"A terrible piece of bad metaphysics"? Towards a history of abstraction in nineteenth- and early twentieth-century probability theory, mathematics and logic

Verburgt, L.M.

Publication date
2015

Document Version
Final published version

Citation for published version (APA):
1. Von Mises’s axiomatization of probability theory

The Austrian-born applied mathematician Richard Edler von Mises (1883-1953) published his ‘axiomatics’ for the ‘foundations of the calculus of probability as a mathematical discipline’\(^1\) in the second of two papers of 1919 entitled ‘Grundlagen der Wahrscheinlichkeitsrechnung’.\(^2\) In the other paper of that year, ‘Fundamentalsätze der Wahrscheinlichkeitsrechnung’,\(^3\) Von Mises had described the then-present situation in probability theory in the following words:

‘He who follows the development of the calculus of probability in the last decades cannot deny that this branch of the science of mathematics is behind all others in two respects. The analytical theorems [...] are lacking – except for few works by Russian mathematicians\(^4\) – [and] there is, in spite of some valuable beginnings, almost no clarity over the foundations of the calculus of probability as a mathematical discipline; this is all the more

---


\(^{2}\) For biographical accounts of Von Mises’ life and career see Kadıog˘lu & Erginöz, 2011; Siegmund-Schultze, 2004; Šišma, 2002; Vogt, 2007.

\(^{3}\) The ‘Fundamentalsätze’ paper ‘was concerned with the general theorem [...] for which, a year later, Georg Pólya was to propose the now well known name “the central limit theorem”’ (Cramer, 1953, p. 657).

\(^{4}\) Both in the ‘Fundamentalsätze’ paper and in an unpublished manuscript, entitled ‘Erwiderung auf die Bemerkungen des Herrn Pólya zu meiner Arbeit über die Fundamentalsätze der Wahrscheinlichkeitsrechnung’, submitted to the Mathematische Zeitschrift in November 1919 (Von Mises, 1919c), Von Mises referred to Chebyshev and Markov (but not to Liapunov).
surprising as we are living not only in an age of vivid interest in questions of axiomatics in mathematics but also in a period of increasing use of the calculus of probability in various fields of application'.

The main outlook of his ‘axiomatics’, which Von Mises would defend throughout the rest of his life, was that as the mathematical study of ‘mass phenomena’ or ‘repetitive events’ probability theory is concerned with their probabilities as the limits of relative frequencies in so-called ‘collectives’ (‘Kollektiv’) satisfying two ‘postulates’ (‘Forderungen’) or axioms – the existence of limits and irregularity or randomness:

‘Let (e) be an infinite sequence of objects of thought [gedachter Dinge], which we will briefly refer to as ‘elements’ e₁, e₂, e₃ ... To each element corresponds as a ‘label’ (‘Merkmal’) a system of values of the k real variables x₁, x₂ ... eₖ, i.e. a point of the k-dimensional [real] ‘label space’ (‘Merkmalraum’) and we assume that not all elements and not even all except finitely many elements have the same label. We call such [an infinite] sequence of elements ‘Kollektiv’ K, if the correspondence between the elements and the label satisfies the following postulates (axioms) I and II’:

‘[I] (L): the existence of the limit of the sequence of relative frequencies of the appearance of any subset of the label space (= probability of that subset), which results [i.e. is derived] from the sequence of events

6 For example, it can be found both in his Wahrscheinlichkeit, Statistik und Wahrheit of 1928, his Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und theoretischen Physik of 1931 and his posthumously published Mathematical Theory of Probability and Statistics of 1964.
7 This means that, according to Von Mises, it is possible to speak about probabilities only in reference to a properly defined collective – or, as his famous (infamous?) slogan goes, ‘First the collective – then the probability’.
8 In other words, ‘[t]here must be at least two labels, to both of which an infinity of elements is ordered’ (Von Plato, 1994, p. 183).
[...], 10 [II] (IR): the *irregularity* or *randomness* of the collective secured by stipulating that the limit (the probability as defined in (L)) will not change as a result of such [place]-selections of sub-sequences which [can] be defined without regard to the outcome (label) of the [event].

For Von Mises, a ‘collective’ represented an infinite sequence of events with values (labels) in some space $M$ such that it holds that (L) each subset $A$ of $M$ appears with a certain limiting (asymptotic) relative frequency in both the sequence itself as well as in each sub- or partial-sequence formed (IR) without knowledge of its future values. Thus, ‘each collective induces a probability distribution' $W$ in the space $M$ [and] this distribution is invariant under the selection of [...] sub-sequences' by some ‘regular’ mathematical law.

---

10 Siegmund-Schultze notes that Von Mises ‘is talking about two different kinds of sequences here, the first being a (derived) sequence of numbers (relative frequencies), the second the ‘collective’ itself’ (Siegmund-Schultze, 2004, p. 356).

11 In other words, if the relative frequencies of the attributes must have limiting values (i.e. must tend to fixed limits), these limiting values must be such that they remain invariant under the (place) selection of sub-sequences from a given original sequence (i.e. ‘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’ (Von Mises, 1928 [1981], p. 29). In modern terminology, a place selection is ‘a procedure for selecting a subsequence of the given sequence $x$ in such a way that the decision to select a term $x_n$ does not depend on the value of $x_n$’ (Van Lambalgen, 1987, p. 725). Von Mises referred to this (IR) axiom in terms of the ‘Unmöglichkeit eines Spielsystems’ (‘impossibility of a gambling system’) (Von Mises, 1919b, p. 58).

12 Siegmund-Schultze, 2006, p. 441.

13 Von Mises’s definition of a ‘probability distribution’ was the following: ‘If, for instance, six players bet, each on one of the six different sides of a die, the chances are ‘distributed’ in such a way that the relative chance of each of the players is equal to the probability of the side which he has chosen. If the die is an unbiased one, all the chances are equal; they are then uniformity ‘distributed’’ (Von Mises, 1928 [1981], p. 35).


15 The core of Von Mises’s probability theory consisted of mathematically studying how new collectives with new distributions are produced from given collectives with certain given distributions by means of the combination of four ‘fundamental operations’ (‘selection’, ‘mixing’, ‘partition’ and ‘combination’) in ways practically relevant for games of chance, statistics and theoretical physics (e.g. Von Mises, 1928 [1981], p. 38). Or, as Karl Popper succinctly put it in his *Logic of Scientific Discovery*: ‘The task of the calculus of probability consists, according to von Mises, simply and solely in this: [...] to calculate probabilities which are not given from probabilities which are given’ (Popper, 1935 [2004], p. 141).
1.1 The non-mathematical foundations of Von Mises’s mathematical probability theory

1.1.1 Von Mises’s 1912 paper

Von Mises’s ‘collective’ first appeared in a non-axiomatic way in a paper of 1912 on Gustav Theodor Fechner’s (1801-1887) and Heinrich Bruns’ (1848-1919) ‘Kollektivmasslehre’, the theory studying the frequency distributions of the values of the attributes of an unknown number of randomly varying ‘individual specimens’ that together form a ‘collective object’ (‘Kollektivgegenstand’) or ‘collective’ (‘Kollektiv’) in so far as they are held together by a ‘kind’, ‘species’ or ‘genus’. Von Mises, whose teacher at the Technische Hochschule (1901-1906) Emanuel Czuber (1851-1925) had developed Kollektivmasslehre from a Laplacian perspective (!) in the first volume of his *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung* of 1908, followed the Machian Fechner and Bruns in several respects. Firstly, the theory of measuring ‘collectives’ is considered as an ‘empirical theory of chance alongside science based on causal laws’ that is equal to or nothing other than probability theory, with both probabilities and ‘collective objects’ being applicable to mass phenomena. Secondly, it differs from these other sciences because its laws are ‘laws of chance’ concerned with distributions rather than individual cases – that is, if these laws can determine the frequency with which the values of individual members (‘specimens’) of a collective are distributed,

---

16 In the year 1912, Von Mises worked as Associate Professor of applied mathematics in Strasbourg.
17 See Von Mises, 1912.
18 See Bruns, 1898; Bruns, 1906; Fechner, 1897. For accounts of the ‘Kollektivmasslehre’ see, for example, Heidelberger, 2004, chapter 8; Sheynin, 2004.
19 A similar definition of a ‘collective’, albeit without the condition of random variation, in terms of ‘individuals’ of a ‘species’ was given not only by, for example, Gustav Rümelin (1815-1889) in 1863 (Rümelin, 1863), but also by the English logician John Venn (1834-1923) in his famous *Logic of Chance* of 1866. Stöltzner has also drawn attention to the fact that the Rector of the University of Vienna Franz-Serafin Exner (1849-1926) defended the ‘Kollektivmasslehre’ as early as in the year 1908 (Stöltzner, 2003, pp. 214-215).
20 Czuber, 1908. See also below (‘An afterword’).
21 See Heidelberger, 2004, section 8.5 for a discussion.
they cannot determine ‘how large this or that individual specimen is’. Although they were not in full agreement about the import of ‘chance’, Fechner and Bruns both made a fundamental distinction between viewing probability laws as ‘descriptions of nature or as a methodical means for avoiding [observational] error’. Thus, in so far as the theory was presented, thirdly, as a generalizing transformation of the Gaussian law of normally (symmetrically) distributed errors of observation – with the errors being deviations from a single real value of a physical quantity into a law of non-normal (asymmetrical) distributions of ‘objects’ as the ‘real [and] true’ deviations from a purely mathematical

