induces a complete ordering to the unit interval which acquires the structure of a completely ordered set. Since for Keynes probabilities may not be numerical, a different interpretation of the relation "being more probable than or equally probable to" expressing the comparability of probabilities is required. In the class of probabilities, Keynes defines a relation of 'between':

$$A \text{ is between } B \text{ and } C, (A, B, C)$$

where, for any three probabilities A, B, C the relation, if satisfied, is satisfied by a unique ordered triple (A, B, C). He identifies two distinguished probabilities, impossibility, O, and certainty, I, between which *all* other probabilities lie. Finally, he used the relation of betweenness to compare probabilities:

If A is between O and B, the probability B is said to be greater than the probability A.

To illustrate these relations among probabilities, Keynes suggested the following diagram. In this diagram, all probabilities comparable in terms of the 'greater than' relation are connected with a continuous path:
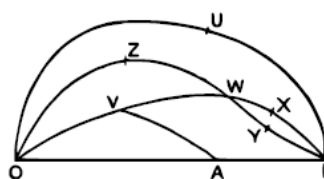
In Keynes's (1921: 39) words:

O represents impossibility, I certainty, and A a numerically measurable probability intermediate between O and I; U, V, W, X, Y, Z are nonnumerical probabilities, of which, however, V is less than the numerical probability A, and is also less than W, X, and Y. X, and Y are both greater than W, and greater than V, but are not comparable with one another, or with A. V and Z are both less than W, X, and Y, but are not comparable with one another; U is not quantitatively comparable with any of the probabilities V, W, X, Y, Z. Probabilities which are numerically comparable will all belong to one series, and the path of this series, which we may call the numerical path or strand, will be represented by OAI.

b. The Principle of Indifference

To have numerical probabilities between alternative cases, Keynes (1921: 41) believed that equiprobability of the alternatives is required:

> In order that numerical measurement may be possible, we must be given a number of equally probable alternatives.

And:

> It has always been agreed that a numerical measure can actually be obtained in those cases only in which a reduction to a set of exclusive and exhaustive equiprobable alternatives is practicable. (1921: 65)

In the terminology of the mathematical theory of probability, Keynes stipulates that a real number $p(E|H)$ denotes the *numerical probability of an event E given the truth of some hypotheses H*, assigned by a function p satisfying Kolmogorov's axioms, only if $p(E|H)$ *can be deduced by or it can be reduced to* some initial numerical probabilities $p(A_k|H)$ assigned to the members of a partition $\{A_k\}_{k=1\ldots n}$ of the event space S that satisfy the equiprobability condition:

$$p(A_k|H) = p\big(A_j|H\big), \qquad k, j = 1, \ldots, n.$$

What is the basis of equiprobability and how can it be justified? Keynes (1921: 45) suggested that the justification of equiprobability follows from the Principle of Indifference which states that:

> if there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability. Thus, equal probabilities must be assigned to each of several arguments, if there is an absence of positive ground for assigning unequal ones.

The term 'Principle of Indifference' was coined by Keynes in the *Treatise on Probability*. According to Ian Hacking (1971), this principle can be traced back to Leibniz's paper "De incerti aestimatione" (1678). In this, Leibniz, anticipating Laplace, claimed that:

> Probability is the degree of possibility. Hope is the probability of having. Fear is the probability of losing.

Leibniz considered the above claim as an axiom—something very similar to the Principle of Indifference:

> Axiom. If players do similar things in such a way that no distinction can be drawn between them, with the sole exception of the outcome, there is the same proportion of hope to fear.

Moreover, he suggested that we understand this axiom as having its source in metaphysics, which seems to be an allusion to the Principle of Sufficient Reason and, in particular, to the claim that God does, or creates, nothing without a sufficient reason. Applying this metaphysical principle to the expectations of rational agents, i.e., 'players', we get the foregoing axiom, as Hacking suggested (1975:126):

> If several players engage in the same contest in such a way that no difference can be ascribed to them (except insofar as they win or lose) then each player has exactly the same ground for 'fear or hope'.

Keynes, however, traces the principle of indifference to Jacques (James) Bernoulli's *Principle of Non-Sufficient Reason* (1921: 41). Bernoulli in his *Ars Conjectandi*, attempted to calculate the "degree of certainty, or probability, that the argument generates" [Notice that by 'argument' he meant a piece of evidence.] and he assumed that "all cases are equally possible, or can happen with equal ease." There are examples, however, in which a case happens more 'easily' than others. Then, according to Bernoulli (1713: 219), we need to make a correction:

> For any case that happens more easily than the others as many more cases must be counted as it more easily happens. For example, in place of a case three times as easy I count three cases each of which may happen as easily as the rest.

Thus, Bernoulli suggested that to save equiprobability we should consider a finer partition of the sample space by subdividing the ill-behaved case into distinct cases.

Keynes was aware that the principle faces a number of difficulties which take the form of a paradox: it predicted contradictory evaluations of probabilities in specific cases. To resolve these paradoxes and avoid ill cases, he attempted to provide restrictions to the application of the principle of indifference.

The first paradox is known as the Book Paradox. Consider a book of unknown cover color. We have no reason to believe that its color is red rather than not red. Hence, by the principle of indifference the probability of being red is $\frac{1}{2}$. In a similar vein, the probability of being green, yellow or blue are all $\frac{1}{2}$ which contradicts the theorem of probability that the sum of probabilities of mutually exclusive events is less than or equal to 1.

The second paradox is the Specific Volume Paradox. Consider the specific volume v of a given liquid and assume that $1 \leq v \leq 3$ in some system of units. Given that there is no reason to assume that $1 \leq v \leq 2$, rather than $2 \leq v \leq 3$, by the principle of indifference it is equally likely for the specific volume to lie in each one of these intervals. Next, consider the specific density $d = \frac{1}{v}$. Given our original assumption, we are justified to infer that $\frac{1}{3} \leq d \leq 1$. Similarly, the principle of indifference

maintains that it is equally likely for the specific density to have a value, $\frac{1}{3} \leq d \leq \frac{2}{3}$, or to have a value, $\frac{2}{3} \leq d \leq 1$. Turning now to considerations about specific volume we find that it is equally likely that $1 \leq v \leq \frac{3}{2}$ or $\frac{3}{2} \leq v \leq 3$. But we have already shown that it is as likely v to lie between 1 and 2 as between 2 and 3.

The third paradox that seems to challenge the principle of indifference is Bertrand's paradox. Bertrand in his *Calcul des Probabilités* (1888) argues that the principle of indifference can be applied in more than one way in cases with infinitely many possibilities giving rise to contradictory outcomes regarding the evaluation of probabilities. In support of his argument he presented, among other examples, his famous paradox: We trace at random a chord in a circle. What is the probability that it would be longer than the side of the inscribed equilateral triangle? Here are some different ways to apply the principle of indifference to solve the problem, each leading to different probability values. The first solution assumes that one end of the requested chord is at a vertex of the triangle and the other lies on the circumference. The circumference is divided in three equal arcs by the vertices of the triangle. From all possible chords traced from the given vertex, only those that lie in the arc which subtends the angle at that vertex are longer than the side of the equilateral triangle. Therefore, the probability is $\frac{1}{3}$. For the second solution, we assume that the chord is parallel to a side of the triangle. From these parallel chords only the ones with distance less than one-half of the circle's radius will have a length greater than the side of the inscribed equilateral triangle. Thus, the requested probability is $\frac{1}{2}$. Finally, we yield a third solution by assuming that the chord is defined by its midpoint. Then a chord is longer than the side of triangle if its midpoint falls within a concentric circle of a radius one-half of the outer circle. The probability is calculated as the ratio of the areas of the two circles and is found $\frac{1}{4}$. Notice that Bertrand's Paradox can undermine the principle of indifference if and only if the problem at hand is a determinate problem with no unique solution. But there is no agreement on that! Many believe that the problem is ambiguous or underspecified and, in this sense indeterminate. They claim that once we select the set of chords from which we draw one at random, the problem has a unique solution by applying the principle of indifference. [For an interesting discussion, see Shackel, 2007].

To address the Book and the Specific Volume Paradoxes, Keynes suggested that we should place a restriction to the application of the Principle of Indifference. We should require that given our state of knowledge, the partition of the sample space, i.e., the number of alternative cases, is finite, and each alternative cannot be split up further into a pair of mutually exclusive sub-alternatives which have non-zero probability to occur (see 1921: 60). Now it is obvious that the class of books with a non-red cover can be further subdivided into the class of books with a blue cover and those with a non-blue cover and so on; thus the adequacy condition for the application of the principle is not satisfied. Similarly, in the case of the ranges of values of the specific volume and the specific density, the principle does not apply since there is no range of values which does not contain within itself two similar ranges. Finally, for Bertrand's paradox, since areas, arcs and segments can be subdivided further into non-overlapping parts without a limit, the principle of indifference is not applicable (see 1921: 62). Yet, for the geometric example, Keynes suggested a solution. Instead of considering as an alternative a point in a continuous line, we may divide that line into a finite number of m segments, no matter how small, and take as an alternative

the segment in which the point under consideration lies. Then we can apply the principle of indifference to the m alternatives which we consider indivisible. However, Keynes solution is not at all clear. Number m can be as great as one desires on the condition that we keep it finite. Hence, who decides what is the number of alternatives to which the principle of indifference is applied? If, on the other hand, we allow m to increase indefinitely then we get the continuous case we sought to avoid. (see Childers 2013: 126)

c. Keynes on the Problem of Induction
For Keynes, probability is the part of logic that deals with rational but inconclusive arguments; and since inductive reasoning is both inconclusive but rational, induction becomes inductive logic. The key question, of course, is the following: on what grounds one is justified to believe that induction is rational?

According to Keynes, though Hume's skeptical claims are usually associated with causation, the real object of his attack is *induction* i.e., the inference from past particulars to future generalizations (see 1921: 312).

Keynes's argument is the following:

(1) A constant conjunction between two events has been observed in the past. This is a fact. Hume does not challenge this at all.
(2) What Hume challenges is whether we are justified to infer from a past constant conjunction between two events that it will also hold in the future.
(3) This kind of inference is called *inductive*.
(4) So, Hume is concerned with the problem of induction.

To see Keynes's reaction to the problem of induction, let's first clarify what is for him an inductive argument: (1921: 251)

> It will be useful to call arguments inductive which depend in anyway on the methods of Analogy and Pure Induction.

Arguments from analogy are based on similarities among the objects of a collection, on their likeness, while Pure Induction is induction by enumeration. As Keynes (*ibid*) put it:

> [w]e argue from … Pure Induction when we trust the number of the experiments.

Keynes criticized Hume for not taking into account the analogical dimension of an inductive argument by considering the observed instances which serve as premises, as absolutely uniform (see 1921: 252). Instead, Keynes suggested that the basis of Pure Induction is the likeness of instances in certain respects (positive analogies) and their dissimilarity in others (negative analogies). Only after having verified such a likeness, we can single out some features and predict the occurrence of other features or infer a generalization of the sort "all A is B". Hence (1921: 253):

> In an inductive argument, therefore, we start with a number of instances similar in some respects AB, dissimilar in others C. We pick out one or more respects A in which the instances are similar, and argue that some of the other

respects B in which they are also similar are likely to be associated with the characteristics A in other unexamined cases.

So, assume that a finite number, n, of instances exhibits a certain group of qualities, $a_1, ..., a_r$ and single out two subgroups:

$$a_1, a_2, a_3 \text{ and } a_{r-1}, a_r$$

An inductive argument, for Keynes, would conclude that in *every instance* of a_1, a_2, a_3, qualities a_{r-1}, a_r are also exhibited. Or that a_{r-1}, a_r "bound up" with qualities a_1, a_2, a_3. (1921: 290) This account of induction presupposes, claims Keynes (*ibid*), that *qualities in objects are exhibited in groups* and "a sub-class of each group [is] an infallible symptom of the coexistence of certain other members of it also."

However, the world may not co-operate to the success of an inductive argument. Keynes identifies three "open possibilities" that would compromise inductive generalization:

(1) Some quality a_{r-1} or a_r, may be independent of all other qualities of the instances, i.e., there are no groups of qualities that contain the said quality and at least some of the others.

(2) There are no groups to which both a_1, a_2, a_3 and a_{r-1}, a_r belong.

(3) a_1, a_2, a_3 belong to groups that include a_{r-1}, a_r and to other groups that do not include them.

In any of the three cases, "All a_1, a_2, a_3' are a_{r-1}, a_r" fails. Hence induction fails.

Keynes (1921: 291) suggested an assumption of probabilistic nature that would save us from such 'pathological' cases and would lead to a successful induction; namely:

> If we find two sets of qualities in coexistence there is a finite probability that they belong to the same group, and a finite probability also that the first set specifies this group uniquely.

If we grant this assumption, then inductive methodology aims to increase the prior probability and make it large, in the light of new evidence. But to this point we will return later.

Keynes discusses the justificatory ground of this assumption and shows that it requires an *a priori* commitment to the claim that qualitative variety in nature is limited. Although the individuals do differ qualitatively, "their characteristics, however numerous, cohere together in groups of invariable connection, which are finite in number" (1921: 285).

This idea is incorporated in the Principle of Limited Variety of a finite system (PLV), which Keynes (1921: 286) stated thus:

> the amount of variety in the universe is limited in such a way that there is no one object so complex that its qualities fall into an infinite number of independent groups (i.e. groups which might exist independently as well as in conjunction); or rather that none of the objects about which we generalise are as complex as this; or at least that, though some objects may be infinitely complex, we sometimes have a finite probability that an object about which we seek to generalise is not infinitely complex.

The gist behind the role of PLV is this. Suppose that although a group of properties, say A, has been invariably associated with a group of properties, B, in the past, there is an unlimited variety of groups of properties, $B_1, \ldots, B_n$, such that it is logically possible that future occurrences of A will be accompanied by any of the B_i's, instead of B. Then, and if we let n (the variety index) tend to infinity, we cannot even start to say how likely it is that B will occur given A, and the past association of As with Bs. PLV excludes the possibility just envisaged.