---

23 Fechner 1897, p. 6.
24 The difference in opinion will be elaborated below, namely in the context of the second part of the fourth respect in which Von Mises followed Fechner and Bruns.
26 See, for example, Fechner, 1897, p. vi; Von Mises, 1912, p. 9. Von Mises spoke of ‘einer weitergehenden Analyse der Ergebnisse wiederholter Beobachtungen’ (‘a more elaborate analysis of the results of repeated observations’) (Von Mises, 1912, p. 9). Fechner, Bruns and Von Mises followed up on the generalization of the Gaussian error law in the work of, among others, Friedrich Wilhelm Bessel (1784-1846) and Gotthilf Hagen (1798-1884) (see Fischer, 2011, section 3.2).
27 As Fischer explains, ‘[i]n the framework of classical probability theory, the primary objective was to calculate probabilities of certain events, with the aim of making “rational” decisions based on these probabilities. Error or frequency functions only played the role of auxiliary subjects. This paradigm, however, would change fundamentally during the course of the 19th century […] In this context it was the prevailing opinion for a long time that almost all quantities in nature obeyed normal distributions. For a justification of the apparently privileged role of normal distribution, a model was used in most cases which had originally originated from error theory: the hypothesis of elementary errors. A random quantity obeying this hypothesis was assumed to be additively composed of a very large number of independent elements, each of them being insignificant compared with the total sum […] In error calculus, the main object was to give an “optimal” estimate of the true value of a physical quantity and to minimize random deviations from this value as far as possible’ (Fischer, 2011, p. 75, p. 108).
28 Fechner, 1897, p. 16.
mean, it has to be established as a new discipline.\textsuperscript{29,30} Given that these objects also had to be created, the theory was premised, fourthly, on the explication of two ‘axioms’ for ‘collectives’.\textsuperscript{21} On the one hand, in order to mathematically represent frequency distributions it must be assumed that collectives are, in some sense, infinitely large – which, in the case of Fechner\textsuperscript{32} meant that collectives consist of an infinitely large amount of attributes and, in that of Bruns,\textsuperscript{33} that collectives, as arithmetical counterparts of density curves,\textsuperscript{34} are described by an infinite series\textsuperscript{35} with successive derivations as terms. Von Mises, in showing the convergence (or ‘development by successive approximations’)\textsuperscript{36} to the normal distribution of the series that Bruns had introduced in response to Fechner’s problem of finding distributing laws for the ‘deviations from the arithmetical mean exhibited by the \textit{many} specimens of a collective object’,\textsuperscript{37}

\textsuperscript{29} See Heidelberger, 2004, p. 303.
\textsuperscript{30} For example, Fechner wrote that if the distributions of (the randomly varying ‘specimens’ of) collectives are governed by ‘the general laws of random probability [...] when determining dimensions in astronomy the same laws of probability are only used \textit{secondarily} for defining the certainty of the calculated average measurements, and thus play a different and much less significant role than in the theory of measuring collective objects [for] the specimens of a collective object, no matter how much they deviate from an arithmetical average or any other main value [are] equally real and true, and preferring one above another for a reason that is insignificant to them all naturally makes no sense’ (Fechner, 1897, p. 16). Put differently, Fechner observed that ‘most frequency functions encountered outside the physical sciences [e.g. in lotteries, anthropology, botany, and meteorology] are asymmetric and he aims at supplementing the classical error theory taking this fact into account’ (Hald, 2002, p. 24).
\textsuperscript{31} Or, as Heidelberger puts it, ‘[t]he first important step [for the theory] consists of examining whether or not a given compilation of data is actually a collective object’ (Heidelberger, 2004, p. 303).
\textsuperscript{32} Fechner ‘used the error-theoretic term ‘true value’ of the constant sought’ which he equated ‘to the limit of the arithmetic mean as the number of observations [of ‘attributes’] increases indefinitely’ (Sheynin, 2004, p. 57).
\textsuperscript{33} Bruns, who thought that ‘a collective was an arithmetical counterpart of a density curve [or ‘function’]’ (Sheynin, 2004, p. 69), held that the application of probability theory ‘presupposes that objects exist that at least approximately realize the concepts of random events and theoretical frequency distributions’ (Hald, 2002, p. 25).
\textsuperscript{34} See Bruns, 1898, pp. 342-343;
\textsuperscript{35} Von Mises, in the section of coefficients of the ‘Bruns-series’, wrote of ‘formally writing down an infinite series’ (‘Wir schreiben [...] formal die unendliche Reihe [...]’) (Von Mises, 1912, p. 12).
\textsuperscript{36} Von Mises wrote of a ‘durch sukzessive Approximationen gebildete Entwicklung nach “Normal-funktionen”’ (Von Mises, 1912, p. 16).
\textsuperscript{37} Fechner, 1897, p. 64.
assumed the ‘ideal-typical’ (‘idealen Typus’) existence of an ideal distribution (‘ideale Verteilung’) of values ‘with which not only the immediately given real distribution, but (in a certain, yet to be further specified way)* [also] the deviation between reality (‘der Wirklichkeit’) and the ideal-case (‘dem Idealfall’) can be compared’. On the other hand, in order to constitute a collective a set of randomly varying observational data must exhibit a ‘random variation’. Fechner wrote that this condition is satisfied when it can be demonstrated that the individual elements (or ‘specimens’) ‘are not related to one another by natural law dependency’ but by ‘pure randomness’ or ‘chance’. Although he developed several tests for discovering the influence of natural laws on the individual specimens, because he was aware of the philosophical difficulties involved in his understanding of chance as a an indeterminate objective phenomenon existing in the real world Fechner settled for a negative definition of the criterion of random variation. Bruns, who thought that ‘it is absolutely unnecessary to consider chance when calculating probabilities’, allowed for the natural-scientific ‘abstraction’ of the ‘apparently random events of reality [from] the causal relations which actually prevail, even though we may not understand them’.

---

38 ‘Wir gehen von der Vorstellung aus, dass es irgendeine ideale Verteilung [...] gibt, mit der nicht nur die unmittelbar gegebene wirkliche Verteilung [...] sondern [...] auch die Abweichung zwischen der Wirklichkeit und dem Idealfall zu vergleichen ist’ (Von Mises, 1912, pp. 9-10).
40 The basic idea of these tests was the following: ‘Take a series of data for which we know that they “vary by chance” [...] meaning that it is irregular (or at random). Using this random series, then compare the group of data to be investigated and measure the degree of deviation and thus also the degree of dependency or independence among individual specimens’ (Heidelberger, 2004, p. 304).
41 For an account of Fechner’s views on indeterminism see Heidelberger, 1987; Heidelberger, 2004, section 8.1
42 Fechner believed that in so far as the condition of randomness cannot be precisely defined ‘[w]e must be content to name the factual point of more negative than positive a nature that underlies what is to follow. By random variation of specimens I mean a variation that is just as independent of any arbitrariness in determining dimensions as it is from a natural law that governs the relations holding between dimensions. Although the one or the other may partake in defining the objects, the only real random changes are those that are independent of them’ (Fechner, 1897, p. 6). Sheynin has drawn attention to the fact that Fechner ‘believed that an attempt to consider randomness from a philosophical standpoint would bear little fruit, remarking that the random variation of the [collective] objects was neither arbitrary nor regular’ (Sheynin, 2004, p. 63).
43 Bruns, 1906, p. 7.
Von Mises, for his part, did not mention the condition of random variation for collectives in his 1912 paper – and in light of this fact it may be observed that by that time Von Mises approached probability theory from the mathematical side, leaving the question of the direct empirical support of a condition like this to ‘theories of a physical nature’.

Where Bruns wrote of the axioms of probability calculation as ‘abstractions’ whose approximate validity for reality had somehow to be decided empirically, Von Mises aimed to introduce ‘Bruns-series’ into the measuring theory of collectives ‘in a decidedly self-evident way, instead of apodictically claiming their applicability’. He, thus, had to specify what justifies the assumption that an ‘ideal distribution’ (‘ideale Verteilung’) (with its ‘Gaussian function’ (‘Gaussche Funktion’)) is comparable, in the sense of statistics, with a ‘real distribution’ (‘wirkliche Verteilung’) (the continuous function \( f(x) \)) and with the deviation between reality and the ideal-case. For this purpose, the constants \( c_v \) of the infinite series, \( c_1 \phi_1 + c_2 \phi_2 + c_3 \phi_3 + \ldots = f(x) \), that is uniformly convergent in the entire interval, are said to mean (‘bedeuten’) ‘moments’ (‘Momente’) of the function \( f(x) \).

Von Mises did not explicitly refer either to the analytic theory of moments of Pafnuty Chebyshev (1821-1894), A.A. Markov (1856-1922) and Thomas Johannes Stieltjes (1856-1894) or to the theory of empirical moments of Francis Ysidro Edgeworth (1845-1926) and Karl Pearson (1857-1936).

---

45 ‘[A]uf eine gewisse natürliche Weise enzuführen, statt ihre Anwendung einfach apodiktisch hinzustellen’ (Von Mises, 1912, p. 9).
46 This is Von Mises’s notion.
47 Von Mises also, and repeatedly, made use of ‘moments’ in his Fundamentalsätze paper of 1919. The important point is that in the 1912 paper ‘moments’ were introduced in order to justify one of the ‘axioms’ of measuring theory of collectives.
48 For an account of the history of the analytical theory of moments and the quasi-analytical theory of empirical moments see, for example, Fischer, 2011, section 3.4.2.3, section 4.3-4.6.
Be that as it may, the mechanics-inspired idea of quantitative ‘moments’ was here put forward to account for the convergence in normal distribution of the series expansions by means of justifying it in terms of a sequence of moments with a one-to-one correspondence, so to say, to ‘theoretical’ moments (‘constants’).