But as PLV stipulates there are no infinitely complex objects; alternatively, the qualities of an object cannot fall into an infinite number of independent groups. For Keynes, the qualities of an object are determined by a finite number of primitive qualities; the latter (and their possible combinations) can generate all apparent qualities of an object. Since the number of primitive qualities is finite, the number of groups they generate alone or by being combined is finite. Hence, for any two sets of apparent properties, Keynes (1921: 292) concludes, there is, "in the absence of evidence to the contrary, a finite probability that the second set will belong to the group specified by the first set."

In any case, Keynes takes it that a generalization of the form 'All As are Bs' should be read thus 'It is probable that any given A is B' rather than thus 'It is probable that all As are Bs'. So, the issue is the next instance of the observed regularity and not whether it holds generally (1921: 287-288).

The absolute assertion of the finiteness of a system under consideration as expressed by the Principle of Limited Variety is called *Inductive Hypothesis* (IH) (1921: 299), and provides one of the premises of an inductive argument; namely, that the *a priori* probability of our conclusion, $p(C|IH)$, has a finite value. Keynes distinguished (IH) from *Inductive Method* (IM) which amounts to the process of increasing the *a priori* probability of the conclusion, $p(C|IH)$, by taking into account the evidence e:

$$p(C|e\&IH) > p(C|IH).$$

[For the mathematics of Keynes's account of inductive method and the emergence of the need for the inductive hypothesis in order that new evidence strengthen our belief in the truth of the conclusion of an inductive argument, the reader may consult Appendix 6.c]

Significantly, Keynes adds that the Inductive Method may be used to strengthen the Inductive Hypothesis itself. Since IH is a hypothesis and since IM is indifferent to the content/status of the hypothesis it applies to, it can be applied to IH itself. In other words, IM brings some evidence to bear on the truth of IH. What Keynes suggests is this:

$$p(IH|e'\&IH') > p(IH|IH'),$$

where IH' is another general hypothesis, "more primitive and less far-reaching" than IH such that $p(IH|IH')$ has a finite value, and e' other evidence. The argument is non-circular since the justification of the inductive hypothesis is not accomplished by the hypothesis itself but in terms of some other hypothesis more fundamental, by means of inductive method. Of course, the account runs the risk of exchanging circularity for infinite regress unless there exist some *primitive inductive hypothesis.* But what would such a primitive inductive hypothesis be? We are left in the dark:

We need not lay aside the belief that this conviction gets its invincible certainty from some valid principle darkly present to our minds, even though it still eludes the peering eyes of philosophy. (1921: 304)

However, in the end of the day, Keynes simply argues that a non-zero (finite) *a priori* probability is assigned to the inductive hypothesis *IH* (which is equivalent to PLV). What would be the reason to assign an a priori non-zero probability to the inductive hypothesis *IH*? Keynes answer, honest to the bone, shows the limitations of all attempts to satisfy the inductive sceptic: "It is because there has been so much repetition and uniformity in our experience that we place great confidence in it." (1921: 289-290)

It seems we cannot do better than relying on past experience. The Inductive Hypothesis that supports induction, PLV in Keynes's case, is neither a self-evident logical axiom nor an object of direct acquaintance (1921: 304). But nevertheless, he insists that it is true of some factual systems. How do we know this? By past experience!

d. On the Rule of Succession

Before we leave Keynes let us consider his critique of Laplace's Rule of Succession, i.e., the theorem of mathematical probability which claims that if an event has occurred m times in succession, then the probability that it will occur again is $\frac{m+1}{m+2}$. As discussed elsewhere [see our entry in IEP on The Problem of Induction (Psillos and Stergiou, 2022)] Venn had reasons not to "take such a rule as this seriously." (1888: 197), but Keynes's criticism goes well beyond these reasons.

The crux of Keynes' criticism consists in that the derivation of the rule of succession combines two different methods for the determination of the probability of an event which yield different probability values. Thus, their combination is inconsistent and it includes a latent contradiction.

Consider several possible events $E_1, E_2, \ldots, E_n$ that are alternatives, i.e., they are mutually exclusive and exhaustive of the sample space, and choose any one of them, E_i.

The first method stipulates that "when we do not know anything about an alternative, we must consider all the possible values of the probability of the alternative; these possible values can form in their turn a set of alternatives, and so on. But this method by itself can lead to no final conclusion." (1921: 426) Let the probability of the alternative be $p(E_i)$. The method stipulates that we should consider all probability values of E_i assigned by any admissible probability functions p. These probability values for E_i form another set of alternatives, say, $p_1(E_i), \ldots, p_n(E_i), \ldots$ And the same process may be repeated, again and again, involving us in an infinite regress. Thus, the first method is inconclusive.

The second method applies the principle of indifference stipulating that "when we know nothing about a set of alternatives, we suppose the probabilities of each of them to be *equal*." (*ibid*) Thus, the second method concludes that, $p(E_1), = \cdots = p(E_n)$. Consider the event that E_1: "the sun will rise tomorrow" and its alternative that the E_2: "the sun will not rise tomorrow". If we apply the first method only, we reach no conclusion about probability and we are involved in infinite regress. Secondly, if we apply the second method only, we obtain $p(E_1) = p(E_2) = \frac{1}{2}$. Finally, in deriving the rule of succession both methods are applied subsequently. Namely, the probability of

E_1 is unknown, and any probability value is possible according to the first method. Thus, we form a set of alternatives for the probability of E_1 which, at a second stage are reduced to the equal probability case by applying the second method. This reasoning is presupposed by the rule of succession.

The latent contradiction included in the rule of succession is that for its derivation it is assumed that the *a priori* probability of the event can be any number in the interval [0,1], with all numbers being equally probable, while by application of the rule the *a priori* probability, calculated in the absence of any observations (N=0) is $\frac{1}{2}$. In Keynes's (1921: 430) own words:

The principle's conclusion is inconsistent with its premises. We begin with the assumption that the a priori probability of an event, about which we have no information and no experience, is unknown, and that all values between 0 and 1 are equally probable. We end with the conclusion that the a priori probability of such an event is $\frac{1}{2}$ … this contradiction was latent, as soon as the Principle of Indifference was superimposed on the principle of unknown probabilities.

4. Carnap's Inductive Logic
a. Two Concepts of Probability
Carnap presented his views of probability and induction mainly in the two books entitled the *Logical Foundations of Probability* (1950) and *The Continuum of Inductive Methods* (1952) and in his papers "A basic system of inductive logic, I, II" (1971 and 1980, respectively) and "Replies and Systematic Expositions" (1963). For Carnap, the theory and principles of inductive reasoning, inductive logic, is the same as probability logic (1950, v) and the primary task to be set toward an account of inductive logic is the *explication of probability*.

Explication, according to Carnap (1950: 3)*,* is the transformation of an inexact, possibly prescientific concept, the *explicandum*, into a new exact concept, the *explicatum*, that obeys explicitly stated rules for its use. By means of this transformation a concept of ordinary discourse or a metaphysical concept may be incorporated into a well-structured body of logico-mathematical or empirical concepts. Explication has a long history as a philosophical method that, in a wide sense, may be traced back even to Plato's investigations on definitions. Strictly speaking, however, Carnap borrowed the term "Explikation" from Kant and Husserl while Frege may be considered his precursor in this method of philosophical analysis and Goodman, Quine and Strawson among his prominent intellectual inheritors. [For a general presentation of the notion explication, consult IEP's entry on Explication, (Cordes and Siegwart 2019).]

Two concepts are distinguished as *explicanda* of probability according to Carnap: the logical or inductive probability, called 'probability$_1$' and the statistical probability, called 'probability$_2$'. Both concepts are important for science and lack of recognition of this fact, Carnap claimed, has fueled many futile controversies among philosophers. The meaning of probability$_2$ is that of relative frequency of a kind of event in a long sequence of events, and in science it is applied to the description and statistical analysis of mass phenomena. All sentences about statistical probability are factual, empirical.

The logical concept of probability, probability$_1$, is the basis for all inductive reasoning. For Carnap (1950: 2), the problem of induction is the problem of the logical relation between a hypothesis and some confirming evidence for it and

"inductive logic is the theory based upon what might be called the degree of inducibility, that is, the degree of confirmation." Hence, by taking probability$_1$ to mean "the degree of confirmation of a hypothesis h with respect to an evidence statement e, e.g., an observational report" (1950: 19) Carnap made it the basis of inductive logic. As for any logical sentence, the truth or falsity of sentences about probability$_1$ is independent of extralinguistic facts.

In addition, logical probability is an objective concept, i.e., "if a certain probability$_1$ value holds for a certain hypothesis with respect to a certain evidence, then this value is entirely independent of what any person may happen to think about these sentences, just as the relation of logical consequence is independent in this respect."(1950: 43) The objectivity of probability$_1$, Carnap recognized it in the views of Keynes and Jeffreys who interpreted probability in terms of rational degrees of beliefs as distinguished from subjective, actual degrees of belief a person might bear on the truth of a sentence given some evidence. Later, he (1963: 967) came to accept the interpretation of probability$_1$ as "the degree to which [one]… is rationally entitled to believe in h on the basis of e."

b. C-functions

Carnap suggested three different concepts of confirmation. The *classificatory concept of confirmation*, which expresses a logical relation between a piece of evidence e and a hypothesis h and, if satisfied, it qualifies the former as a confirming instance of the latter. To signify the *explicatum* of this concept, Carnap used the symbol '$\mathfrak{C}$'and $\mathfrak{C}(h, e)$ corresponds to "h is confirmed (or, supported) by e". The second concept of confirmation he employed is the *comparative concept* which compares the strength by which a piece of evidence e_1 confirms a hypothesis h_1 with the corresponding strength by which e_2 confirms h_2. Thus, comparative confirmation requires the underlying classificatory confirmation and it is, in general, a tetradic relation. Its explicatum is symbolized by '$\mathfrak{MC}$', where $\mathfrak{MC}(h_1, e_1, h_2, e_2)$ corresponds to the statement "h_1 is confirmed by e_1 at least as strongly (i.e., either more, or equally, strongly) as h_2 by e_2". Finally, there is a *quantitative (or, metrical) concept* of confirmation, the *degree of confirmation,* which assigns a numerical value to the degree to which a hypothesis h is supported by given observational evidence e. The explicatum of this concept is symbolized by '$\mathfrak{c}$', where 'the degree of '$\mathfrak{c}(h, e) = r$' is the statement, "the degree of confirmation of h with respect to e is r", where h and e are sentences and r a real number in the unit interval.

In this context, Carnap points out that Keynes's objective conception of probability is similar to the comparative concept of confirmation and only in some special cases, when the principle of indifference is applicable, it can be interpreted quantitatively similar to his concept of degree of confirmation (1950: 45 & 205). Moreover, notice that all three conceptions of confirmation Carnap (1950: 19) suggested are *semantical:*

> The concepts of confirmation to be dealt with in this book are semantical, i.e., based upon meaning, and logical, i.e., independent of facts.

The inductive relation the three concepts of confirmation attempt to explicate is not determined by the form of the sentences, as Hempel required in his syntactic account of confirmation (1945), nor depend on the users of a language, as Goodman suggested in his pragmatic solution of the new riddle of induction (1955) (See also our other entry in IEP on The Problem of Induction (Psillos and Stergiou, 2022)). Rather:

[O]nce h and e are given, the question mentioned requires only that we be able to understand them, that is, to grasp their meanings, and to establish certain relations which are based upon their meanings (1950: 20).

Carnap begins with the construction of the language(s) in which inductive logic is to be applied. He defines several language systems each one characterized by the number of names (constants) it contains (1950: 58). Each name refers to individuals in the corresponding universe of discourse, be they things, events, or the like. Thus, he considered an infinite language system $\mathfrak{L}_\infty$, having an infinite number of names and a sequence $\mathfrak{L}_1, \mathfrak{L}_2, \dots, \mathfrak{L}_N, \dots$ of language systems each one characterized by the index N that runs through all positive integers indicating the number of names the system includes. Hence, $\mathfrak{L}_1$ contains only 'a_1'; $\mathfrak{L}_2$ contains 'a_1' and 'a_2'; etc. Notice that any sentence of $\mathfrak{L}_\infty$ is contained in an infinite number of finite language systems of the hierarchy since if 'a_N' is the name with highest subscript that appears in that sentence, then this sentence will be represented in any language system $\mathfrak{L}_n$ with $n \geq N$. Apart from names, $\mathfrak{L}_\infty$ contains a finite number of primitive (atomic) predicates of any degree (unary, binary etc.) designating properties and relations among individuals in the universe of discourse. Carnap considered only three connectives as primitive for his language systems: the negation '$\sim$', the conjunction '&' and the inclusive disjunction '$\vee$' – and he defined implication and biconditional in terms of these three. Each language system contains an infinite number of variables, $x, y, z, x_1, x_2 \dots$, and two quantifiers, the existential '$(\exists x)$' and the universal one, '(x)'. The sentence '$(x)Px$' is taken to be logically equivalent to '$Pa_1 \& Pa_2 \dots \& Pa_N$' in a language $\mathfrak{L}_N$, according to the semantics adopted. The same is not true for the case of $\mathfrak{L}_\infty$ since in this case the conjunction of an infinite number of sentences is not a well-formed formula of the language. Apart from the atomic predicates, molecular predicates may be defined. They are formed by atomic or more basic molecular predicates with the help of connectives. For example, if P_1, P_2, P_3 are atomic predicates, then '$\sim P_1$' or '$P_1 \& P_2$' or '$P_1 \vee P_3$' are molecular predicates understood as follows: for any variable x, $(\sim P_1)x$ stands for '$\sim(P_1 x)$'; $(P_1 \& P_2)x$ for '$P_1(x) \& P_2(x)$'; and $(P_1 \vee P_3)x$ for '$P_1(x) \vee P_3(x)$'. Finally, language systems contain an equality symbol '=' designating identity of individuals in the universe of discourse and a tautological sentence 't'. As any language, these language systems are equipped with some rules for the formation of well-formed formulas (sentences) and some rules of truth, i.e., a semantics.