1.1.2 Von Mises’s 1919 papers

After having worked as the mechanics assistant of the professor of mathematics at the Technical University of Brünn (now Brno) Georg Hamel (1877-1954) between 1906-1909, Von Mises, who participated in the short-lived ‘First Vienna Circle’, was appointed ‘extraordinary’ professor of applied mathematics at the University of Strassburg (Strasbourg) – a position which he officially held until 1918, the year in which the university was returned to France. For a few months in the year 1919, Von Mises occupied the new chair of hydro- and aerodynamics at Dresden Technical College, only to hold the new chair of applied mathematics and to arrange for becoming the director of the new institute of Applied Mathematics at the University of Berlin in 1920. Together with his colleagues Issai Schur (1875-1941) and Ludwig Bieberbach (1886-1982), Von Mises managed to turn Berlin, with its tradition of mathematicians of the stature of Karl Weierstrass (1815-1897) and Ernst Eduard Kummer (1810-

---

49 If this generally acknowledged in so far as ‘moments’ represent mass in mechanics, Von Mises himself compared ‘moments’ to coefficients in the case of spherical harmonics (‘Kugelfunktion’) (see Von Mises, 1912, pp. 12-13). Reflecting on Von Mises’s 1919 papers, Siegmund-Schultze writes not only that ‘Von Mises as a theoretical mechanist was [...] prepared to take over Stieltjes’s integrals into probability, as it had close connections with moment problems in mechanics’ (Siegmund-Schultze, 2010, p. 212, f. 26), but also that for his proof of the central limit theorem Von Mises did not make use of ‘continued fractions which had had a central place in the [analytic!] Chebyshev-Markov “moment methods”’ (Siegmund-Schultze, 2006, p. 456). Von Mises himself would later refer to his own methods in his early papers as being ‘roundabout’.

50 Von Mises explained that if the constants cv ‘mean’ (‘bedeuten’) moments $M_v$ of the function $f(x)$, conversely, the moments $M_v$ can be put forward by means of ‘durch’ the constants cv (see Von Mises, 1912, p. 12).

51 In the pre-war years, Philipp Frank (1884-1966), Hans Hahn (1879-1934), Otto Neurath (1882-1945) and Von Mises regularly met in a Vienna coffee house to discuss the philosophy and scientific method of Mach and the French conventionalists, Pierre Duhem (1861-1916) and Henri Poincaré (1854-1912). See Haller, 1985.
1893), into a world-renowned center of mathematics that could compete, institutionally and theoretically, with David Hilbert’s (1862-1943) and Felix Klein’s (1844-1925) Göttingen. Von Mises himself attributed the success of both Göttingen and Berlin to the effort of bridging the gap between ‘pure’ and ‘applied’ mathematics – a point which he made explicit in his contribution to the first issue of the journal that he himself established in 1921, the Zeitschrift für angewandte Mathematik und Mechanik (ZAMM).

It was shortly after the First World War, during which Von Mises had worked as a planner, designer, test pilot and lecturer in the Austro-Hungarian air force, that probability theory ‘began to be discovered as a field for ambitious analysts, even outside Russia’. Von Mises was one of the first to emphasize the need for the clarification of the analytical foundations of the central limit theorem and its classic special cases, so to say, by means of transforming them from a tool for (‘error-theoretical’) probability distributions in, for example, astron-

---

52 See, for example, Biermann, 1988; Rowe, 1998; Vogt, 1982 on the history of Berlin mathematics.
53 See, for example, Bernhardt, 1980 on the development of Berlin mathematics during the 1920s.
54 See Von Mises, 1921a; Von Mises, 1921c.
55 Von Mises founded the ZAMM in 1921 and ‘until 1933, he was not only an editor of this journal, but also one of its most active contributors’ (Šišma, 2002, 180).
57 Von Mises referred to ‘das Bernoullische und Poissonsche Problem und seine Verallgemeinerung, das Gaussische Fehlergesetz, das Gesetz der grossen Zahl […] das Bayessche und Laplace-Bienaymésc Problem und seine Verallgemeinerung, einen zweiten Satz der Fehlertheorie und ein zweites Gesetz der grossen Zahl’ and described the goal of the first of his two 1919 papers as that of settling ‘a large amount of individual problems of probability theory […] from a unified and general analytic point of view’ (‘eine grosse Reihe von Enzelproblemen der Wahrscheinlichkeitsrechnung […] von einem einheitlichen und allgemeinen analytischen Gesichtspunkt aus zu bewaltigen’) (Von Mises, 1919a, p. 2).
omy and economics, into objects within mathematics itself. And his lengthy *Fundamentalsätze* paper, submitted on 31 August 1918 and published in 1919, was indeed prompted exactly by the lagging behind of probability theory in respect of ‘precision of formulation and proof that has become self-evident for long in other parts of analysis’.

Von Mises formulated limit theorems for distributions (*Verteilungen*) $Vx$ here introduced as so-called monotonically increasing (real) functions with value (limit) 0 when $x = -$ and value (limit) 1 when $x = +$ for certain linear combinations of independent random variables (*linearen Faltungen von Kollektives*).

Referring to these distributions, Von Mises explained the use of the then relatively new analytic tool of the so-called ‘Stieltjes integral’ for the representation of probabilities by writing that it was analytically ‘superior to the usual

---

58 Given him being a ‘technician [whose] mathematical competence [was] recognized by all his colleagues in the field’ (Hamel 1912 quoted in Šišma, 2002, p. 188), this priority of ‘discovery’ may come as no surprise. For example, Vogt characterizes Von Mises as follows: Vogt characterizes Von Mises: ‘Richard von Mises was not only qualified as an excellent mathematician [but] [h]e had also enormous experience in solving practical problems, problems of mechanics and fluid dynamics, engineering problems, and last but not least, problems of aerodynamics and the construction of airplanes’ (Vogt, 2007, p. 11).

59 Hereby, Von Mises established as a fait accompli that ‘the study of limit theorems for probability distributions was the only aspect of the emerging [‘modern’] form of probability theory that was linked in any significant way to the results produced in the 19th century’ (Fischer, 2011, p. 208).

60 This fact seems to have been important for the following reason: ‘The start of [Von Mises’s] more detailed studies of probability theory […] probably also owed something to the fact that this was an area that was relatively poorly researched from the mathematical point of view, so that he saw it as offering opportunities to distinguish himself – as he was afraid of losing his professorship in Strassburg toward the end of the First World War’ (Fischer, 2011, p. 209).

61 ‘Präzision der Formulierung und Beweisführung, die in anderen Teilen der Analysis längst zur Selbstverständlichkeit geworden ist’ (Von Mises, 1919a, p. 1).

62 These are functions defined on a subset of the real numbers that preserve the given order between two ordered sets in such a way that if for all $x$ and $y$ such that $x \leq y$ it holds that $f(x) \leq f(y)$.

63 See Von Mises, 1919a, p. 76.

64 For technical discussions of the core of Von Mises’s *Fundamentalsätze* paper see, for instance, Fischer, 2011, section 5.2.2; Siegmund-Schultze, 2006.
one in the calculus of probability’,\textsuperscript{65} namely the mean of Bruns-series or the mathematical expectation of Markov.\textsuperscript{66} His ‘rigorous and general presentation of the analytic aspects of basic probabilistic principles’\textsuperscript{67} in terms of the then up-to-date standards and vocabulary of analysis seems to have been suggestive not only of the paper’s purely analytical aim, but also of Von Mises’s (instrumental)\textsuperscript{68} ‘tendency towards [...] mathematics (instead of merely mechanics).’\textsuperscript{69} around the late 1910s. But Von Mises remained an ‘applied mathematician’, and he also wanted his results on the integral central limit theorem to contribute to the establishment of new axiomatic (albeit non-mathematical and application-oriented) foundations for probability theory.

The eclecticism of Von Mises’s own attitude as an ‘applied’ or ‘practical’\textsuperscript{70} mathematician was reflected in the axioms ((L) and (IR))\textsuperscript{71} for collectives found in the \textit{Grundlagen} paper of 1919 – a foundational system for which Bruns,

\textsuperscript{65} ‘The notation [...] which is introduced here is superior to the usual one in the calculus of probability such as $D [\varphi(x)]$ (Mean ['Durchschnitt'] of $\varphi$, Bruns) or $m \cdot H \cdot \varphi(x)$ (mathemat. expectation of $\varphi$, Markov) because it shows explicitly the distribution, related to which the mean is taken. Also as to simplicity it should not be wanting and, above all, it is in better agreement with the notation elsewhere in analysis’ (‘Die hier eingeführte Bezeichnungsweise [...] ist den in der Wahrscheinlichkeitsrechnung gebräuchlichen, wie $D [\varphi(x)]$ (Durchschnitt von $\varphi$, Bruns) oder $m \cdot H \cdot \varphi(x)$ (mathemat. Hoffnung (mathemat. Hoffnung von $\varphi$, Markoff) dadurch überlegen, dass die Verteilung bezüglich deren der Durchschnitt gebildet wird, ausdrücklich angeführt erscheint. An Einfachheit dürfte sie auch nicht nachstehen und vor allem sich der sonstigen Schreibweise der Analysis besser anpassen’) (Von Mises, 1919a, p. 20, f. 3, translated into English in Siegmund-Schultze, 2006, p. 440).

\textsuperscript{66} Fischer notes that Von Mises ‘made explicit reference to [...] the work of Chebyshev (as described by the Chebyshev biographer Vasilev [...]), and to Markov’s papers (as contained in the appendix of the German translation of the latter’s ‘Probability Theory’). However, Von Mises’s knowledge of the work of Chebyshev and Markov was merely superficial, as evidenced [for example] by the fact that he cited it only partially’ (Fischer, 2011, p. 212).