A *state description* $\mathfrak{B}$ is an explication of the vague concept of a state of affairs relativized to a given language system $\mathfrak{L}$ (1950: 70ff). It purports to describe possible states of the universe of discourse of $\mathfrak{L}$. A state description describes for every individual designated by some name 'a' and for every property designated by an atomic predicate 'P' of $\mathfrak{L}$ whether or not this individual has that property, and similarly for relations. Thus, a state description will contain exactly one sentence from the pair 'Pa', '$\sim Pa$': either 'Pa' or '$\sim Pa$' but not both, and no other element (similarly for relations). In the case of a finite language system $\mathfrak{L}_N$, a state description has the form of a conjunction of sentences of the aforementioned sort while in the case of an infinite language system $\mathfrak{L}_\infty$, a state description is a class of sentences that contains at most one sentence of the aforementioned sort. In both cases nothing more is included in a state description. The class of all state descriptions in a given system $\mathfrak{L}$ is designated by '$V_\mathfrak{B}$' while the null class by '$\Lambda_\mathfrak{B}$'.

For example, consider a language system $\mathfrak{L}_3$ with names, 'a', 'b' and 'c' and a single atomic unary predicate symbol 'P'. The complete set of state descriptions is the following:

$\mathfrak{B}1$ 'Pa&Pb&Pc'	$\mathfrak{B}5$ '$\sim Pa$&$\sim Pb$&Pc'
$\mathfrak{B}2$ '$\sim Pa$&Pb&Pc'	$\mathfrak{B}6$ '$\sim Pa$&Pb&$\sim Pc$'
$\mathfrak{B}3$ 'Pa&$\sim Pb$&Pc'	$\mathfrak{B}7$ 'Pa&Pb&$\sim Pc$'
$\mathfrak{B}4$ 'Pa&Pb&$\sim Pc$'	$\mathfrak{B}8$ '$\sim Pa$&$\sim Pb$&$\sim Pc$'

The adequacy of a language system $\mathfrak{L}$ for inductive logic requires compliance with two important conditions: the requirement of logical independence and the requirement of completeness. The first condition aims at restricting the language system to bar contradictory state descriptions. The requirement of logical independence stipulates (i) that atomic sentences (i.e. sentences that consist of an n-place predicate and n names) are logically independent, i.e. a class containing atomic sentences (e.g. sentences of the form Pa for a predicate 'P' and a name 'a') and the negations of other atomic sentences does not entail logically entail another atomic sentence or its negation; (ii) names in $\mathfrak{L}$ designate different and separate individuals; (iii) atomic predicates are interpreted to designate logically independent attributes.

The requirement of completeness of language stipulates that the set of the atomic predicates of $\mathfrak{L}$ be sufficient for expressing every qualitative attribute of the individuals in the universe of discourse of $\mathfrak{L}$. This requirement seemed absolutely necessary for the Carnapian system, since the language systems affect the $\mathfrak{c}$-values in the theory of inductive logic. For the time being, all we need to stress is that this requirement implies that a language system $\mathfrak{L}$ *mirrors* its universe of discourse. Whatever there is in it can be exhaustively expressed within $\mathfrak{L}$. Here is Carnap's example (1950: 75). Take a language system $\mathfrak{L}$ with only two predicates, 'P_1' and 'P_2' interpreted as Bright and Hot. Then, every individual in the universe of discourse of $\mathfrak{L}$ should differ only with respect to these two attributes. If a new predicate 'P_3', interpreted as Hard, were added, the $\mathfrak{c}$-values of hypotheses concerning individuals in $\mathfrak{L}$ would change. Even if this simple scheme holds (or might hold) in a simple language, can it be adequate for the language of natural sciences? A similar requirement had been proposed by Keynes, in the form of the Principle of Limited Variety (see section 3c above).

Later on, Carnap abandoned this requirement and replaced it with the following: The value of the confirmation function $\mathfrak{c}(h, e)$ remains unchanged if further families of predicates are added to the language (see 1963: 975). According to this requirement, the value of $\mathfrak{c}(h, e)$ depends *only* on the predicates occurring in h and e. Hence, the addition of new predicates to the language does not affect the value of $\mathfrak{c}(h, e)$. This new idea amounts to what Lakatos (1968: 325) called the *minimal language requirement*, according to which the degree of confirmation of a proposition depends only on the minimal language in which the proposition can be expressed.

Another important concept defined by Carnap is that of the *range* of a sentence or of a collection of sentences (1950: 78). The range of a sentence i, $\mathfrak{R}(i)$, is the class of those state descriptions in which that sentence holds. A (molecular) sentence of the form 'Pa or $\sim Pa$' for a atomic predicate 'P' and some name 'a' holds in a state description $\mathfrak{B}$ if it is either a conjunct in $\mathfrak{B}$'s defining conjunction or it belongs to the class of sentences that define $\mathfrak{B}$. Analogously, if a sentence is a conjunction of sentence, then all components of the conjunction should hold for a state description

while if it is a disjunction, at least one disjunct should hold in a state description – so that the state description partake of the sentence's range. Notice that a tautology holds in all state descriptions. For instance, in the previous example, the range of $Pa\&Pb$ is $\Re(Pa\&Pb) = \{\mathfrak{B}1, \mathfrak{B}4, \}$ while the range of $Pa \vee Pb$ is $\Re(Pa \vee Pb) = \{\mathfrak{B}1, \mathfrak{B}2, \mathfrak{B}3, \mathfrak{B}4, \mathfrak{B}6, \mathfrak{B}7\}$. Finally, the range of a class of sentences is the class of state descriptions in which every sentence of class holds.

As a final step before defining the $\mathfrak{c}$-function, we present Carnap's account of logical concepts in a system $\mathfrak{L}$ in terms of state descriptions and the concept of range: a sentence i is *L-true* in $\mathfrak{L}$ if and only if $\Re(i)$ is $V_{\mathfrak{B}}$ while it is *L-false* in $\mathfrak{L}$ if and only if $\Re(i)$ is $\Lambda_{\mathfrak{B}}$; a sentence i *L-implies* j in $\mathfrak{L}$ if and only if $\Re(i) \subset \Re(j)$; i is *L-equivalent* to j in $\mathfrak{L}$ if and only if $\Re(i) = \Re(j)$; $j_1, j_2, \dots, j_n$ $(n \geq 2)$ are *L-disjunct* with one another in $\mathfrak{L}$ if and only if $\Re(j_1) \cup \Re(j_2) \cup \dots \cup \Re(j_n)$ is $V_{\mathfrak{B}}$; i is *L-exclusive of j* in $\mathfrak{L}$ if and only if $\Re(i) \cap \Re(j)$ is $\Lambda_{\mathfrak{B}}$; a class of sentences is *L-exclusive in pairs* if and only if every pair of the class is L-exclusive of every other sentence of that class. L-truth is the explicatum for logical truth or analytical truth while L-false for contradiction. L-implication is the explicatum for logical entailment while L-equivalence explicates mutual deducibility and it is the same as mutual L-implication. L-disjunctness applied to a set of sentences explicates the idea that at least one of those sentences is true and L-exclusion explicates logical incompatibility or logical impossibility of joint truth.

For the sake of simplicity, in this presentation we focus on finite language systems. Thus, $\mathfrak{m}$ is a *regular measure function* (briefly, a regular $\mathfrak{m}$-function) for $\mathfrak{B}$ in $\mathfrak{L}_N$ if and only if it fulfills the following two conditions: (a) for every $\mathfrak{B}_i$ in $\mathfrak{L}_N$, $\mathfrak{m}(\mathfrak{B}_i) \in \mathbb{R}$; (b) the sum of the values of $\mathfrak{m}$ for all $\mathfrak{B}$ in $\mathfrak{L}_N$ is 1, $\sum_{\mathfrak{B}_i} \mathfrak{m}(\mathfrak{B}_i) = 1$. The regular $\mathfrak{m}$-function for $\mathfrak{B}$ can be extended to a *regular $\mathfrak{m}$-function for the sentences* in $\mathfrak{L}_N$ by requiring the following: (a) for any L-false sentence j in $\mathfrak{L}_N$, $\mathfrak{m}(j) = 0$; (b) for any non-L-false sentence j, $\mathfrak{m}(j) = \sum_{\mathfrak{B} \in \Re(j)} \mathfrak{m}(\mathfrak{B})$ (Carnap 1950: 295).

In the example of the language system $\mathfrak{L}_3$ considered previously, a regular $\mathfrak{m}$-function for state descriptions is defined as follows:

$$\mathfrak{m}(\mathfrak{B}_i) = \frac{1}{12}, \text{ for } i = 1,3,4,7 \quad \mathfrak{m}(\mathfrak{B}_i) = \frac{1}{6}, \text{ for } i = 2,5,6,8.$$

It is extended to a regular $\mathfrak{m}$-function for sentences that assigns numerical values to sentences, e.g.,

$$\mathfrak{m}(Pa\&{\sim}Pa) = 0 \; ; \; \mathfrak{m}(Pa \vee {\sim}Pa) = 1$$
$$\mathfrak{m}(Pa\&Pb) = \sum_{i=1,4} \mathfrak{m}(\mathfrak{B}_i) = \frac{1}{6} \; ; \; \mathfrak{m}(Pa \vee Pb) = \sum_{i=1,2,3,4,6,7} \mathfrak{m}(\mathfrak{B}_i) = \frac{2}{3}.$$

A regular confirmation function is defined as a two-argument function for sentences on the basis of a regular $\mathfrak{m}$-function for sentences in $\mathfrak{L}_N$. Namely, let $\mathfrak{m}$ be a regular $\mathfrak{m}$-function for sentences in $\mathfrak{L}_N$, then $\mathfrak{c}$ *is a regular confirmation function* (briefly, a regular $\mathfrak{c}$-function) *for sentences* in $\mathfrak{L}_N$ if and only if for any sentences e, h in $\mathfrak{L}_N$,

$$\mathfrak{c}(h, e) = \frac{\mathfrak{m}(e\&h)}{\mathfrak{m}(e)},$$

where $\mathfrak{m}(e) \neq 0$ and $\mathfrak{c}(h, e)$ has no value, where $\mathfrak{m}(e) = 0$ (Carnap 1950: 295).

In the aforementioned example, if e stands for the L-false sentence '$Pa\&\sim Pa$', $\mathfrak{c}(h, e)$ is not defined for any hypothesis h. L-false sentences cannot be evidence for or against any hypothesis. However, if an L-false sentence, e.g., '$Pa\&\sim Pa$', is taken as hypothesis h, then $\mathfrak{c}(h, e) = 0$, for any admissible piece of evidence e. Consider an L-true sentence, such as '$Pa \vee \sim Pa$', as hypothesis h. Then $\mathfrak{c}(h, e) = 1$ no matter what the admissible evidence might be; no evidence can increase or decrease the degree of confirmation of a logical truth (obviously, e is not L-false). In other cases, e.g., for the hypothesis h, 'Pa' and the evidence e, 'Pb', $\mathfrak{c}(Pa, Pb) = \frac{\mathfrak{m}(Pa\&Pb)}{\mathfrak{m}(Pb)} = \frac{1/6}{1/2} = \frac{1}{3}$.

A regular $\mathfrak{c}$-function is a conditional probability function in the common parlance of mathematical theory of probability since it satisfies Kolmogorov's axioms. This was a *desideratum* for Carnap who stipulated that an adequate concept of degree of confirmation should fulfill the following conditions (1950: 285):

(a) *L-equivalent evidences.* If e and e' are L-equivalent, then $\mathfrak{c}(h, e) = \mathfrak{c}(h, e')$.
(b) *L-equivalent hypotheses.* If h and h' are L-equivalent, then $\mathfrak{c}(h, e) = \mathfrak{c}(h', e)$.
(c) *General Multiplication Principle.* $\mathfrak{c}(h\&j, e) = \mathfrak{c}(h, e) \cdot \mathfrak{c}(j, e\&h)$.
(d) *Special Addition Principle.* If $e\&h\&j$ is L-false, then $\mathfrak{c}(h \vee j, e) = \mathfrak{c}(h, e) + \mathfrak{c}(j, e)$
(e) *Maximum Value.* For any not L-false e $\mathfrak{c}(t, e) = 1$,

where h, h', e, e', j are any sentences in $\mathfrak{L}_N$ and t is a logical truth. Conditions, (a) and (b) demand that the *explicatum* of the degree of confirmation should respect logical equivalence. The General Multiplication Principle is derived mathematically directly from the definition of conditional probability. The Special Addition Principle is recognized as the additivity axiom in Kolmogorov's formulation which gives rise to the finite additivity condition and the Maximum Value condition corresponds to the fact probability of the sample space is 1.

To recover *unconditional* probability functions for sentences in $\mathfrak{L}_N$, Carnap suggested to consider the probability of any sentence conditionally to a tautology. Namely, if $\mathfrak{c}$ is a regular confirmation function for $\mathfrak{L}_N$, then for every sentence j in $\mathfrak{L}_N$, the *null confirmation* $\mathfrak{c}_0$ is $\mathfrak{c}_0(j) = \mathfrak{c}(j, t)$. Moreover, he showed that $\mathfrak{c}_0(j) = \mathfrak{m}(j)$. The null confirmation represents the prior probability of a sentence in the absence of any evidence (1950: 307-8).

In the example of the language system $\mathfrak{L}_3$ considered previously we suggested a regular $\mathfrak{m}$-function that assigns different real numbers to different state descriptions, i.e., to different states in the universe of discourse. However, is there any reason to believe that these numbers should be unequal? Is there any reason to believe that one state description weighs more than any other? Rather, by application of the principle of indifference, it seems that we should demand equal distribution of weight to all state descriptions,

$$\mathfrak{m}^+(\mathfrak{B}) = \frac{1}{\zeta}$$

where ζ is the number of the state descriptions in $\mathfrak{L}_N$ (Carnap, 1950: 564). Moreover, it is easy to show that for any given piece of evidence e and for every pair of state description $\mathfrak{B}_i, \mathfrak{B}_j$ compatible with e, it holds:

$$\mathfrak{c}^+(\mathfrak{B}_i, e) = \mathfrak{c}^+(\mathfrak{B}_j, e).$$

Of course, the principle of indifference entails equiprobability only for state descriptions and not for all sentences, in a way that Keynes would appreciate, since he was the first to suggest restricted application of the principle of indifference to possibilities that are mutually exclusive and exhaustive of the sample space, to avoid the Book paradox. Salmon (1966: 72) notes that Carnap's "…explication of probability in these terms has been thought to preserve the 'valid core' of the traditional principle of indifference".

Nevertheless, Carnap has shown that to suggest a regular $\mathfrak{m}$-function for $\mathfrak{B}$ in $\mathfrak{L}$ that assigns equal weight to all state descriptions, although intuitively plausible, has deeply undesirable consequences: it inhibits learning from experience. To see why consider a language $\mathfrak{L}_{N+1}$, with a single unary atomic predicate P. We want to calculate the degree of confirmation of the hypothesis that the $(N+1)^{\text{th}}$ individual will have the property P, i.e., h: 'Pa_{N+1}', given the evidence that *all* individuals examined so far had the property P, i.e., e: '$Pa_N \& \dots \& Pa_1$'. The number of state descriptions is 2^{N+1}, hence, the $\mathfrak{m}^+$ regular $\mathfrak{m}$-function assigns equal weight to all state descriptions, $\mathfrak{m}^+(\mathfrak{B}) = \frac{1}{2^{N+1}}$. First, notice that $h\&e$ and $\sim h\&e$ are state descriptions; hence, $\mathfrak{m}^+(h\&e) = \mathfrak{m}^+(\sim h\&e) = \frac{1}{2^{N+1}}$. Second, sentences e and $(h\&e) \vee (\sim h\&e)$ are L-equivalent and $\mathfrak{m}^+(e) = \mathfrak{c}_0^+(e) = \mathfrak{c}^+(e, t)$. By the L-equivalent-hypotheses condition, $\mathfrak{m}^+(e) = \mathfrak{c}^+\big((h\&e) \vee (\sim h\&e), t\big)$; and by the Special Addition Principle, $\mathfrak{m}^+(e) = \mathfrak{c}^+(h\&e, t) + \mathfrak{c}^+(\sim h\&e, t) = \mathfrak{c}_0^+(h\&e) + \mathfrak{c}_0^+(\sim h\&e) = \mathfrak{m}^+(h\&e) + \mathfrak{m}^+(\sim h\&e) = \frac{1}{2^{N+1}} + \frac{1}{2^{N+1}} = \frac{2}{2^{N+1}}$. Hence,

$$\mathfrak{c}^+(h, e) = \frac{\mathfrak{m}^+(h\&e)}{\mathfrak{m}^+(e)} = \frac{\frac{1}{2^{N+1}}}{\frac{2}{2^{N+1}}} = \frac{1}{2}$$

Moreover, by a simple calculation

$$\mathfrak{c}_0^+(h) = \mathfrak{m}^+(h) = \sum_{\mathfrak{B} \in \mathfrak{R}(h)} \mathfrak{m}^+(\mathfrak{B}) = 2^N \frac{1}{2^{N+1}} = \frac{1}{2}.$$

i.e.,

$$\mathfrak{c}^+(h, e) = \mathfrak{c}_0^+(h).$$

The last equality yields the desired conclusion: the degree of confirmation of a hypothesis is independent of the evidence collected in a given population. No matter how many positive instances of a given property one observes in a population, their guess regarding the appearance of the property in the next individual is not better justified than if no observations were made; thus learning does not come from experience (1950: 564-5).

To avoid this difficulty, Carnap suggested to apply the principle of indifference in a different way. Instead of distinguishing states of affairs in terms of properties and relations instantiated by certain individuals, Carnap grouped all states of affairs

instantiating the same properties and relations, *independently* of the individuals that instantiated them, and distinguished only among these classes. Hence, we should not focus anymore on state descriptions describing possible states of the universe of discourse for a language system but on classes of such state descriptions in which any two state descriptions are *isomorphic* to one another. Two sentences i, j in $\mathfrak{L}_N$ are isomorphic if j is formed from i by replacing each individual constant occurring in i by its correlate with respect to a one-to-one relation among all individual constants in $\mathfrak{L}_N$. These classes are called *structure descriptions,* $\mathfrak{Str}$. They describe the common structure attributed to the realm of individuals by a class of state descriptions. For instance, a structure description may express the fact that there are exactly two individuals in the universe of discourse possessing a given property P or that none of the individuals bears the relation R to itself, or that relation R is satisfied by pairs of individuals non-symmetrically - i.e., if for all individual constants a, b Rab and $\sim Rba$ are both satisfied - etc. Now the principle of indifference applies in two stages: firstly, following the principle we assign equal weight to all structure descriptions and, secondly, within each structure description we assign equal weight to all isomorphic state descriptions. Thus, for a state description $\mathfrak{B}_i$ in a language system $\mathfrak{L}_N$, if τ is the number of structure descriptions $\mathfrak{Str}$ and ζ_i the number of all state descriptions that are isomorphic to $\mathfrak{B}_i$, we define (1950: 564) the regular $\mathfrak{m}$-function for $\mathfrak{B}$:

$$\mathfrak{m}^*(\mathfrak{B}_i) = \frac{1}{\tau \cdot \zeta_i}.$$

To illustrate the relation between state descriptions and structure descriptions and the difference between the values of $\mathfrak{m}^+, \mathfrak{m}^*$ regular $\mathfrak{m}$-functions, consult the following table which represents the example of $\mathfrak{L}_3$ with a single predicate P:

STATE DESCRIPTIONS	WEIGHT	STRUCTURE DESCRIPTIONS	WEIGHT	$\mathfrak{m}^+$	$\mathfrak{m}^*$
$Pa\&Pb\&Pc$	1/8	All Ps, no$\sim P$s	1/4	1/8	1/4
$\sim Pa\&Pb\&Pc$	1/8			1/8	1/12
$Pa\&\sim Pb\&Pc$	1/8	2 Ps, 1 $\sim P$	1/4	1/8	1/12
$Pa\&Pb\&\sim Pc$	1/8			1/8	1/12
$\sim Pa\&\sim Pb\&Pc$	1/8			1/8	1/12
$\sim Pa\&Pb\&\sim Pc$	1/8	1 P, 2 $\sim P$	1/4	1/8	1/12
$Pa\&\sim Pb\&\sim Pc$	1/8			1/8	1/12
$\sim Pa\&\sim Pb\&\sim Pc$	1/8	No Ps, all $\sim P$s	1/4	1/8	1/4

Let's now revisit the problem of determining the degree of confirmation of the hypothesis that the $(N + 1)^{\text{th}}$ individual will have the property P, i.e., h: 'Pa_{N+1}', given the evidence that *all* individuals examined so far had the property P, i.e., e: '$Pa_N\& ...\&Pa_1$' in a language $\mathfrak{L}_{N+1}$ with a single unary predicate P. Since our language contains $N + 1$ individual constants, a structure description is determined by the number of instances of the property P we find in the universe of discourse disregarding the identity of the individuals that instantiate the property. Thus, all state descriptions that are isomorphic to '$Pa_{N+1}\&Pa_{N-1}\& ...\&Pa_1$' correspond to the same structure description characterized by $N + 1$ property instances in the universe of discourse, while all state descriptions that are isomorphic to '$\sim Pa_{N+1}\&\sim Pa_{N-1}\& ...\&\sim Pa_1$' correspond to the same structure description characterized by 0 property instances in the universe of discourse. Thus, we have different structure description corresponding to $0, 1, ..., N + 1$ occurrences of P and

the total number of structure descriptions is $\tau = N + 2$. To calculate the number ζ_k of state descriptions that are isomorphic to $\mathfrak{B}_k$, let us take k to denote the number of occurrences of P in $\mathfrak{B}_k$, i.e., $k = 0,1, \dots, N + 1$. Then ζ_k is the number of the different ways that N individuals can form k-tuples, i.e., $\binom{N + 1}{k} = \frac{(N+1)!}{k!(N+1-k)!}$. Thus, we find that

$$\mathfrak{m}^*(\mathfrak{B}_k) = \frac{k!\,(N + 1 - k)!}{(N + 2)!}$$

for $k = 0,1, \dots, N + 1$.

The degree of confirmation of the hypothesis h given evidence e is

$$\mathfrak{c}^*(h, e) = \frac{\mathfrak{m}^*(h\&e)}{\mathfrak{m}^*(e)}.$$

Notice that $h\&e$ is isomorphic to any state description $\mathfrak{B}_{N+1}$ and

$$\mathfrak{m}^*(h\&e) = \mathfrak{m}^*(\mathfrak{B}_N) = \frac{(N + 1)!}{(N + 2)!} = \frac{1}{N + 2}$$

while $\sim h\&e$ is isomorphic to any state description $\mathfrak{B}_N$ and

$$\mathfrak{m}^*(\sim h\&e) = \mathfrak{m}^*(\mathfrak{B}_N) = \frac{N!}{(N + 2)!}.$$

As before, sentence e is L-equivalent to $(h\&e) \vee (\sim h\&e)$ and

$$\mathfrak{m}^*(e) = \mathfrak{m}^*(h\&e) + \mathfrak{m}^*(\sim h\&e) = \frac{N!}{(N + 1)!} = \frac{1}{N + 1}$$

Thus,

$$\mathfrak{c}^*(h, e) = \frac{\mathfrak{m}^*(h\&e)}{\mathfrak{m}^*(e)} = \frac{N + 1}{N + 2}.$$

Using the same reasoning, we may calculate, more generally, the degree of confirmation of the hypothesis that the $(r + 1)$-th individual a_{r+1} will exhibit property P, i.e., h: 'Pa_{r+1}' given the evidence that r individuals of the universe of discourse have exhibited so far the same property P, i.e. e : '$Pa_r\& \dots \&Pa_1$',

$$\mathfrak{c}^*(h, e) = \frac{\mathfrak{m}^*(h\&e)}{\mathfrak{m}^*(e)} = \frac{r + 1}{N + 2}.$$

These results amount to the celebrated Laplace's Rule of Succession, which in Carnap's theory of inductive logic has become a theorem.

c. The Continuum of Inductive Methods
In the examples so far, we have examined three different regular $\mathfrak{c}$-functions: one determined by arbitrarily assigning weight to state descriptions in $\mathfrak{L}_3$; the other two,

$\mathfrak{c}^+, \mathfrak{c}^*$, determined by assigning equal weight to state and structure descriptions, respectively, on the basis of the principle of indifference. There are many alternative ways to assign such a weight to the different possibilities and each one of them results in a different regular $\mathfrak{c}$-function yielding a different degree of confirmation $\mathfrak{c}(h, e)$ for a given hypothesis h and evidence e in a language system $\mathfrak{L}$. Thus, there are many different inductive methods, actually, a continuum of such possible methods (Carnap, 1952). For a given language system each inductive method is characterized by the value of a non-negative real parameter λ. For a given λ the degree of confirmation $\mathfrak{c}(h, e)$ is fixed for any hypothesis h and with respect to any evidence e and any two inductive methods have the same λ only if they agree on the value of $\mathfrak{c}(h, e)$.

To understand how the degree of confirmation is defined in terms of the λ-parameter, we need first to explain the concept of *logical width* of a property (1950: 126-127). Consider any language system $\mathfrak{L}_N$ having π unary atomic predicates. We may form molecular predicates by taking the conjunction of π predicates which are either the atomic predicates or of their negations. In this way we form $\kappa = 2^\pi$ molecular predicates (Q-predicates). Then any property F expressible in $\mathfrak{L}_N$ is represented either by a Q-predicate or by a disjunction of two or more Q-predicates. Logical width characterizes the logical complexity of a property F. The greater the logical width of a property, the greater is the number of possible (non-contradictory) properties it admits. For example, the property $P_1 \vee P_2$ is wider than P_1 since property $\sim P_1 \& P_2$ is admitted by the first but excluded by the second. Thus, the logical width of a contradictory property is 0 while the logical width of a property represented by a Q-predicate is 1. Any property F that is expressed as a disjunction of Q-predicate has a logical width $\kappa \geq w > 1$ equal to the number of disjuncts. Moreover, the *relative width F* is the ratio w/κ. Notice that the relative width varies from 0, for a contradictory property, through ½ , for any property represented by a atomic predicate, to 1 for a logically necessary property.

Let e be the sentence expressing that out of s individuals examined, s_F had property F and h be the hypothesis that a given individual different that those examined so far had also F, then the degree of confirmation $\mathfrak{c}(h, e)$ is

$$\mathfrak{c}(h, e) = \left(\frac{s}{s + \lambda}\right)\frac{s_F}{s} + \left(\frac{\lambda}{s + \lambda}\right)\frac{w}{\kappa},$$

where s_F/s is the relative frequency of observed instances of the property F and λ a non-negative real number (Burks, 1953). The relative frequency of observed instances, s_F/s, is an empirical fact while the relative width of the property is a logical fact depending on the language system and the predicate that represents the property. Hence, the degree of confirmation is determined as a mixture of a logical factor and of an empirical factor (1952: 24):

$$\mathfrak{c}(h, e) = (1 - a)\frac{s_F}{s} + a\frac{w}{\kappa},$$

where $a = \frac{\lambda}{s+\lambda}$. If no observation has taken place, i.e., $s = 0$, then $\mathfrak{c}(h, e) = \frac{w}{\kappa}$, and the degree of confirmation is determined on logical grounds. As the number of observations increases relative frequency of observed instances acquires significance and the degree of confirmation tends toward $\frac{s_F}{s}$. Exactly how fast we learn from experience, that is how fast $\mathfrak{c}(h, e)$ tends to $\frac{s_F}{s}$, depends on λ. In the following table we

have summarized the degrees of confirmation that correspond to different characteristic values of λ

λ	$\mathfrak{c}(h, e)$
0	$\dfrac{s_F}{s}$
κ	$\dfrac{s_F + w}{s + \kappa}$
$\lambda \to \infty$	$\dfrac{w}{\kappa}$

For $\lambda = 0$, we have the *straight rule* which stipulates that the observed relative frequency is equal to the probability that an unobserved individual has the property in question. Carnap says that the straight rule is problematic since it yields complete certainty ($\mathfrak{c} = 1$), if all examined individuals are found to possess the relative property ($s_F = s$) – a conclusion that may be accepted if the size s of the sample is quite large but not otherwise (1950: 227). The second row in our table ($\lambda = \kappa$) is better interpreted if we assume that our language system consists of one atomic unary predicate only. Then $w = 1$ and $\kappa = 2$, and we get Laplace's rule of succession, $\mathfrak{c}(h, e) = \mathfrak{c}^*(h, e)$. Finally, with the same assumptions about the language system, for $\lambda \to \infty$ the logical factor reigns and $(h, e) = \mathfrak{c}^+(h, e) = \frac{1}{2}$, as calculated for equiprobable state descriptions.