\textsuperscript{67} Fischer, 2011, p. 217.

\textsuperscript{68} See footnote 57.

\textsuperscript{69} Siegmund-Schultze, 2006, p. 435.

\textsuperscript{70} Von Mises apparently once called the author of one of the major textbooks of modern set-theory Felix Hausdorff (1868-1942) an ‘unpractical mathematician’ (see Siegmund-Schultze, 2006, p. 442, f. 35).

\textsuperscript{71} These axioms were given in section 1.
Antoine-Augustin Cournot (1801–1877)\textsuperscript{72} and Jean-Marie Le Roux (1863–1949)\textsuperscript{73} are mentioned as ‘transitional’\textsuperscript{74} predecessors. On the one hand, Von Mises ‘apparently felt obliged, due to the general conditions of [mathematical] research in the post-Weierstrassian era, to strive for the utmost analytic rigor and generality without considering any aspects outside of mathematics’.\textsuperscript{75} Thus, probability theory is described as a ‘branch of mathematical science’ (‘Zweig der mathematischen Wissenschaft’), or ‘mathematical discipline’ (‘mathematischen Disziplin’), whose foundations have, surprisingly enough,\textsuperscript{76} not been axiomatized, whose analytic theorems lack the rigor of other parts of analysis and whose object of study and fundamental problem-situation (‘grundsätzliche Problemstellung’)\textsuperscript{77} are still ‘shrouded in darkness’ (‘in tiefel Dunkel gehüllt’).\textsuperscript{78} Taken together, albeit all the while somewhat rephrasing the theoretical situation, if it had to be acknowledged that the theory was not yet a mathematical discipline,\textsuperscript{79} Von Mises attempted at its construction (‘Aufbau’)\textsuperscript{80} by means of an axiomatic system (‘Axiomen-system’)\textsuperscript{81} with (L) and (IR) as the axioms for a mathematical version (or ‘mathematische Präzisierung’) of the ‘Fechner-Brunssian “collective objects”’ (‘Fechner-Brunsschen “Kollektiv-Gegenstandes”’),\textsuperscript{82} namely

\begin{itemize}
\item \textsuperscript{72} Von Mises referred to Cournot’s \textit{Exposition de la théorie des chances et des probabilités} of 1843 – the ‘last book of moral science in the style of the Enlightenment’ (Hacking, 1990, 96).
\item \textsuperscript{73} Von Mises referred to Le Roux’s ‘Calcul des probabilités’ of 1906, a two-page article explaining Czuber’s work.
\item \textsuperscript{74} It is not entirely correct, as, among others, Von Plato suggests (Von Plato, 1994, p. 183), that these figures were predecessors of Von Mises’s own foundational system. Rather, their work formed a bridge between (in the sense of ‘Überleitung zu’ (Von Mises, 1919b, p. 53)) the classical definition of probability in terms of ‘equally likely cases’ and the new frequency definition as found in Von Mises’s 1919 papers and 1928 textbook.
\item \textsuperscript{75} Fischer, 2011, p. 217.
\item \textsuperscript{76} Von Mises was startled (he wrote ‘was um so erstaunlicher erscheint’ (Von Mises, 1919a, p. 1)) about the fact that there was no clarity over the foundations of the calculus of probability in a time of such a lively interest for axiomatic questions within mathematics (‘in einer Zeit lebhaften Interesses für axiomatische Fragen innerhalb der Mathematik’ (Von Mises, 1919a, p. 1))
\item \textsuperscript{77} Von Mises, 1919b, p. 53.
\item \textsuperscript{78} Von Mises, 1919b, p. 53.
\item \textsuperscript{79} Von Mises wrote that ‘man den gegenwärtigen Zustand kaum anders als dahin kennzeichnen [kann], dass die Wahrscheinlichkeitsrechnung heute \textit{eine mathematische Disziplin nicht ist}’ (Von Mises, 1919b, p. 52, emphasis in original).
\item \textsuperscript{80} Von Mises, 1919b, p. 53.
\item \textsuperscript{81} Von Mises, 1919b, p. 54.
\item \textsuperscript{82} Von Mises, 1919b, p. 54.
\end{itemize}
the ‘collective’. On the other hand, because probability theory is ‘a natural science, in the same way as geometry or theoretical mechanics’, the ‘axiomatics’ was not a formal system, but a result of an idealized abstraction from the external world (‘Außenwelt’) that went ‘beyond’ the formal realm. Where the attempts at mathematically establishing probability of Émile Borel (1871-1956), Georg Bohlmann (1869-1928) and Ugo Broggi (1880-1965) remained formalizations, Von Mises’s own exposition (‘Darstellung’) assumed that

‘[the] goal [of the theory] is to ‘re-present’ (‘wiederzugeben’) the inter-relations (‘Zusammenhänge’) and dependencies of well-determined (‘bestimmter’), observable phenomena, not as a truthful image (‘getreues Abbild’) of the external world, but as its abstraction and idealization. It must start from logically unambiguous concept-constructions (‘Begriffskonstruktionen’), for whose construction it is decisive to take into consideration the fact [represented] (‘darzustellende Objekt’) in the external world. [W]hen the construction has succeeded by means of a system of axioms, then the results are obtained in purely deductive ways, the practical usefulness of the theory being determined by their applicability to the objects of experience (‘Gegenstände der Erfahrung’).’

Here, Von Mises followed Hilbert not only in mentioning probability theory alongside geometry and theoretical mechanics and in characterizing it as a...

---

83 ‘Eine Naturwissenschaft gleicher Art wie Geometrie oder die theoretische Mechanik’ (Von Mises, 1919b, p. 53).
84 Von Mises criticized the attempts at mathematically establishing probability in the work of Borel, Georg Bohlmann (1869-1928) and Ugo Broggi (1880-1965) for not having moved beyond, in some sense, the realm of the formal: ‘Die bisher unternommen Versuche […] (Bohlmann, Broggi, Borel) scheinen mir durchaus im Formalen stecken geblieben zu sein’ (Von Mises, 1912, p. 53, my emphasis).
85 Von Mises referred to Borel, 1909; Bohlmann, 1901; Broggi, 1907.
86 ‘Sie hat das Ziel, die Zusammenhänge und Abhängigkeiten bestimmt, beobachtbarer Erscheinungen wiederzugeben, nicht als getreues Abbild der Außenwelt, sondern als deren Abstraktion und Idealisierung. Sie muss von logisch eindeutigen Begriffskonstruktionen ausgehen, bei deren Aufbau die Rücksicht auf das darzustellende Objekt der Außenwelt maßgebend ist; ist aber einmal der Aufbau durch ein Axiomensystem erfolgt, so werden auf rein deduktivem Wege die Ergebnisse gewonnen, deren Anwendbarkeit auf die Gegenstände der Erfahrung allein über die praktische Brauchbarkeit der Theorie entscheidet’ (Von Mises, 1919b, pp. 53-54). Author’s translation.
natural or physical science, but also in his empiricist distinction between the ('genetic') origin and ('logical') use of the axioms of a system. At the same time, Von Mises was in disagreement with Hilbert and his fellow Göttingen mathematicians Bohlmann and Broggi as far as the status of this origin and use was concerned. He dismissed the idea of 'probabilities' as the primitive observable (or 'intuitive') phenomena to be expressed by a collection of concepts to which the inner-mathematical theorems derived from the axioms could later be traced back by means of interpretation. The reason for which was that the ontological basis for the foundational law that grounds probability theory, namely that of large numbers, is such that this average law is only realized with exactness in the limit of an unobservable infinity of occurrences of events. Because an observable sequence of events (or 'empirical collective' ("empirischen" Kollektivs)) about which experience teaches that it is characterized by two features (stability of relative frequency and a lack of order), called the 'Urphänomen', is necessarily finite, in order to mathematically calculate real world probabilities the sequence must first be carefully transformed into (or re-presented in the form of) a calculable infinite sequence ('mathematical collective') with the two laws as axioms ((L) and (IR)). These two

87 Hilbert’s ‘sixth problem’ of his famous 1900 lecture proposed that the ‘[t]he investigations on the foundations of geometry’ suggest the problem of treating ‘in the same manner, by means of axioms [...] the theory of probabilities and mechanics’ (Hilbert, 1900 [2000], p. 418).

88 Corry has shown that the ‘arch-formalist’ Hilbert actually held an empiricism position vis-à-vis the axiomatic foundation of the physical sciences in general and, for example, geometry in specific. See, for example, Corry, 2004, section 3.1.

89 Von Mises, 1919b, p. 60.

90 In Von Mises’s own words: ‘Die Erfahrung lehrt erstens, dass, [...] die relative Häufigkeit jedes [...] Merkmale nähernd konstant bleibt [...] Die zweite Erfahrung [...] besteht darin, dass die annähernde Konstanz der relative Häufigkeiten bei genügend großer Beobachtungsreihe fortbesteht, und zwar mit unveränderten Werten der Konstanten [...] Aber nicht nur dies, man weiß auch, dass das Häufigkeitsverhältnis zweier Merkmale bei einer solchen Auswahl unverändert bleibt [...] Mit anderen Worten heißt das: die Chancen zweien Spieler, die etwa auf die Zahlen 1 und 6 setzen, ändern sich nicht, wenn sie, ohne die Wurfergebnisse vorher zu kennen, aus der Gesamtheit aller Würfe irgendeine – genügend große – Gruppe von Würfen herausgreifen’ (Von Mises, 1919b, p. 61).

91 Stöltzner notes that ‘[s]till in those days [around 1919] empirical investigations into such simple phenomena [such as lotteries, birth rates, etc.] were very common; Frank, for instance, reports in detail a statistical investigation of the number of pedestrians within a small strip of the wide walk’ (Stöltzner, 2003, p. 217).
axioms for ‘mathematical collectives’, thus, arise (‘entsprechen’)\(^{92}\) and are valid (‘gelten’)\(^{93}\) as the exact idealizations, rationalizations or schematizations of the two laws governing certain ‘well-defined observable appearances’ (‘ganz bestimmte beobachtbare Erscheinungen’),\(^{94}\) namely mass phenomena. The relation between (‘Zusammenhang’)\(^{95}\) these two collectives is similar to that between geometrical concepts and the images of empirical space arising from them:\(^{96}\) for Von Mises, the abstract representations of the natural sciences exist as imperfect instruments reconstructing the observable phenomena in empirical reality as completely and as simply as possible. Von Mises distinguished the two collectives as two parts or stages of a single scientific activity in which statements about observations (‘empirical collectives’) are turned into an axiomatic system (‘mathematical collectives’) from which all known laws of probability theory can be mathematically derived and whose applicability to phenomena is somehow secured by the sake of the idea that the axioms are ‘scientific’ only in so far as they are not merely tautological but provide useful descriptions of phenomena.\(^{97}\) Von Mises dismissed the problem of the coordination of observations of mass phenomena to the axioms and referred to a ‘double transition’ (‘zweimaliger Übergang’)\(^{98}\) between ‘theory and reality’ (‘Theorie und Wirklichkeit’)\(^{99}\) – in brief, that of specifying probability distributions and providing pre-dictions or explanations of observable phenomena on the basis of the results of calculation (‘Rechnungsergebnissen’)\(^{100}\) – the actual burden of which is said to fall outside the ‘”sphere of thought”’ (‘Gedankenkreise’))\(^{101}\) of probability theory.

\(^{92}\) Von Mises, 1919b, p. 60.
\(^{93}\) Von Mises, 1919b, p. 60.
\(^{94}\) Von Mises, 1919b, p. 60.
\(^{95}\) Von Mises, 1919b, p. 60.
\(^{96}\) As Von Mises wrote, ‘Der Zusammenhang zwischen unsren Definition und der Wirklichkeit ist derselbe wie der zwischen geometrischen Begriffen wie Körper, Fläche, Linie usw. und den ihnen entsprechenden Gebilden des empirischen Raumes’ (Von Mises, 1919b, p. 60).
\(^{97}\) The problem of ‘applicability’, that is that of finding out how the ‘correspondence’ between mathematical or logical theories and empirical reality works, would haunt Von Mises during his entire career. For example, in 1938 Von Mises wrote that it is ‘a big problem’ and concluded that ‘[b]etween the exact theories and reality there lies an area of vagueness and “unspeakability”’ (Von Mises, 1939 [1968], p. 123).
\(^{98}\) Von Mises, 1919b, p. 63.
\(^{99}\) Von Mises, 1919b, p. 63.
\(^{100}\) Von Mises, 1919b, p. 63.
\(^{101}\) Von Mises, 1919b, p. 63.
Taken together, the ‘pragmatic probabilists’\textsuperscript{102} Von Mises justified the choice and features of his ‘axiomatics’ not in inner-mathematical terms of formal (‘Hilbertian’) criteria, but with extra-mathematical reference to the fact that as the necessary consequence of the frequency theory (embodied in the ‘empirical collectives’) it is the best guarantee of their scientific description.\textsuperscript{103} The fact that he also allowed for deductively drawing conclusions from the more or less arbitrary and tautological ‘axiomatics’ that could possibly also be expressed in the language of set- and measure-theory explains why there is no contradiction in the fact that [Von Mises] dismissed an exclusively formalistic [or] “modern” notion of mathematical fundamentals, while at the same time working [on] the modernization of probability’.\textsuperscript{104,105}

\subsection*{1.1.2.1 Von Mises and set- and measure theory}

In his 1919 papers, Von Mises criticized not only the traditional Laplacean definition of probability as the quotient of the number of favorable cases over the number of all possible cases, but also Borel’s number-theoretical and Felix}

\begin{table}[h]
\begin{tabular}{ll}
\textsuperscript{102} & Hochkirchen, 1999, p. 164. \\
\textsuperscript{103} & On the one hand, for example Van Lambalgen writes that ‘Kollektives are a necessary consequence of the frequency interpretation’, in the sense that if one interprets probability as limiting relative frequency, then infinite series of outcomes will exhibit Kollektiv-like properties. Therefore, if one wants to axiomatise the frequency interpretation, these properties have to be built in’ (Van Lambalgen, 1996, p. 354). It may here be remarked that John von Neumann put forward exactly this viewpoint in his famous 1932 book on the foundations of quantum mechanics – where he wrote that ‘collectives, are absolutely necessary for establishing probability theory as the theory of frequencies’ (‘Kollektive [...], sind überhaupt notwendig, um die Wahrscheinlichkeitsrechnung als Lehre von den Häufigkeiten begründen zu können’ (Von Neumann, 1932 [1971], p. 298, f. 156) (Thanks to Arianna Borrelli for this reference). On the other hand, for example Siegmund-Schultze observes that to Von Mises ‘the axioms seemed to be the best guarantee of describing genuinely probabilistic situations, relevant to real applications’ (Siegmund-Schultze, 2010, p. 225).
\textsuperscript{104} & Fischer, 2011, p. 192. \\
\textsuperscript{105} & See Gillies, 2000 [2006], pp. 90-92 for a helpful diagram of Von Mises’s views on these connections.
\end{tabular}
\end{table}
Hausdorff’s (1868-1942) (then-recent) measure-theoretical definitions.\textsuperscript{106} Both Borel and Hausdorff are said to ‘even further extend the notion “probability”, which is already burdened by ambiguities’,\textsuperscript{107} such that it becomes ‘too unspecified for a theory […] to be built upon it, because it [does] not contain a hint of the occurrence of events in [a] series of trials and no direct reference to randomness’.\textsuperscript{108} Von Mises had no ‘principled mathematical reservations’\textsuperscript{109} against Borel’s and Hausdorff’s contributions as either, in the case of Borel, an investigation into ‘improper probabilities’ (‘uneigentlichen Wahrscheinlichkeiten’)\textsuperscript{110} that is of merely mathematical interest,\textsuperscript{111} or, in that of Hausdorff, a specific unpractical proposal for the tautological or formal side of probability theory. But he emphasized that neither Borel nor Hausdorff was able to make their work bear upon probability theory’s scientific study of empirical reality.

Von Mises himself did attempt to express (not define!), for the sake of convenience,\textsuperscript{112} his own ‘mathematical collectives’ in set-theoretical terms and even ‘tried to indicate the possibility of the construction of a bridge toward measure

\begin{itemize}
  \item \textsuperscript{106} Where Borel, thus, wrote of ‘denumerable probabilities’ as an investigation of the asymptotic behavior of the relative frequency with which individual digits occur in a representation of real numbers, Hausdorff, in his \textit{Grundzüge der Mengenlehre} of 1914, defined probability as ‘the quotient of the measure of a point set divided by the measure of the set in which that point set is contained’ (‘den Quotienten des Masses einer Punktmenge durch das Mass einer Menge, in der sie enthalten ist’) (Von Mises, 1919b, p. 66).
  \item \textsuperscript{107} ‘[D]en schon ohnehin durch mancherlei Vieldeutigkeiten belasteten Gebrauch des Ausdruckes “Wahrscheinlichkeit” […] noch weiter auszudehnen’ (Von Mises, 1919b, p. 66).
  \item \textsuperscript{108} Siegmund-Schultze, 2010, p. 208.
  \item \textsuperscript{109} Siegmund-Schultze, 2010, p. 208, emphasis in original.
  \item \textsuperscript{110} Von Mises, 1919b, p. 66.
  \item \textsuperscript{111} Von Mises mentioned not only theoretical physics, statistics and games of chance as the fields of application of probability theory, but also the pure mathematical problem of the limiting relative frequencies of the digits 0’s and 1’s. This problem, which was recently being expressed in terms of probability theory (‘in neuerer Zeit unter Verwendung der Ausdrucksweise der Wahrscheinlichkeitsrechnung zur Sprache gebracht wurden’ (Von Mises, 1919b, p. 65)), Von Mises understood as dealing with sequences that satisfy axiom (L), but not axiom (IR).
  \item \textsuperscript{112} See, for example, Von Mises 1920 where Von Mises wrote of ‘der Bequemlichkeit halber eingeführte’ representation in set- and measure-theoretical terms of a notion such as distribution.
\end{itemize}
theory’ despite his wish to avoid its ‘exotic’ analytic sophistications. Given the increasing variety and complexity of probability distributions, the modern mathematical theories forced themselves on Von Mises, especially because the generalization of probability theory that these theories made possible was needed to account for the distributions found in applications in several scientific disciplines that could not be reduced to discrete ones and were non-normal. Von Mises’s endeavor, essentially, took the form of ‘fitting’ set- and measure-theoretical concepts to the ‘axiomatics’ (proposition (‘Satz’) 1-7) and distributions (proposition (‘Satz’) 8-14) of his Grundlagen and hoping that they could be shown to possess the properties implied in these concepts. In other words, Von Mises assumed for limits of relative frequencies and probability distributions certain set- and/or measure-theoretical properties that ‘one should intuitively expect’ and not worried too much about counter-intuitive technicalities. When it turned out that these ‘technicalities’ undermined such assumptions – e.g. the limits of relative frequencies, presented as set-functions, are defined for arbitrary subsets only in the case of trivial discrete (‘arithmetic’) distributions and the probability distributions, ‘imagined’ (‘vorstellen’) as a measure function for each set, are not countably additive –, Von Mises tended to maintain the property in a weakened sense for practically relevant cases (or sets).