How can we decide which of the uncountable infinity of inductive methods is the appropriate one? Carnap's answer is based on two important elements: (a) adopting an inductive method is a matter of *choice* that we make; (b) this choice is made on *a priori* grounds. Carnap agreed with Burks' suggestion to apply to induction the internal-external distinction concerning the adoption of frameworks (1963: 982). Thus, while the degree of confirmation for a given hypothesis on given evidence is an *internal* question, it presupposes the adoption of a $\mathfrak{c}$-function, the choice of which is an *external* one; i.e., it is raised outside any inductive system and has to do with the choice of a framework similar to the choice of a language system. Richard Jeffrey (1992: 28) pointed out that:

> Carnap counted the specification of $\mathfrak{c}$-functions among the semantical rules for languages. Choice of a language was a framework question, a practical choice that could be wise or foolish, and lucky or unlucky, but not true or false.

The pragmatic (i.e., non-cognitive) nature of the scientist's choice of an inductive method becomes apparent in the passage below:

> *X* may change this instrument [i.e., their inductive method] just as he changes a saw or an automobile, and for similar reasons. (Carnap 1952: 55)

It is up to the scientists to make up their minds and to choose among them the one that they feel are the more appropriate for their purposes. They can change them as they change their automobiles!

Assuming that a choice of an inductive method has been made and a particular $\mathfrak{c}$-function has been defined, any statement of the sort "$\mathfrak{c}(h, e) = p$" for specified sentences h, e, is *analytic,* if true (and contradictory, if false), i.e., their truth or falsity

rests on definition and pure logic. This fact raises additional problems regarding the justification of the applicability of the inductive methods to practical issues: "The question is", says Salmon (1966:76), "*How can statements that say nothing about any matters of fact serve as 'a guide of life'?*" The observation that non-trivial empirical content is introduced by the synthetic sentence *e* expressing evidence of past experience, does not improve things very much. For, one may further require a justification of considering past evidence and logico-mathematical facts about the degree of a confirmation as a guide to predictions and our *future* conduct. On what grounds do we deem such a practice *rational?* Nevertheless, these last questions seem to get us outside the limits of any framework since they are reformulations of the external question about the choice of a particular c-function, and can be answered neither from reason nor from experience.

Where does all this leave Carnap's project? The project of specifying *the* inductive logic falls apart. There is no uniquely rational way to determine the relations between evidence and hypotheses. Instead, Carnap's attitude seems to be captured by the following paraphrase of Chairman Mao's famous dictum: 'Let a hundred inductive methods bloom'. But even if we were to argue that we end up with a plurality of inductive methods, they would still fall short of being inductive *logics*. As we saw, the c-function depends on the parameter λ. But, as Howson and Urbach (1989: 55) have stated, the very idea of an adjustable parameter λ "calls into question the fundamental role assigned to his systems of inductive logic by Carnap. If their adequacy is itself to be decided empirically, then the validity of whatever criterion we use to assess that adequacy is in need of justification, not something to be accepted uncritically".

5. Subjective Probability and Bayesianism
a. Probabilities as Degrees of Belief
Subjective theory is a theory of inductive probability proposed by the Cambridge Apostle F. P. Ramsey in his paper "Truth and Probability", written in 1926 and published in 1931, and, *independently*, by the Italian mathematician, Bruno de Finetti, who proposed it somewhat later, in 1928, and published it in a series of papers in 1930. In this conception, *probability is the degree of belief of an individual at a given time*. The inductive nature of the account is reflected in de Finetti's (1972: 21) that:

> [t]he subjectivists … maintain that a probability evaluation, being but a measure of someone's beliefs, is not susceptible of being proved or disproved by the facts …

A major assumption of the theory is that beliefs, commonly conceived as psychological states, are *measurable*, otherwise as Ramsey put it "all our inquiry will be vain" (1926:166). Thus, one needs to specify a method of measuring belief to consider the sentence 'the degree of belief of X, at time t, is *p*' meaningful. Ramsey examined two such methods. The first one is based on the fact that the degree of belief is perceptible by its owner, since one ascribes different intensities of feelings of conviction to different beliefs that they hold. However, as Ramsey noted, we do not have strong feelings for things we take for granted, actually, such things are practically accompanied by no feeling; thus, this way of measuring degree of belief seems inadequate. The second method rests on the supposition that the degree of belief is a causal property and:

the difference [in the degree of belief] seems to me to lie in how far we should act on these beliefs (*ibid*: 170).

To measure beliefs as bases for actions Ramsey (*ibid*: 172) suggested:

to propose a bet and see what are the lowest odds which… [the agent] will accept.

In a similar vein, de Finetti (1931) characterized probability "the psychological sensation of an individual" and also suggested to use bets to measure degrees of belief.

A *bet* on a hypothesis h, with *betting quotient p*, at *stake S*, $bet(h, p, S)$, is defined by the following conditions:

(a) if hypothesis h is true, the gambler wins $(1 - p)S$;
(b) if hypothesis h is false, the gambler loses pS,

where p is any real number in the unit interval and S any sum of money. We say that the *odds* in a bet on h at stake S are $R : Q$ whenever the betting quotient $p = R/(R + Q)$.

h	AGENT PAYS	AGENT RECEIVES	NET PAYOFF FOR THE AGENT
T	pS	S	$(1 - p)S$
F	pS	0	$-pS$

The actions that measure an agent's degree of belief in a hypothesis h are the buying and selling of a bet on h. In particular, the *degree of belief* of an individual X in a hypothesis h is a number p_0 which, expressed in monetary values, $\$p_0$, is (i) the *highest* price X is willing to buy a bet that returns $\$1$ if h is true, and $\$0$ if h is false, and, (ii) the *lowest* price, X is willing to sell that same bet.

To better understand this definition, consider the set of all bets on h at stake $\$1$. It can be characterized in terms of the betting quotients as follows:

$$\{p \in \mathbb{R} : bet(h, p, \$1)\}$$

To buy any bet from this collection the bettor should pay $\$p$. But depending on h they are not willing to pay any amount of money; on the contrary they seek to pay the least possible. The definition assumes that the amount of money the agent is willing to pay to buy the bet is bounded from above and its least upper bound is $\$p_0$. Similarly, the money an agent could earn from selling the bet is bounded from below and the greatest lower bound is also $\$p_0$. This number p_0 is the degree of belief of an agent in h.

On this view, the conditional degree of belief of an individual X in a hypothesis h given some statement e, $b_X(h|e) = p_0$ is defined in terms of the following bet:

(a) if hypothesis $h\&e$ is true, the bettor wins $\$(1 - p_0)$;
(b) if hypothesis e is false, the bettor wins $\$p_0$

The idea for this bet is that it is called off in case e is false and the agent gets a refund of $\$p_0$. (Jeffrey 2004: 12)

The degree of belief p_0 of an individual X in a hypothesis h is confined within the unit interval. To see this, assume, first, that $p_0 < 0$ and consider the agent *selling* a

bet to the bookie that pays \$1 if h is true, and \$0 if h is false, for \$$p_0$. Independently of the truth-value of h, this bet is a loss for the agent: the agent has a net gain of \$$(-1 + p_0) < 0$ in case h is true and \$$p_0 < 0$ in case h is false. In a similar vein, if $p_0 > 1$, an agent *buying* a bet from the bookie that pays \$1 if h is true, and \$0 if h is false, for \$$p$, gains \$$(1 - p_0) < 0$ if h is true, and \$$- p_0 < 0$ if h is false, and the bet is, again, a loss for the agent. Hence, if an agent assigns to any of their beliefs degrees that are either negative or greater than 1, they are exposed to a betting situation with guaranteed loss independently of the truth or the falsity of that belief. Such an unwelcome bet or set of bets which "will with certainty result in a loss" (de Finetti, 1974: 87) for the agent is called *Dutch book*. It is conjectured that the term can be traced back to the introduction of the Lotto game in the Low Countries, at the beginning of the 16th century where in the so-called "Dutch Lotto", the organizer had, in any event, a positive gain (de Finetti, 2008: 45). Hence, to avoid a Dutch book, one should confine degrees of belief within the interval [0,1].

A *degree of belief function* b_X is an assignment of degrees of belief of a person X's beliefs as represented by propositions (or, classes of logically equivalent sentences, in a language dependent context):

$$\mathcal{S}_L \ni h \mapsto b_X(h) \in [0,1].$$

For an agent X with an assignment of degrees of belief described by the function b_X, we may define the *expected winnings of a $bet(h, p, S)$ for X*, as a convex combination of the gains and losses of the agent on this bet with coefficients determined by their degree of belief in h :

$$EW[\, bet(h, p, S), X] = b_X(h)V(h) + \big(1 - b_X(h)\big)V(\sim h).$$

where $V(h)$ is the net payoff for the agent if h is true and $V(\sim h)$, the net payoff if h is false. To understand this concept, think of $V(h)$ and $V(\sim h)$ as the possible states in which an agent that their belief function assigns 1 and 0 to h, respectively, expects to be found if the bet offered is accepted. Namely, an agent that is certain of the truth of h, expects to gain $V(h)$ an agent that is certain of the falsity of h, expects to gain $V(\sim h)$ by accepting the bet. If the agent's belief function assigns any other number in the unit interval to h, they will occupy an intermediate state. Geometrically, $V(h)$ and $V(\sim h)$ may be thought as the extremities of a line segment and any other state a point between these extremities. Next, assume that the agent is placed on the midpoint of the segment, equidistant from its extremities. Then the bet doesn't give any prevalence beforehand to the truth or the falsity of the hypothesis for that particular agent and it is fair. If the agent's belief function places them closer to either of the extremities, $V(h)$ or $V(\sim h)$, then the gives an unfair advantage for or against h, for this agent. Thus, for $b_X(h) = p_0$, the expected winnings of a $bet(h, p, S)$ *for X* is:

$$\big(p_0 - p\big)S$$

and it measures how much fair or unfair is the bet for that particular agent. In this understanding, no commitment to a probabilistic view of the belief function is required. It is sufficient to treat belief quantitatively, to consider the degree of belief on a hypothesis a number in the closed interval and to interpret the values 0 and 1 in terms of the belief in the falsity and truth of the hypothesis respectively.

Accordingly, we may now give the following definitions:

- We call $bet(h, p, S)$ a *fair bet for X* if and only if $EW[\,bet(h, p, S), X] = 0$.
- We call $bet(h, p, S)$ *advantageous for X* if and only if $EW[\,bet(h, p, S), X] > 0$.
- We call $bet(h, p, S)$ *disadvantageous for X* if and only iff $EW[\,bet(h, p, S), X] < 0$.

Notice that the Dutch book in which we would be vulnerable were we to consider degrees of belief outside the unit interval, is fair, since it is defined in terms of buying and selling $bet\big(h, p_0, S\big)$ – a fact that makes its bite even worse.

b. Dutch Books

Ramsey identified a connection between Dutch books and the laws of mathematical probability. In "Truth and Probability" we read that (1926: 182):

> If anyone's mental condition violated these laws [of probability] … [h]e could have a book made against him by a cunning bettor and would then stand to lose in any event.

And conversely,

> Having degrees of belief obeying the laws of probability implies a further measure of consistency, namely such a consistency between the odds acceptable on different propositions as shall prevent a book being made against you (1926: 183).

Instead of Ramsey's 'consistency', de Finetti (1974: 87) has spoken of 'coherence' of degrees of beliefs. The degrees an agent assigns to his beliefs are said to be *coherent* :

> if among the combinations of bets which [y]ou have committed yourself to accepting there are none for which the gains are all uniformly negative.

Thus, if an agent is not vulnerable to a Dutch book with betting quotients equal to their degrees of belief, the agent is said to have coherent degrees of belief. In addition, an agent has coherent degrees of belief if and only if their degrees of belief satisfy the axioms of probability. This is the celebrated Ramsey – de Finetti or Dutch-Book theorem:

> *Let $b_X: \mathcal{S}_L \longrightarrow \mathbb{R}$ be a degree of belief function of a person X. If b_X does not satisfy the axioms of probability, then there is a family of fair bets $bet(h_i, p_i, S_i)$, with $h_i \in \mathcal{S}_L$, $p_i = b_X(h_i)$ and $S_i \in \mathbb{R}$, for every $i = 1, \ldots, n$ (or ∞) which guarantees that the agent will result in an overall loss, independently of the truth-values of the hypotheses h_i.*

The converse of that theorem has also been shown:

Let $b_X : \mathcal{S}_L \to \mathbb{R}$ be a degree of belief function of a person X. If b_X satisfies the axioms of probability, then there is no family of fair bets $bet(h_i, p_i, S_i)$, with $h_i \in \mathcal{S}_L$, $p_i = b_X(h_i)$ and $S_i \in \mathbb{R}$, for every $i = 1, \dots, n$ which guarantees that the agent will result in an overall loss, independently of the truth-values of the hypotheses h_i.

We have already discussed the application of the Ramsey-de Finetti theorem in the case of violation of the axiomatically imposed constraint that probability values lie within the unit interval. The next example illustrates how an agent will experience an overall loss if they hold degrees of belief that do not comply with the finite additivity axiom.