For example, Von Mises wrote that ‘the totality of $W_A$ [probability- or limit-] values for all point sets $A$ of the label space constitutes the distribution; it is sufficient to consider only such $A$ that are subsets of $M$. One can imagine the function $W_A$ as a (partly continuous, partly discontinuous) distribution of the

---

115 Siegmund-Schultze writes that ‘[n]o doubt von Mises saw that increasingly complicated probability distributions occurred in applications in biology, physics, astronomy, statistics, and insurance mathematics [...] In particular, it had become clear that probability distributions that could not be reduced to discrete ones did not necessarily possess representations as classical integrals over “densities”. And even if densities existed, they could not necessarily be reduced to traditional ones such as the Gaussian ‘bell curve’ (Siegmund-Schultze, 2010, p. 210).
117 Von Mises, 1919b, p. 56.
mass 1 over the points of $M$. Although Von Mises was aware, among other things, of the role of sets of probability $\circ$ and emphasized that $\mathcal{W}A = \circ$ is not always equivalent to “impossibility [and] $\mathcal{W}A = 1$ not always to “certainty” of $A$, he, on the one hand, upheld the ‘collective’ ‘as standing behind the set function’, and, on the other hand, assumed $\mathcal{W}A$ as a set function and a measure function ‘which is measurable for each point set’ Von Mises, thereby giving in to the temptation of claiming more advanced set- and measure-theoretical properties for his probabilities on the basis a quite basic ‘parallelism’ between the two, thus expected that since all events have probabilities, probabilities will always exist for any set of labels (sets of elementary events) and they can be expressed in terms of a mathematical notion (a measure function) that assigns (‘outer’) measures to all sets. After it became apparent

---

118 ‘Die Gesamtheit der $\mathcal{W}A$-Werte für alle Punktmengen $A$ des Merkmalraumes bildet die Verteilung; es genügt dabei, nur solche $A$ zu betrachten, die Teilmengen von $M$ sind. Man kann sich die Funktion $\mathcal{W}A$ unter dem Bilde einer (teils stetigen, teils unstetigen) Verteilung der Masse $1$ über die Punkte von $M$ vorstellen’ (Von Mises, 1919b, p. 56, my emphasis).

119 ‘$\mathcal{W}A = 0$ ist nicht immer gleichbedeutend mit “Unmöglichkeit”, $\mathcal{W}A = 1$ nicht immer mit “Sicherheit” von $A$’ (Von Mises, 1919b, p. 56).

120 Siegmund-Schultze, 2006, p. 213.

121 ‘[F]ür die jede Punktmenge Messbarkeit besitzt’ (Von Mises, 1919b, p. 66).

122 Siegmund-Schultze explains that this rash assumption was based on an applied mathematician’s temptation or intuition to find ‘a satisfactory point of departure for a set theoretic founding of his notion of probability in [Constantin Carathéodory’s (1873-1950)] notion of “outer measure”’ (Siegmund-Schultze, 2006, p. 214) the subtlety of which ‘escaped’ Von Mises’s attention (see Siegmund-Schultze, 2006, pp. 214-217).

123 ‘Von Mises […] realized as basic facts the (finite) additivity of probabilities of mutually exclusive events and – parallel to this – the (finite) additivity of limits of relative frequencies of the occurrence of events in trial sequences […] That this parallelism could be extended to infinite [sigma]-additivity […] was clear to von Mises well’ (Siegmund-Schultze, 2006, p. 214).

124 Von Mises referred to Carathéodory’s ‘measure function’ and what is today called ‘Carathéodory (outer) measure’ – or, taken together, to the idea of a function which assigns a nonnegative number $\mu^*$ to any subset (point set) of a set $M$ under certain conditions. What ‘escaped’ Von Mises (see footnote 425), as Hausdorff pointed out, was that ‘each point set has an outer measure (a value assigned) but it is not necessarily “measurable” with respect to that outer measure’ (Siegmund-Schultze, 2006, p. 215).
that this was not the case,\textsuperscript{125} Von Mises simply maintained that ‘it remains valid for arithmetic and geometric distributions, which alone occur in applications [...]’\textsuperscript{126}, that is, ‘in all practical problems of the calculus of probability’.\textsuperscript{127}

Where pure mathematicians such as Hausdorff and many others would define probability as a measure of sets, the ‘eclectic’ applied mathematician Von Mises used set- and measure-theory to refine the mathematically expressed idealizing representation of mass phenomena.\textsuperscript{128} Rather than approaching probability ‘even “more axiomatically” than those probabilists who chose the measure-theoretic way’,\textsuperscript{129} Von Mises’s statement that it was not possible to ‘prove the “existence” of collectives by analytical construction’\textsuperscript{130} is suggestive of his commitment to the view that the relation between the ‘extra-’ and ‘inner-math-

\begin{itemize}
\item \textsuperscript{125} Von Mises admitted this much in his ‘Berichtigung zu meiner Arbeit “Grundlagen der Wahrscheinlichkeitsrechnung”’ of 1920 in which he wrote that ‘[t]he remark in the beginning of [Von Mises, 1919b, p. 66], according to which the probability $W_a$ is a measure function each set is based [...] on error’ (‘Die Bemerkung zu Beginn von [Von Mises, 1919b, p. 66], wonach die Wahrscheinlichkeit $W_a$ Massfunktion fur jede Menge sein soll, berucht [...] auf Irrtum’ (see Von Mises, 1920a, p. 323)).
\item \textsuperscript{126} Von Mises, 1920a, p. 323.
\item \textsuperscript{127} Von Mises, 1919c, p. 5 quoted in Siegmund-Schultze, 2010, p. 460.
\item \textsuperscript{128} For example, as Von Mises wrote in a 1919 book review, because ‘all real observations [...] can obviously only cover finite sequences of elements, [and] a collective, by definition, consists of infinitely many elements, [...] it is impossible to speak of an identity’ (‘[a]lle wirklich Beobachtungen können sich natürlich nur auf endliche Elementfolgen erstrecken, während ein Kollektiv definitioengemäß aus unendlich vielen Elementen besteht. Daraus folgt schon, dass von einer Identität’ (Von Mises, 1919d, p. 172) of probabilities and the calculations approximating them.
\item \textsuperscript{129} Siegmund-Schultze, 2006, p. 226.
\item \textsuperscript{130} (‘[D]ie “Existenz” durch eine analytische Konstruktionen nachweisen’ (Von Mises, 1919b, p. 60)).
\end{itemize}
ematical’ sides of mathematics is part of the ‘axiomatics’ itself. More in specific, even though the ‘axiomatics’, as an idealizing image, is independent from observation (‘äußeren Erfahrung’) its axioms (‘Axiome’) for probability are chosen such that ‘an application of the results to the objects (‘Gegenstände’) of the real external world is possible’.

---

131 For example, in his programmatic paper in the first volume of the *ZAMM* Von Mises wrote that ‘[f]rom abstract-logical investigations, which extend to the realm of philosophy, reaching to the rational considerations of everyday life which are directed to measure and number, there is put up a chain of repeatedly intertwined links, which includes what we denote as mathematics in a general sense. Everybody among us […] is put at a specific place in this chain, where he usually sees only a small part of the whole. Within this small part he draws arbitrarily a border and calls that which is lying to the left of it, pointing to the more abstract, “pure” mathematics, while he calls what lies to the right and mediates the connection to practical life, “applied”. Here, no rigid separation is possible, no part of the whole can be fully separated, if the chain is not to lose its tension’ (Von Mises, 1921a, p. 2). Von Mises repeated this ‘Machian’ view (as he himself termed it) in less ‘un-philosophical’ terms, namely by arguing for it in light of the work of Hilbert, Bertrand Russell (1872-1970) and L.E.J. Brouwer (1881-1966), in his *Kleines Lehrbuch des Positivismus* of 1938 (see Von Mises, 1938 [1968], chapter 10 & 11).

132 Von Mises, 1919d, p. 172.

2. **Von Mises and early-1920s mechanical and statistical physics: the Forman-thesis**

The fundamental reason for the initial success\(^\text{134}\) of the 1919 papers has often been sought in Von Mises’s ability to present his probability theory as a (Machian) solution to the crisis in mechanical physics of the early-1920s. And the adoption of Von Mises’s approach to probability by mathematical physicists such as Max von Laue (1879-1960), Leonid Mandel’shtam (1879-1944) and John von Neumann (1903-1957), indeed, seems to justify this explanation. Aleksandr Iakovlevich Khinchin (1894-1959) would capture the spirit of the reception of the work of Von Mises during the 1920s and early 1930s in terms of it being ‘inadmissible to reject on principle, because of its […] purely formal imperfections, this theory which so brilliantly conforms to the essence and requirements of scientific practice’.\(^\text{135}\) In light of what follows it is important to also draw attention to the fact that Von Mises himself hoped to find the more general influence of his work in its introduction of a certain ‘enlargement of mathematical thinking’ that is needed ‘in order to master the problems of probability’\(^\text{136}\).

---

\(^{134}\) Von Mises claimed victory for his presentation of the foundations of probability theory in 1932 when he, in his lecture for the *International Congress of Mathematicians* in Zurich, said that it is ‘generally acknowledged today’ (Von Mises 1932 quoted and translated into English in Siegmund-Schultze, 2004, p. 358). The defeat of his axiomatization was sealed at a 1937 conference in Geneva where all the objections against his axiomatization and the advantages of Kolmogorov’s 1933 axiomatization were enumerated.

\(^{135}\) See Khinchin, 1929 [2015].

In his ‘Spenglerian’ or ‘Rilkean’ paper entitled ‘On the present crisis in mechanics’ the fundamental question was whether classical Newtonian deterministic mechanics with its rigid causal structure and differential equations can explain ‘all phenomena of motion and equilibrium which we observe in visible bodies’. The question was answered negatively with reference to the fact that ‘within the purely empirical mechanics there are phenomena […] which will forever escape an explanation on the basis of [classical mechanics] and stand in need of the construction of a closed theory of mechanical statistics’, such as the movement of liquids (‘Flüssigkeiten’), the elasticity of solid bodies (‘festen Körper’) and Brownian motion. Von Mises’s new problem then became that of determining the boundary (‘Grenze’) between the domains of validity (‘Geltungsbereich’) of classical and statistical mechanics. Where, for instance, Albert Einstein (1879-1955) had earlier

---

137 As a strong case for what is known as the ‘Forman Thesis’ (see Carson, Kolevnikov, Trischler, 2011), Forman has written that ‘a direct influence [upon Von Mises] of [Oswald] Spengler [(1880-1936) can […] be established by September 1921. When at this time von Mises added an appendix to the republication of his lecture of February 1920, his tone had changed entirely, his optimism and enthusiasm had disappeared. Von Mises had largely, and explicitly, adopted Spengler’s perspective and assumptions’ (Forman, 1971, p. 51). The case is such that in February 1920 Von Mises had given an address that took for granted causality and that ‘when one turns to the [1921] appendix […] one finds his attitude to causality […] entirely transformed’ (Forman, 1971, p. 81). Where Forman attributes this sudden transformation to a Spenglerian capitulation to acausality others have shown the difficulties involved in this causal explanation (e.g. Nola, 2003, pp. 239-241).

138 Von Mises, who owned the largest private Rilke collection in the world, was deeply interested in the poet Rainer Maria Rilke and his conversion to Catholicism (Von Mises himself converted to Catholicism sometime between 1909 and 1914). It has been suggested that this could have provided an emotional-intellectual influence away from causality and determinism (e.g. Feuer, 1974 [1989], p. 233, f. 290).

139 See, for example, Hochkirchen, 1999, pp. 167-173; Von Plato, 1994, pp. 189-192 for a treatment of this paper.

140 ‘[A]lle Bewegungs- und Gleichgewichts-Erscheinungen, die wir an sichtbarem Körpern beobachten’ (Von Mises, 1921b, p. 427).


142 Von Mises, 1921b, p. 427.

143 Von Mises, 1921b, p. 428.

144 See Von Mises, 1921b, pp. 428-429.

145 Von Mises, 1921b, p. 429.

146 Von Mises, 1921b, p. 429.
attempted to either ground statistical mechanics on classical dynamics or use statistical mechanics as a bridge between classical and modern (or ‘new’) physics.\textsuperscript{147} Von Mises developed a purely probabilistic approach to statistical mechanics, dubbed ‘mechanical statistics’ (‘mechanische Statistik’),\textsuperscript{148} for which classical mechanics is no longer a relevant framework.\textsuperscript{149} It was Von Mises’s own frequency probability theory that furnished this approach with a suitable ontology, that is, ‘with its own independent object of physical theorizing in the same vein as Newtonian point particles’.\textsuperscript{150} Where the Newtonian dynamical laws determined the motions of such individual particles, the statistical laws concerns random (mechanical, electrical, etc.) mass phenomena. And where Newtonian mechanics calculated outcomes from specified initial states, statistical mechanics calculates probabilities, that is, limiting relative frequencies, by deriving them from initial probabilities (‘gegebenen Wahrscheinlichkeiten’).\textsuperscript{151} This means that in so far as the probability laws of statistical mechanics never make definite statements about the sequence (in time) (‘zeitlichen Ablauf’\textsuperscript{152}) of single events (‘Einzelvorgänge’)\textsuperscript{153} it cannot ‘come into direct conflict with a result of [classical] mechanics or the rest of deterministic physics’.\textsuperscript{154} Von Mises’s general proposal for a solution to the crisis in physics was, thus, to advocate a probabilistic approach of (pre-quantum mechanical) fundamental and irremovable indeterminacy – albeit without opting for a final decision between determinism and indeterminism.\textsuperscript{155}

Importantly, in the Grundlagen of 1919 Von Mises had written that in the application to theoretical physics neither the given probabilities nor the resulting probabilities (‘Resultate der Wahrscheinlichkeitsberechnung’\textsuperscript{156}) immediately ‘border’ reality because ‘physical hypotheses (‘Hypothesen von physikalischer Natur’)

\begin{footnotesize}
\textsuperscript{147} See, for instance, Kox, 2014; Renn & Rynasiewicz, 2014, section 5.
\textsuperscript{148} Von Mises, 1921b, p. 429.
\textsuperscript{149} See Von Plato, 1994, section 6.3
\textsuperscript{150} Stöltzner, 2003, p. 216.
\textsuperscript{151} Von Mises, 1921b, p. 429.
\textsuperscript{152} Von Mises, 1921b, p. 429.
\textsuperscript{153} Von Mises, 1921b, p. 429.
\textsuperscript{154} ‘[K]ann so niemals in unmittelbare Konkurrenz treten mit einem Ergebnis der Mechanik oder der übrigen deterministischen Physik’ (Von Mises, 1929b, p. 429, my emphasis).
\textsuperscript{155} For this point see Heidelberger, 2004, p. 315; Stöltzner, 2003, p. 200, pp. 211-223.
\textsuperscript{156} Von Mises, 1919b, p. 63.
\end{footnotesize}
lie between them’. On the one hand, ‘the initial values (‘Ausgangswerte’) of the probability calculation (‘Wahrscheinlichkeitsberechnung’), for instance the assumption (‘Annahme’) of equiprobability of each velocity for every molecule in a gas at complete rest (‘ganzen ruhenden Gases’), belong to the hypothetical part of the physical image the justifiability [of which] is examined with reference to the results of the theory [i.e. probability theory]. On the other hand, ‘not the immediate results of the [calculation] experience a check (‘erfahren eine Kontrolle’) by means of observation […] but only those conclusions which are connected to the […] distributions via certain specified hypotheses’. The fact that these hypotheses, the existence of whose entities (atoms, molecules etc.) Von Mises, the orthodox Machian, questioned, somehow determined the ontological status of probabilities differed from the presentation of the theory as belonging to the indeterministic part of theoretical physics dealing directly with probabilistic phenomena in the 1921 paper. Von Mises, apparently, found the specification of the given probabilities in initial collectives and the probabilities resulting from calculation ‘decisive for the scientific import of probability calculus and for an ontology suitable to statistical physics’. For example, his dismissal of the Boltzmannian ergodic hypothesis (and the related notions of time-average and ensemble), as it had appeared in Einstein’s early attempt to use statistical mechanics to find evidence for a deterministic ‘microworld of

159 ‘[N]icht die unmittelbaren Resultate der Wahrscheinlichkeitsberechnung erfahren eine Kontrolle durch die Beobachtung […] sondern nur Folgerungen, die erst durch bestimmte Hypothese mit den […] Verteilungen verknüpft sind’ (Von Mises, 1919b, p. 63).
160 For accounts of Mach (or Machianism) and atomism see, for example, Brush, 1968; Laudan, 1981, chapter 13.
161 Stöltzner, 2003, p. 219, my emphasis.
162 For a treatment of the place and meaning of these notions in Boltzmann’s work see, for example, Brush, 2006; Von Plato, 1991.
atoms’, involved the physical argument that classical physics and genuinely probabilistic phenomena are in contradiction with each other – and this because, on the one hand, a statistical viewpoint ‘in the large’ (i.e. of observable macrosystems) is irreconcilable with an unscientific determinism ‘in the small’ (i.e. of unobservable microsystems) and, on the other hand, it is impossible to derive statistical statements from differential equations. The ergodic hypothesis, the assumption that during its evolution a deterministic mechanical system will eventually pass through all exact and equally probable (but unobservable) microstates, had been introduced to provide a justification, in terms of statistical averaging, of the intuitive statement that a gas passes from a very improbable state ‘through ever more probable states, reaching finally the most probable of all, i.e. thermal equilibrium’. Boltzmann attempted to prove that a gas during a motion of unlimited duration corresponds to the so-called Maxwell-Boltzmann distribution by means of the idea that if probability gives ‘the relative time one molecule is in a given state during a very long time’ and if this is equal to ‘the relative number of all the molecules in that state at a single time’, it follows that ‘[w]ith a great number of molecules, their relative numbers in different states would approximate a continuous [velocity] distribution over the states’. Von Mises’s fundamental problem with the mechanical interpretation of the ergodic hypothesis (and, later, ‘theory’) for statistical mechanics was twofold: on the one hand, that the ‘premises do not refer to empirically determined probabilities, whereas the conclusion does (in the form of time-av-