Consider the tossing of a die and assume that the degrees of belief assigned by a person X to the beliefs that they will obtain: '6' in a single toss is q; '3' in a single toss is r; and, either '6' or '3' is k. Moreover, let $k < r + q$, i.e., finite additivity axiom is violated. Then we may consider the following family of fair bets, suggested to the agent:

$$bet('6', q, 1), bet('3', r, 1), bet('6' or '3', k, -1).$$

The agent *buys* from the bookie $bet('6', q, 1)$ that pays \$1, if "'6' is obtained" is true, and \$0, if false, for \$q. Next, the agent *buys* the second bet, $bet('3', r, 1)$, that pays \$1, if "'3' is obtained" is true, and \$0, if false, for \$r. Finally, in the third bet, the agent *sells* to the bookie $bet('6' or '3', k, -1)$ that pays \$1, if "'6' or '3' is obtained" is true, and \$0 if false, for \$k. In the following table, is calculated the net gain for the agent in this betting sequence:

"'6',"	"'3',"	"'6', OR '3',"	NET GAIN FOR THE AGENT
T	F	T	$(1 - q) \cdot 1 + (-r) \cdot 1 + (1 - k)(-1) = k - (r + q)$
F	T	T	$(-q) \cdot 1 + (1 - r) \cdot 1 + (1 - k)(-1) = k - (r + q)$
F	F	F	$(-q) \cdot 1 + (-r) \cdot 1 + (-k)(-1) = k - (r + q)$

As we can see, this sequence of bets results in an overall loss for the agent. Thus, as the Ramsey-de Finetti theorem demands, an agent whose degree of belief function violates the axiom of finite additivity is exposed to a Dutch book.

One could obtain a similar result for the violation of countable additivity axiom. In this case they need to employ a countable infinite family of bets. However, a criticism that follows such an assumption is that it is unrealistic for any agent to be engaged in infinitely many bets. (Jeffrey, 2004: 8)

There have been attempts to extend the requirement of coherence from the *synchronic* case, as expressed by the compliance of the degrees of belief with the axioms of probability theory, to *diachronic coherence* by stipulating rules for belief updating. Learning from experience requires that the agent should change their assignment of degree of belief (probability) on a given hypothesis in response to the result of experiment or observation. The simplest, and most common, rule for updating is the following:

In the light of new evidence, the agent should update their degrees of beliefs by *conditionalizing* on this evidence.

Thus, assume that the belief function of a person X before new evidence e is acquired is $b_{X_{old}}$ and $b_{X_{new}}$ is the belief function after the acquisition of new evidence. The transition from the old degree of belief to the new one is governed by the rule:

$$b_{X_{new}}(h) = b_{X_{old}}(h|e)$$

where e is the total evidence, and $b_{X_{old}}(h|e)$ is the posterior probability as determined by Bayes's Theorem if we identify the degree of belief function with the probability function.

This form of conditionalization is called *strict conditionalization* and it takes the probability of the learned evidence to be unity, i.e., $b_{X_{new}}(e) = 1$. Jeffrey found out that certainty is a very restrictive condition that does not conform with the uncertainties of real empirical research in science and everyday life. To show that Jeffrey suggested the example of observing the color of a piece of cloth by candlelight. The agent gets the impression that the observed color is green, but they concede that it maybe blue or less probably violet. The experience causes as to change our degrees of belief in propositions about the color of the object but does not cause us to change them to 1. Hence, strict conditionalization is inapplicable for updating our degrees of belief. Jeffrey suggested another form of conditionalization that tackles the problem, known as *Jeffrey-conditionalization* (or, *probability kinematics,* as Jeffrey called it), which considers evidence as providing probabilities to a partition of our set of beliefs. In this case, the new degree of belief function is calculated in terms of the old one,

$$b_{X_{new}}(h) = \sum_i^n b_{X_{old}}(h|e_i)p_i,$$

where $\{e_i\}_{i=1}^n$ is a partition of our set of beliefs consisting mutually exclusive and jointly exhaustive propositions and $p_i = b_{X_{new}}(e_i)$, $i = 1, \dots, n$, are the probabilities assigned to propositions e_i by new evidence. As before, $b_{X_{old}}(h|e_i)$ is calculated as the posterior probability in Bayes's Theorem.

One difficulty with Jeffrey's conditionalization is that while strict conditionalization provides an assurance to convergence to truth, Jeffrey's conditionalization generally doesn't. There is a family of theorems, known as convergence theorems, with the most well-known being that of Gaifman and Snir (1982), which claim that, under reasonable assumptions, the probability of a hypothesis conditional on available evidence converges to 1 in the limit of empirical research, if the hypothesis is true. These theorems provide a vindication of Bayesianism showing that it is guaranteed to find the truth eventually by applying successively strict conditionalization.

Conditionalizing on the evidence is purely *logical* updating of degrees of belief. It is not ampliative. It does not introduce new content, nor does it modify the old one. It just assigns a new degree of belief to an old opinion. The justification for the requirement of conditionalization is supposed to be a diachronic version of the Dutch-book theorem. It is supposed to be a canon of rationality (certainly a necessary condition for it) that agents should update their degrees of belief by conditionalizing on evidence. The penalty for not doing this is liability to a *Dutch-book strategy*: the agent can be offered a set of bets *over time* such that a) each of them taken individually will seem fair to them at the time it is offered; but b) taken collectively, they lead them to suffer a net loss, come what may.

c. Bayesian Induction

In this context, induction rests on the degree of belief one assigns to a hypothesis given a body of confirmatory evidence and on the process of updating in the light of new evidence. Hence, the problem of justification of induction gives way to the problem of justifying conditionalization on the evidence. In general, Bayesian theories of confirmation maintain the following theses:

(a) Belief is always a matter of degree; degrees of belief are probability values and degree of belief functions are probability functions.

(b) Confirmation is a relation of positive relevance, viz., a piece of evidence confirms a hypothesis if it increases its probability;
$$e \text{ confirms } h \text{ iff } p(h|e) > p(h),$$
where p is a probability function.
Similarly, we may define *disconfirmation* of a hypothesis by a piece of evidence in terms of negative relevance $(p(h|e) < p(h))$, as well as *neutrality* of a hypothesis with respect to a piece of evidence in terms of irrelevance $(p(h|e) = p(h))$.

(c) The relation of confirmation is captured by Bayes's theorem which dictates the change of the degree of belief in a given hypothesis in the light of a piece of evidence.
$$p(h|e) = \frac{p(e|h)p(h)}{p(e)}, \text{ where } p(h), p(e) > 0,$$

(d) The only factors relevant to confirmation of a hypothesis are its prior probability $p(h)$, the likelihood of the evidence given the hypothesis $p(e|h)$; and the probability of the evidence $p(e)$.

(e) The specification of the prior probability of (*aka* prior degree of belief in) a hypothesis is a purely subjective matter.

(f) The only (logical-rational) constraint on an assignment of prior probabilities to several hypotheses should be that they obey the axioms of the probability calculus.

(g) The reasonableness of a belief does not depend on its content; nor, ultimately, on whether the belief is made reasonable by the evidence.

d. Too Subjective?

In 1954, Savage discussed a criticism of subjective Bayesianism based on the idea that science or scientific method aims at finding out "what is probably true, by criteria on which all reasonable men agree." (1954:67). By applying intersubjectively accepted criteria, scientific method is supposed to lead to an agreement between any two rational agents on the probability for the truth of a hypothesis given the same body of evidence. According to Savage this demand for intersubjectivity has its source either in considering probabilistic entailment as a generalization of logical entailment, or in considering probability an objective property of certain physical systems. Yet, the criticism goes, complete freedom in the choice of prior probabilities for a hypothesis by two agents may yield different posterior probabilities for that hypothesis given the same body of evidence. This fact compromises the desideratum of intersubjectivity of criteria since it makes room for the intrusion of idiosyncratic elements, non-cognitive values, or any other source of subjective preferences, reflected in the disagreement of the agents in the choice of priors, and, ultimately, in the value of posterior probability of a hypothesis. Hence, what is "probably true" is not evaluated by "criteria on which

all people agree". In a nutshell, it is claimed that purely subjective prior probabilities fail to capture the all-important notion of rational or reasonable degrees of belief and that subjective Bayesianism is *too* subjective to offer an adequate theory of confirmation.

In defense of subjective probability, Savage claims that although this view

> incorporates all the universally acceptable criteria for reasonableness in judgement… [these criteria] do not guarantee agreement on all questions among all honest and freely communicating people, even in principle (*ibid*),

considering disagreements a non-distressful situation. Moreover, anticipating what later became known as *convergence-to-certainty* or *merger-of-opinions* theorems, he showed that:

> …in certain contexts any two opinions, provided that neither is extreme in a technical sense, are almost sure to be brought very close to one another by a sufficiently large body of evidence. (1954: 68; see also 46f)

Yet, as Hesse (1975; see Earman 1992:143) objected, Savage's argument makes assumptions that are valid for the flipping of a coin case but are not typically valid in scientific inference. Gaifman and Snir (1982) have shown important results which overcome the limitations of Savage's account. They have shown (Thm. 2.1) that for an infinite sequence of empirical questions, $\varphi_1, \ldots, \varphi_n, \ldots,$ formulated in a given language that satisfies certain conditions:

- *Convergence-to-certainty:* The limiting probability of a true sentence ψ in that language, given all empirical evidence collected in our world w, in response to empirical questions stated, $\varphi_1^w, \ldots, \varphi_n^w, \ldots,$ equals to 1, $\lim_{n\to\infty} \Pr(\psi|\&_{i\leq n}\varphi_i^w) = 1$. For a false proposition, the respective probability is 0, $\lim_{n\to\infty} \Pr(\psi|\&_{i\leq n}\varphi_i^w) = 0$.
- *Merger-of-opinions:* The distance between any two probability functions that agree to assign probability 0 to the same sentences, i.e., they are equally dogmatic, converges to 0, in the limit of empirical research, i.e., $\lim_{n\to\infty} sup_\psi |\Pr_1(\psi|\&_{i\leq n}\varphi_i^w) - \Pr_2(\&_{i\leq n}\varphi_i^w)| = 0$.

Merger-of-opinions theorem is supposed to mitigate the excessive subjectivity of Bayesianism in the choice of prior probabilities: the actual values assigned to prior probabilities do not matter much since they 'wash out' in the long run.

Unfortunately, several criticisms of the theorem showed that the objection of subjectivism is not fully addressed. Let us briefly review some of these criticisms: The first objection is related to the asymptotic character of convergence and merging and the fact that the speed of convergence is unknown. The results do not apply to the divergences of opinion induced by small and medium-sized sets of evidence that have practical importance. The second objection is related to the language-dependent nature of the theorems restricting them to cases in which the predicates of the language are fixed. The theorems cannot guarantee washing out the priors assigned by agents in different linguistic contexts, as before and after a scientific revolution.

An important criticism stems from the fact that convergence in the theorems is obtained *almost everywhere,* i.e., for all worlds w, the actual world included, which belong to some set of possible worlds with probability 1. In the authors' own words:

> … with probability 1, two persons holding mutually nondogmatic initial views will, in the long run, judge similarly… Also the convergence is guaranteed with probability 1, where "probability" refers to the presupposed prior. (I) and (II) [referring to the two parts of the theorem] form an "inner justification" but they do not constitute a justification of the particular prior.

So, the theorem guarantees convergence to truth and merging of opinions in every world except for some pathological cases that form small sets of worlds of measure zero. But who decides what those sets of worlds of measure zero would be? The Bayesian agent themselves through the choice of priors who is compelled to assign probability zero to 'unpleasant' scenarios. On these grounds, Earman claims that the "impressiveness of these results disappears in the light of their narcissistic character… 'almost surely' sometimes serves as a rug under which some unpleasant facts are swept" (1992:147).

Extending on this criticism, Belot (2013; 2017) has argued that in problems of convergence to truth, there are *typical* cases – their typicality being defined in a topological sense without measure-theoretic presuppositions – in which convergence to truth is unsuccessful, a fact that a Bayesian agent is bound to ignore by assigning prior probability zero to such cases. Thus, Belot, concludes, convergence – merger theorems "constitute a real liability for Bayesianism by forbidding a reasonable epistemological modesty" (2013)

Belot's arguments have prompted a variety of responses: some philosophers were critical of Belot's topological considerations as being irrelevant to probability theory (Cisewski *et al.* 2018; Huttegger 2015). Others focused on imprecise probabilities and finitely additive probabilities to escape the charge of immodesty (Weatherson 2015; Elga 2016; Nielsen and Stewart 2019). Huttegger (2021) has shown using non-standard analysis that "convergence to the truth fails with (non-infinitesimal) positive probability for certain hypotheses … [a fact] that creates a space for modesty within Bayesian epistemology." As regards the countable additivity of the probability function, convergence-to-certainty and merger-of opinions theorem relies essentially on this axiom. Prominent subjective Bayesians, on the other hand, such as de Finetti and Savage, explicitly reject countable additivity axiom despite its theoretical fecundity. Yet Savage, as mentioned above, has explored the possibility of theorems that despite their shortcomings attempt to mitigate the extreme subjectivism of Bayesianism. Recently, Nielsen (2021) has shown that there are uncountably many merely finitely additive probabilities that converge to the truth almost surely and in probability. As a general comment, we would say that the area convergence and merger theorems seems to have many open problems to capture the interest of researchers.

e. Some Success Stories

Bayesian theory has a record of successful justifications of some important common intuitions about confirmation – such as the belief that a theory is confirmed by its observational consequences or the belief that a theory is better confirmed if subject to strict tests – and it has provided a solution to the famous 'raven paradox'.

It is straightforward to show that hypotheses are confirmed by their consequences. Assume that $h \vdash e$, then the likelihood of e given h is $p(e|h) = 1$ and according to Bayes theorem, $p(h|e) = \frac{p(e|h)\,p(h)}{p(e)} = \frac{p(h)}{p(e)} > p(h)$, given that e is not trivially true ($p(e) < 1$); hence, e confirms h. This result justifies the inference of the truth of a hypothesis on the basis of its observational consequences as the hypothetico-deductive method of confirmation suggests. Although the inference commits the

formal fallacy of affirming the consequent, if considered inductively, through the lenses of Bayes's theorem, it is fully justified and the confirmatory nature of the hypothetico-deductive method is explained. This is what Earman recognized as an important "success story" of the Bayesian approach (1992: 233)

Another common methodological intuition that may be justified on Bayesian grounds is related to the scientific practice of subjecting a hypothesis to severe tests on the basis of improbable consequences. As Deborah Mayo (2018: 14), following Popper, suggested in her Strong Severity Principle:

> We have evidence for a claim C just to the extent it survives a stringent scrutiny. If C passes a test that was highly capable of findings flaws or discrepancies from C, and yet none or few are found, the passing result, x, is evidence for C.

Now, as before, consider a logical consequence e of a hypothesis h. i.e., $h \vdash e$. A severe test of h would be one in which $p(\sim e)$ is high and, consequently, $p(e)$ is low. In this case e would be evidence for h. Hence, a necessary condition for collecting evidence for a hypothesis according to the aforementioned principle, would be to test its improbable consequences. Indeed, following Bayes's theorem:

$$p(h|e) = \frac{p(e|h)\,p(h)}{p(e)} = \frac{p(h)}{p(e)}.$$

Thus, the more improbable the consequence e is, the greater the degree of confirmation, as measured by the ratio $\frac{p(h|e)}{p(h)}$, is.

Another piece in the collection of trophies of the Bayesian account is the resolution of the ravens paradox. This is a paradox of confirmation, first noted by Carl Hempel, which took its name from the example that Hempel used to illustrate it viz., *all ravens are black*. The paradox emerges from the impossibility of having jointly satisfied three intuitively compelling principles of confirmation. The first is Nicod's principle [named after the French philosopher Jean Nicod]: a universal generalization is confirmed by its positive instances. So, that all ravens are black is confirmed by the observation of black ravens. *Second*, the principle of logical equivalence: if a piece of evidence confirms a hypothesis, it also confirms its logically equivalent hypotheses. *Third*, the Principle of relevant empirical investigation: hypotheses are confirmed by investigating empirically what they assert.

To set up the paradox, take the hypothesis h: All ravens are black. The hypothesis h': *All non-black things are non-ravens* is logically equivalent to h. A positive instance of h' is a white piece of chalk. Hence, by Nicod's condition, the observation of the white piece of chalk confirms h'. By the *principle of equivalence*, it also confirms h, that is that all ravens are black. But then the principle of relevant empirical investigation is violated. For, the hypothesis that all ravens are black is confirmed not by examining the colour of ravens (or of any other birds) but by examining seemingly irrelevant objects (like pieces of chalk or red roses). So at least one of these three principles should be abandoned, if the paradox is to be avoided.

To resolve the ravens paradox, a Bayesian may show that there is no problem with accepting all three principles of confirmation since the degree of confirmation conferred on the hypothesis h by an instance of a non-raven-non-black object is negligible in comparison with how much the hypothesis is confirmed by an instance

of a black object.[According to Howson and Urbach (2006: 100) a Bayesian analysis could also challenge the adequacy of Nicod's criterion as a universal principle of confirmation.]

To see that consider hypotheses $h: \forall x(Rx \to Bx)$ and $h': \forall x(\sim Bx \to \sim Rx)$ and evidence $e: Ra\&Ba$ and $e': \sim Ba\&\sim Ra$ which are positive instances of h, h' respectively. We calculate the ratio $p(h|e)/p(h|e')$ which according to Bayes's theorem and the easily verifiable equality of likelihoods of e and e' given h, $p(e|h) = p(e'|h)$, is $\dfrac{p(h|e)}{p(h|e')} = \dfrac{p(e)}{p(e')}$. But $p(e') \gg p(e)$ because there are very many more things which are non-Black and non-Ravens than Black Ravens. Hence, $p(h|e) \gg p(h|e')$, i.e e confirms h a lot more than e' confirms h'.

We are closing this presentation of subjective probability and Bayesian confirmation theory by referring to what has become known as the *old evidence problem*. The problem has been identified for the first time by Glymour (1980) and it underlines a potential conflict between Bayesianism and scientific practice. Suppose that a piece of evidence e is already known (i.e., it is an old piece of evidence relative to the hypothesis h under test). Its probability, then, is equal to unity, $p(e) = 1$. Given Bayes's theorem, it turns out that this piece of evidence does not affect at all the posterior probability, $p(h|e)$, of the hypothesis given the evidence; the posterior probability is equal to the prior probability, i.e., $p(h|e) = p(h)$. This, it is argued, is clearly wrong since scientists typically use known evidence to support their theories. This fact is demonstrated by the use of the anomalous precession of Mercury's perihelion, discovered in the nineteenth century, as confirming evidence for Einstein's General Theory of Relativity. Therefore, the critics conclude, there must be something wrong with Bayesian confirmation. Some Bayesians have replied by adopting a counterfactual account of the relation between theory and old evidence (Howson and Urbach 2006: 299). Suppose, they argue, that K is the relevant background knowledge and e is an old (known) piece of evidence—that is, e is actually part of K. In considering what kind of support e confers on a hypothesis h, we subtract counterfactually the known evidence e from the background knowledge K. We therefore presume that e is not known and ask: what would the probability of e given $K\backslash\{e\}$? This will be less than one; hence, the evidence e can affect (that is, raise or lower) the posterior probability of the hypothesis.

6. Appendices
a. Lindenbaum algebra and probability in sentential logic.

In this appendix we show how one can assign probabilities, originally defined in set-theoretic framework, to sentences in the language of sentential logic, L. We formulate Kolmogorov's axioms of probability for sentences and some important theorems.

In particular, consider the set of all well-formed formulas (wffs) of L and define for every wff ϕ the equivalence class:

$$[\phi] = \{\psi : \vdash_L \phi \equiv \psi\}.$$

In the set of all equivalence classes $\mathcal{S}_L$, we define set-theoretic operations that correspond to the sentential connectives of the language. Thus, for every two wffs ϕ, ψ:

$$[\phi] \cup [\psi] = [\phi \vee \psi]$$
$$[\phi] \cap [\psi] = [\phi \wedge \psi]$$
$$[\phi]^c = [\sim\phi]$$
$$[\bot] = \emptyset$$
$$[t] = \{\text{wffs of } L\}$$

where "$\bot$" designates a contradiction and "t" a tautology. This way constructed, the set of all equivalence classes, $\mathcal{S}_L$, is a field (and a Boolean algebra) (see section 1a), and it is called *Lindenbaum algebra* (Hailperin 1986: 30ff.). However, since in the language of sentential logic, infinitary operations, like $\phi_1 \vee \ldots \vee \phi_n \vee \ldots$, cannot be applied to wffs ϕ_i to produce other wffs, we cannot define in $\mathcal{S}_L$ the countably infinite union of classes of wffs. As a consequence, $\mathcal{S}_L$ is not a σ-field and the probability function that we are about to define does not satisfy countable additivity. So, this is an account of *elementary* probability theory. To discuss the full axiomatic apparatus of probability theory one needs to work in richer languages, which for present purposes is not deemed necessary.

So, we can define a probability function p that satisfies Kolmogorov's axioms (i)-(iii) on $\mathcal{S}_L$ and assign to each singular sentence of the language L the probability value of its equivalence class. Thus, for any sentences a, b and a tautotology t of L:

i. $p(a) \geq 0$;
ii. $p(t) = 1$;
iii. $p(a \vee b) = p(a) + p(b)$, where $a \vdash_L \sim b$.

As for the conditional probability of a sentences a given the truth of a sentence sentences b, we have:

$$p(a|b) = \frac{p(a \wedge b)}{p(b)}, p(b) \neq 0.$$

It is obvious from the discussion above that logically equivalent sentences have equal probability values:

if $\vdash_L a \equiv b$, then $p(a) = p(b)$.

We conclude this appendix with some useful theorems of the probability calculus which we state in sentence-based formalism, without proof:

1. The sum of the probability of a sentence and of its negation is 1:
$$p(\sim a) = 1 - p(a).$$
2. Contradictions ($\perp$) have zero probability:
$p(\perp) = 0.$
3. The probability function respects the entailment relation:
if $a \vdash_L b$, then $p(a) \leq p(b)$.
4. Probability values range between 0 and 1:
$0 \leq p(a) \leq 1$.
5. Finite Additivity Condition:
$$p(a_1 \vee \ldots \vee a_N) = p(a_1) + \cdots + p(a_N), \quad a_i \vdash_L \sim a_j, \quad 1 \leq i < j \leq N.$$
Corollary:
If $\vdash_L a_1 \vee \ldots \vee a_N$ and $a_i \vdash_L \sim a_j$, $1 \leq i < j \leq N$, $1 = p(a_1) + \cdots + p(a_N)$.
6. Theorem of total probability:
If $p(a_1 \vee \ldots \vee a_N) = 1$, and $a_i \vdash_L \sim a_j$, $i \neq j$, then $p(b) = p(b \wedge a_1) + \cdots + p(b \wedge a_N)$, for any sentence b.
Or in terms of conditional probabilities:
If $p(a_1 \vee \ldots \vee a_N) = 1$, $a_i \vdash_L \sim a_j$, $i \neq j$, and $p(a_i) > 0$ then $p(b) = p(b|a_1)p(a_1) + \cdots + p(b|a_N)p(a_N)$, for any sentence b.
Corollary 1:
If $\vdash_L a_1 \vee \ldots \vee a_N$ and $a_i \vdash_L \sim a_j$, $i \neq j$, then $p(b) = p(b \wedge a_1) + \cdots + p(b \wedge a_N)$.
Corollary 2:
$p(b) = p(b|c)p(c) + \cdots + p(b|\sim c)p(\sim c)$, for any sentence $c, p(c) > 0$.
7. Bayes's Theorem. The famous theorem that took its name after the eighteenth-century clergyman Thomas Bayes.
   - First form (Thomas Bayes):
   $$p(h|e) = \frac{p(e|h)p(h)}{p(e)}, \text{where } p(h), p(e) > 0,$$
   where $p(h|e)$ is called *posterior probability* and expresses the probability of the hypothesis h conditional on the evidence e; $p(e|h)$ is called *likelihood* of the hypothesis and expresses the probability of the evidence conditional on the hypothesis; $p(h)$ is called *prior probability* of the hypothesis; and $p(e)$ is the probability of the evidence.
   - Second form (Pierre Simon Laplace):
   If $p(h_1 \vee \ldots \vee h_N) = 1$ and $h_i \vdash_L \sim h_j$, $i \neq j$, and $p(h_i), p(e) > 0$ then
   $$p(h_k|e) = \frac{p(e|h_k)p(h_k)}{\sum_{i=1}^{N} p(e|h_i)p(h_i)}$$
   - Third form:

   $$p(h|e) = \frac{p(h)}{p(h) + \frac{p(e|\sim h)}{p(e|h)} p(\sim h)}$$

b. A sketch of proof for Laplace's Rule of Succession

Assume that we want to calculate the probability that the sun will rise tomorrow given that the sun has risen for the past N days. We have observation data about the sunrise in the past N days but the probability q of the sunrise is unknown. By application of the principle of indifference, we claim that it is equally likely that the probability of sunrise be any number $q \in [0,1]$. Hence, the distribution of probability values of sunrise is uniform.

We take the sample space to consist of (N+2)-ples of the following type:

$$< \overbrace{S, S, \dots, F, \dots, S}^{N+1}, q >,$$

where S, F stand for 'Success' and 'Failure' of the sunrise, respectively, and q denotes a possible value for the probability of the sun rising.

The subset of the sample space

$$E = \left\{ < \overbrace{S, \dots, S}^{N}, x, q > \;|\; x \in \{S, F\} \; and \; q \in [0,1] \right\},$$

is a random event consistent with observations of the sun rising in the past N days, no matter what is going to happen in the $(N + 1)$ day or what the probability q of the sunrise is.

Since, parameter q takes real values we should not ask what the probability of a given value k of the parameter q is, but what the probability of q to be found within a given interval is:

$$p(q \leq k | E).$$

To calculate this probability, we first apply Bayes' rule:

$$p(q \leq k | E) = \frac{p(q \leq k) \cdot p(E | q \leq k)}{p(E)}$$

Since all values of q in $[0,1]$ are equiprobable:

$$p(q \leq k) = k.$$

Since the sequence of past sunrises is a sequence of independent trials, i.e., whether the sun has risen or not in a given day does not influence the rising of the sun in subsequent days:

$$p(E | q \leq k) = \frac{k^N}{N + 1}$$

and

$$p(E) = \frac{1}{N + 1}$$

Hence:

$$p(q \leq k | E) = k^{N+1}.$$

From here, we can calculate the probability density function for $q = k$ conditional on E:

$$f(k) = (N + 1)k^N.$$

To yield the probability of the sun to rise in the $(N + 1)$ day, given that it has risen in the last N days, no matter what the probability of sunrise might be is given by the following integral:

$$\int_0^1 kf(k)dk = \left.\frac{(N + 1)k^{N+2}}{N + 2}\right|_0^1 = \frac{N + 1}{N + 2}.$$

c. The Mathematics of Keynes's Account of Pure Induction

Consider a generalization h: "$all\ A\ is\ B$" and n positive instances e_i: "$this\ A\ is\ B$" , $i = 1, \dots, n$ that follow logically from h, i.e., $h \vdash e_i$. Let $p(h|K)$ the prior to any evidence probability relative to background knowledge K. Background knowledge is understood as the body of evidence which is related to the truth of the hypothesis with the exception of the evidence that are being considered explicitly. If n positive instances e_i: "$this\ A\ is\ B$" , $i = 1, \dots, n$ and no negative instances have been observed, the posterior probability of h is $p(h|e_1\& \dots \&e_n\&K)$.

To justify inductive inference, Keynes claims, we need to find the conditions on which the posterior probability increases with the accumulation of positive instances and the absence of negative instances so that the inductive argument is strengthened and in the limit of empirical investigation, hypothesis h can be inferred with certainty on the basis of empirical evidence:

$$\lim_{n\to\infty} p(h|e_1\& \dots \&e_n\&K) = 1.$$

From Bayes's theorem we have:

$$p(h|e_1\& \dots \&e_n\&K) = \frac{p(h|K)\, p(e_1\& \dots \&e_n|h\&K)}{p(e_1\& \dots \&e_n|K)}.$$

Since $h \vdash e_i, i = 1, \dots, n$:

$$p(e_1\& \dots \&e_n|h\&K) = 1 \quad (1)$$

$$p(h|e_1\& \dots \&e_n\&K) = \frac{p(h|K)}{p(e_1\& \dots \&e_n|K)} \quad (2)$$

From the law of total probability, we have:

$$p(e_1\& \dots \&e_n|K) = p(e_1\& \dots \&e_n|h\&K)p(h|K) + p(e_1\& \dots \&e_n|{\sim}h\&K)p({\sim}h|K)$$

and by (1),

$$p(e_1\& \dots \&e_n|K) = p(h|K) + p(e_1\& \dots \&e_n|{\sim}h\&K)p({\sim}h|K) \quad (3)$$

Hence, by (2) and (3):

$$p(h|e_1\&\ldots\&e_n\&K) = \frac{p(h|K)}{p(h|K) + p(e_1\&\ldots\&e_n|\sim h\&K)p(\sim h|K)}$$

If $\lim\limits_{n\to\infty} \frac{p(e_1\&\ldots\&e_n|\sim h\&K)}{p(h|K)} = 0$, the requested condition of asymptotic certainty, $\lim\limits_{n\to\infty} p(h|e_1\&\ldots\&e_n\&K) = 1$, is satisfied. Since $p(h|K)$ is the prior probability of the hypothesis which is independent of the evidence accumulated, it is a fixed number.

Hence, the antecedent of the aforementioned conditional can be split into the following two conditions:

$$p(h|K) \neq 0 \ (4)$$

and

$$\lim\limits_{n\to\infty} p(e_1\&\ldots\&e_n|\sim h\&K) = 0 \ (5)$$

Condition (5) can be analyzed in terms of the probability of a positive instance e_j given $j-1$ positive instances for h, $e_1\&\ldots\&e_{j-1}$, and that h is false:

$$p(e_j|e_1\&\ldots\&e_{j-1}\&\sim h\&K) = q_j, \ \ j = 2,\ldots,n$$
$$p(e_1|\sim h\&K) = q_1.$$

The probability of n positive instances and no negative instances given that h is false is:

$$p(e_1\&\ldots\&e_n|\sim h\&K) = q_1\cdot\ldots\cdot q_n.$$

Let $1 > M_n = max\{q_1,\ldots,q_n\}$ then $p(e_1\&\ldots\&e_n|\sim h\&K) \leq M_n{}^n$. The sequence $\{M_n\}_{n\in\mathbb{N}}$ is bounded. If $= sup_{n\in\mathbb{N}} M_n$, $0 < M < 1$, then:

$$\text{for every } n \in \mathbb{N}, \ p(e_1\&\ldots\&e_n|\sim h\&K) \leq M_n{}^n < M^n$$
and (5) follows:

$$\lim\limits_{n\to\infty} p(e_1\&\ldots\&e_n|\sim h\&K) \leq \lim\limits_{n\to\infty} M^n = 0\,.$$

By contraposition we infer that if condition (5) is not satisfied, $\{M_n\}_{n\in\mathbb{N}}$ is not bounded by any number M, $0 < M < 1$. Thus, for every M there is a $n_0 \in \mathbb{N}$ such that $M_{n_0} > M$. Since $M_{n_0} = max\{q_1,\ldots,q_{n_0}\}$, we infer that for every M there is a $k \in \mathbb{N}$, $k < n_0$, such that:

$$1 > p(e_k|e_1\&\ldots\&e_{k-1}\&\sim h\&K) = q_k > M,$$

and

$$\lim\limits_{k\to\infty} p(e_k|e_1\&\ldots\&e_{k-1}\&\sim h\&K) = 1. \ (6)$$

Hence, if (5) is false then (6). But it is reasonable to demand that a negative instance of h, $\sim e_k$, should have non-zero probability no matter how many positive

instances have been observed given the falsity of h. Thus, Keynes (1921: 275) suggested that (6) is false:

> [given that] the generalisation is false, a finite uncertainty as to its conclusion being satisfied by the next hitherto unexamined instance which satisfies its premiss.

Or, as Russell commented referring to condition (5), "[i]t is difficult to see how this condition can fail in empirical material." (1948: 455).

Keynes justified the second condition, (4), by applying the principle of limited independent variety and the principle of indifference (see sections 3.a.1, 3.a.2). According to the principle of limited independent variety, qualities are classified into a finite number of groups so that two qualities that belong in the same group have the same extension, i.e., they are satisfied by the same individuals, and, in this sense, they are equivalent. More precisely, $[A]$ is the set of all qualities that are equivalent to A; it includes all qualities $B \in [A]$ which $(\forall x)(Ax \equiv Bx)$. Thus, generalization h is entailed logically by the assumption that A, B are equivalent properties. Moreover, the principle of limited variety requires that the number of independent qualities that are inequivalent is finite. Hence, if n is the number of independent qualities by the principle of indifference we conclude that the probability of any two properties A, B to belong in the same group is $1/n$. Since, h is a logical consequence of this fact, by a well-known theorem in probability theory (see section 1.a),

$$p(h|K) \geq \frac{1}{n}, \qquad n \text{ fixed counting number}$$

But this is exactly what the demand for finite prior probability, condition (4), requires.

8. References and further reading

1. Belot, G., (2013). "Bayesian Orgulity". *Philosophy of Science* 80 (4): pp.483-503.
2. Belot, G., (2017). "Objectivity and Bias". *Mind* 126(503): pp.655-695.
3. Bernoulli, J., (1713 [2006]). *The Art of Conjecturing.* Baltimore: The John Hopkins University Press.
4. Boole, G., (1854). *An Investigation of The Laws Of Thought, on Which Are Founded The Mathematical Theories Of Logic And Probabilities.* London: Walton – Maberly.
5. Burks, A.W., (1953). "Book Review: *The Continuum of Inductive Methods.* Rudolf Carnap." *Journal of Philosophy* 50 (24):731-734.
6. Carnap, R., (1950). *Logical Foundations of Probability.* London: Routledge and Kegan Paul, Ltd.
7. Carnap, R., (1952). *The Continuum of Inductive Methods.* Chicago: University of Chicago Press.
8. Carnap, R., (1963). "Replies and Systematic Expositions". In Schilpp, P.A., (ed.). *The Philosophy of Rudolf Carnap.* Library of Living Philosophers, Volume XI. Illinois: Open Court Publishing Company, pp.859-999.
9. Carnap, R., (1971). "A basic system of inductive logic, I". In Jeffrey, R., and Carnap, R., (eds.). *Studies in Inductive Logic and Probability.* Los Angeles: University of California Press. pp. 34-165.

10. Carnap, R., (1980). "A basic system of inductive logic, II". Jeffrey, R., (ed.). *Studies in Inductive Logic and Probability.* Berkeley: University of California Press. pp. 2-7.
11. Childers, T., (2013). *Philosophy and Probability.* Oxford: Oxford University Press.
12. Cisewski, J., Kadane, J. B., Schervish, M. J., Seidenfeld, T. and Stern, R., (2018). "Standards for Modest Bayesian Credences". *Philosophy of Science*, 85(1): pp.,53-78.
13. de Finetti, B., (1931)., "Probabilismo. Saggio critico sulla teoria delle probabilità e sul valore della scienza". In: *Logos.* Napoli: F. Pezzella, pp.163-219. English translation in *Erkenntnis* 31 (1989): pp.169-223.
14. de Finetti, B., (1936). "Statistica e Probabilita nella concezione di R. von Mises". *Supplemento Statistico ai Nuovi Problemi di Politica, Storia ed Economia Anno II, Fasc.2-3,* pp. 5-15.
15. de Finetti, B., (1972). *Probability and Induction. The art of guessing.* London: Wiley.
16. de Finetti, B., (1974). *Theory of Probability: A Critical Introductory Treatment.* Chichester: Wiley.
17. de Finetti, B., (2008). *Philosophical Lectures on Probability, collected edited and annotated by A. Mura.* Springer.
18. Earman, J., (1992). *Bayes or Bust: A critical examination of Bayesian Confirmation Theory.* Cambridge, Massachusetts - London, England: The MIT Press.
19. Elga, A., (2016). "Bayesian Humility". *Philosophy of Science*, 83: pp. 305–23.
20. Ellis, R.L., (1842). "On the Foundations of the Theory of Probability". In *The Mathematical and Other Writings of Robert Leslie Ellis,* 1862. Cambridge: Deghton Bell, and Co. pp. 1-11.
21. Gaifman, H., and Snir, M, (1982). "Probabilities Over Rich Languages, Testing and Randomness". *The Journal of Symbolic Logic*, *47(3),* pp. 495-548.
22. Gillies, D., (2000). *Philosophical Theories of Probability*. London and New York: Routledge.
23. Gnedenko, B.V., (1969 [1978]). *The Theory of Probability.* Moscow: Mir Publishers.
24. Goodman, N., (1955 [1981]). *Fact, Fiction and Forecast.* Cambridge, MA: Harvard University Press.
25. Hájek, A., (2019). "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>.
26. Hacking, I., (1971). "Equipossibility Theories of Probability". *The British Journal for the Philosophy of Science*, 22 (4), pp. 339-355.
27. Hacking, I., (1975 [2006]). *The Emergence of Probability: A philosophical study of early ideas about probability induction and statistical inference*. Cambridge: Cambridge University Press.
28. Hailperin, T., (1986). *Boole's Logic and Probability.* Amsterdam: North-Holland.
29. Hausdorff, F., (1914 [1957]). *Set theory.* New York: Chelsea Publishing Company.
30. Hempel, C.G., (1945). "Studies in the logic of confirmation, I". *Mind* 54 (213), pp. 1-26.
31. Hempel, C.G., (1945). "Studies in the logic of confirmation, II". *Mind* 54 (214), pp. 97-121.

32. Hesse, M., (1975). "Bayesian Methods and the Initial Probability of Theories". In, Maxwell, G. and Anderson, R.M., (eds). *Induction, Probability and Confirmation. Minnesota Studies in the Philosophy of Science, vol.6.* Minneapolis: University of Minnesota Press.

33. Hilbert, D., (1902). "Mathematical Problems". *Bull. Amer. Math. Soc*. **8** : pp. 437-479

34. Howson, C. and Urbach, P., (1989/2006). *Scientific Reasoning: The Bayesian Approach.* Chicago and La Salle, Illinois: Open Court.

35. Humphreys, P., (1985). "Why Propensities cannot be Probabilities". *The Philosophical Review* 94(4) pp.557-570.

36. Huttegger, S. M. (2015). "Bayesian Convergence to the Truth and the Metaphysics of Possible Worlds". *Philosophy of Science*, 82: pp. 587–601.

37. Huttegger, S. M. (2021). "Rethinking Convergence to the Truth". *The Journal of Philosophy* 119: pp. 380–403

38. Jeffrey, R., (1992). *Probability and the Art of Judgement.* Cambridge: Cambridge University Press.

39. Jeffrey, R., (2004). *Subjective Probability: The Real Thing.* Cambridge: Cambridge University Press.

40. Kolmogorov, A. N. (1933 [1950]). *Foundations of the Theory of Probability.* New York: Chelsea Publishing Company

41. Keynes, J. M., (1921). *A Treatise on Probability.* London: Macmillan and Co., Limited.

42. Lakatos, I., (1968). "Changes in the problem of inductive logic". In Lakatos, I., (ed.), *The Problem of Inductive Logic: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965, vol.2*. Amsterdam: North Holland Pub. Co. pp.315-417.

43. Laplace, P. S., (1814 [1951]). *A Philosophical Essay on Probabilities.* New York: Dover Publications, Inc.

44. Leibniz, G. W., (1678 [2004]). "On Estimating the Uncertain". *The Leibniz Review* 14.

45. Maher, P., (2006). "The Concept of Inductive Probability". *Erkenntnis* 65, pp.185–206.

46. Nielsen, M., (2021). "Convergence to Truth without Countable Additivity". *Journal of Philosophical Logic,* 50: pp. 395–414.

47. Nielsen, M. and Stewart, R.T., (2019). "Obligation, permission and Bayesian orgulity". *Ergo* 6(3).

48. Popper, K., (1959). "The Propensity Interpretation of Probability". *The British Journal for the Philosophy of Science*, 10, (37), pp. 25-42.

49. Psillos, S. and Stergiou, C. (2022). "The Problem of Induction". *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, https://iep.utm.edu/problem-of-induction/#H8

50. Ramsey, F. P., (1926). "Truth and Probability". In *The Foundations of Mathematics and other Logical Essays.* London and New York: Routledge (1931), pp. 156-198.

51. Reichenbach, H., (1934 [1949]). *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability.* Berkeley and Los Angeles: University of California Press.

52. Russell, B., (1948 [1992]). *Human Knowledge—Its Scope and Limits. London: Routledge.*

53. Salmon, W. C. (1966). *The Foundations of Scientific Inference.* Pittsburgh: University of Pittsburgh Press.
54. Savage, L. J. (1954 [1972]). *The Foundations of Statistics.* New York: Dover Publications. Inc.
55. Shackel, N. (2007), "Bertrand's Paradox and the Principle of Indifference, *Philosophy of Science*, 74 (2), pp. 150–175.
56. Venn, J., (1888). *The Logic of Chance.* London: Macmillan and Co
57. von Mises, R., (1928 [1981]). *Probability, Statistics and Truth.* New York: Dover Publications, Inc.
58. von Mises, R., (1964). *Mathematical Theory of Probability and Statistics.* London and New York: Academic Press.
59. Weatherson, B., (2015). "For Bayesians, Rational Modesty Requires Imprecision". *Ergo*, 2.

Author Information

Chrysovalantis Stergiou
E-mail: cstergiou@acg.edu
The American College of Greece – Deree
Greece

Stathis Psillos
E-mail: psillos@phs.uoa.gr
University of Athens
Greece