---

163 Renn, 2000, p. 148.
164 As is well known, this attempt stood firmly in the Maxwellian tradition of the investigation into the atomistic constitution of gases in specific and matter in general on strictly mechanical principles via probabilities, to which also belonged Boltzmann’s attempt to reduce the second law of thermodynamics to mechanics.
165 Reflecting on Brownian motion, Von Mises wrote that ‘[i]t is entirely irrelevant whether we stick to the assumption that the orbits would be determined if we knew the exact initial conditions and all influences; since we have no prospect of ever achieving this knowledge, this is an assumption of which it can never be decided whether it is true or not, hence an unscientific one’ (‘Es ist ganz gleichgültig, ob wir an der Annahme festhalten, die Bahnen waren bestimmt, wenn wir die genauen Anfangsbedingungen und all Einflüsse kennten; denn da wir keine Aussicht haben, die Kenntnis je zu erlangen, so ist es eine Annahme, von der sich nie entscheiden lasst, ob sie richtig ist oder nicht, also eine nicht wissenschaftliche’) (Von Mises, 1921b, p. 430).
166 For these arguments see Von Mises, 1920b; Von Mises, 1930.
168 For a particularly lucid discussion of Boltzmann’s use of probability see Krüger, 1980.
erages’\textsuperscript{169} and, on the other hand, that probabilistic considerations are said to lead to ‘definite statements about the actual course of single systems’.\textsuperscript{170} The task of his own ‘purely probabilistic’\textsuperscript{171} approach, which squeezed the ‘new physics’ into the indeterminist-inspired axiomatics of probability theory,\textsuperscript{172} thereby giving a justification for physicists to concern themselves with its foundations\textsuperscript{173} and emphasizing the mechanical-statistical reason for its very axiomatization,\textsuperscript{174} was to make predictions about the time development of a system on the basis of a finite number of observable macrostates exhibiting the Markov-property:\textsuperscript{175}

‘Given the present macrostate of the system, the transition probability to the next macrostate depends only on what the present state is, but not on the previous “history” of the system. If the transition probabilities remain the same in time and if they fulfill certain additional conditions, there will be a unique limiting distribution for the macrostates of the system’.\textsuperscript{176}

\textsuperscript{169} Van Lambalgen, 1996, pp. 351-352.
\textsuperscript{170} Von Plato, 1994, p. 191, my emphasis.
\textsuperscript{171} Von Plato, 1994, p. 191.
\textsuperscript{172} Hochkirchen characterizes Von Mises as a convinced indeterminist ‘who axiomatized chance on the basis of this philosophical conviction and then tried to squeeze physics into this framework’ (‘der auf der Basis dieser philosophischen Überzeugung den Zufall axiomatisieren und dann die Physik in dieses Schema pressen wollte’ (Hochkirchen, 1999, p. 164)).
\textsuperscript{173} As Von Mises wrote: ‘in the last decade, the development of probability theory has unfortunately been very much neglected or at least steered in very deceptive directions. One […] has totally lost the idea, that it here concerns a serious natural scientific theory dealing with a well-defined class of natural scientific appearances’ (‘leider ist die Entwicklung der Wahrscheinlichkeitstheorie in den letzten Jahrzehnten sehr nachlässigt oder zumindest auf höchst abwegige Bahnen gelenkt worden. Man […] hat ganz das Gefühl dafür verloren, dass es sich hier um eine ernsthafte naturwissenschaftliche Theorie für eine bestimmte Klasse naturwissenschaftlicher Erscheinungen handelt’) (Von Mises, 1921b, p. 429).
\textsuperscript{174} Hochkirchen writes that Von Mises was definitely of the opinion that ‘probabilistic arguments were needed in physics, and his attempt (\textit{Versuch}) to axiomatize probability calculus was causally inspired by this [need]’ ([Daß] probabilistische Argumente in der Physik benötigt werden, und sein Versuch, die Wahrscheinlichkeitsrechnung zu axiomatisieren, war davor ursächlich motiviert) (Hochkirchen, 1999, p. 173).
\textsuperscript{175} A so-called Markov-chain is a sequence of random variables with the ‘Markov-‘ or ‘memoryless-property’ that given the present state, the future and past states are independent.
\textsuperscript{176} Von Plato, 1994, pp. 191-192.
Von Mises's dismissal of a mechanical interpretation of the ergodic hypothesis for dynamical systems in favor of a probabilistic description of statistical systems was ‘instrumental in formulating the program that led to a purely probabilistic ergodic theory’\(^{177}\) in the work of physicists and applied mathematicians (probabilists) such as Leonid Isaakovich Mandel'shtam (1879-1944), Mikhail Leontovich (1903-1981) and Khinchin from the late-1920s and early 1930s.\(^{178}\)

3. **Von Mises’s 1928 Wahrscheinlichkeit, Statistik und Wahrheit: foundations**

Von Mises’s semi-popular *Wahrscheinlichkeit, Statistik und Wahrheit* was first published in 1928, with a second, third and fourth German edition in 1936, 1951 and 1972, a first and second English edition in 1939 and 1957\(^{179}\) and a Russian edition (under the editorship of Khinchin) entitled *Veroiatnost’ i statistika* (‘Probability and Statistics’) in 1930 (!). The first German edition appeared at ‘Verlag von Julius Springer’ as the first volume (‘Band’) in Moritz Schlick (1882-1936) and Frank’s *Schriften zur wissenschaftlichen Weltauffassung* – the book series of the second Wiener Kreis which represented the viewpoints of, but was ‘not as radical, Machist, or “unity of science” focused’\(^{180}\) as the one of the Verein Ernst Mach put out by Rudolf Carnap (1891-1971) and Neurath.\(^{181}\) Von Mises himself had belonged to the core of the first Wiener Kreis; between 1907 and 1912 he regularly met Frank, Hahn and Neurath in a Viennese coffeehouse to discuss the work of Mach, Boltzmann and the French conventionalists in order to ‘overcome metaphysical philosophy through a synthesis of empiricism and [symbolic] logic’,\(^{182}\) thereby also answering Vladimir Lenin’s (1870-1924) attack on ‘Machism’ in his *Materialism and Empirio-criticism* (‘Materialism und Erkenntnistheorie von der Empirio-Critik’).

---

\(^{177}\) Von Plato, 1994, p. 108.

\(^{178}\) Von Mises's mathematical results and methodological approach ‘were important for the development of statistical mechanics and ergodic theory, at least until the rival Kolmogorovian approach and the general theory of stochastic processes appeared on the scene in the early 1930s’ (Siegmund-Schultze, 2004, p. 359).

\(^{179}\) The second English edition was a retranslation of the third German edition of 1951.

\(^{180}\) Stadler, 1992, p. 373.

\(^{181}\) For an account of the similarities and differences between the second Wiener Kreis and the Verein Ernst Mach see, for example, Siegetsleitner, 2014, section 1.2.2; Stadler, 1985; Stadler, 1985; Stadler, 1992.

\(^{182}\) Stadler, 2007, p. 16.
Richard von Mises’s philosophy of probability and mathematics: a historical reconstruction

Despite him being a committed Machian and ‘moderate positivist’ and the author of the first textbook on positivism (not logical empiricism), Von Mises ‘deliberately choose a certain distance’ to the second Wiener Kreis (1922-1936), the Verein Ernst Mach and Hans Reichenbach’s (1891-1953) Berlin Gesellschaft für Empirische Philosophie (1926-?) – and this on theoretical and political grounds; he dismissed, on the one hand, the ‘metaphysical’ character of their, in his eyes, all too radical reduction of metaphysics to the empirical sciences and, on the other hand, the leftist (non-individualistic, –nationalistic, or -elitist) ambitions allegedly attached to the political epistemology of the logical positivists and empiricists.

Wahrscheinlichkeit, Statistik und Wahrheit consisted of the six parts (‘Abschnitte’) of a talk delivered in January 1914 in Strassburg and in December 1922

183 See Stadler, 2003, p. xiii for this observation.
184 See, for example, Von Mises’s ‘Ernst Mach und die empiristische Wissenschafsauffassung: Zu E. Machs 100. Geburtstag am 18. Feb. 1938’ (Von Mises, 1938).
185 Von Mises, 1951 [1957], p. 146.
186 See Von Mises’s Kleines Lehrbuch des Positivismus of 1939 (Von Mises, 1939a) and, for example, also his ‘Scientific conception of the world. On a new textbook of positivism’ of 1940 (Von Mises, 1940).
187 Von Mises declined the proposal of Neurath, the editor of the Einheitswissenschaften series of the Wiener Kreis, to entitle his textbook (i.e. Von Mises, 1939a) ‘Logical (or scientific) Empiricism’ (see Siegmund-Schultze, 2004, p. 339; Stadler, 1990, p. 33).
189 For an account of the ‘Berlin society’ and/or ‘Berlin group’ see Rescher, 2006 and for a comparison with the second Wiener Kreis see, for example, Milkov, 2013.
190 Von Mises, for example, could write that ‘[a]ll metaphysicians make statements that in some way or other concern reality, the world of experience, and that often enough are intended even to produce changes in the world (via human actions)’ (Von Mises, 1939 [1968], p. 263). It may also be observed that his textbook on positivism contained chapters containing in-depth analyses of topics such as poetry, art, religion and ethics (see Von Mises, 1939a, chapter 23, 24, 27-28).
191 Von Mises could be characterized as a disengaged, individualist and elitist (or aristocratic), liberal with nationalistic leanings. But given him being a liberal in social and cultural matters, ‘there is no way to compare von Mises to the “average” conservative and nationalist German professor whose concern was primarily about his social privileges’ (Siegmund-Schultze, 2004, p. 352).
192 For an account of the political ambitions of the political aims of, for example, the Wiener Kreis see for instance Uebel, 2005. See, for example, Richardson, for a detailed argument in favor of the claim that most of the proponents of logical positivism and empiricism were committed to ‘political neutralism’.
in Berlin\textsuperscript{194} – its main ideas having been put forward in a popular form in a short review-essay\textsuperscript{195} of 1919 and its mathematical substantiation (‘Begründung’) having been provided in the \textit{Grundlagen} of the same year.\textsuperscript{196} \textsuperscript{197} Both in the preface to the first German edition of 8 June 1928 as well as in the preface to, for instance, the first English edition, Von Mises explicitly questioned the title of the series in which the book had originally appeared by distancing himself from the idea that it ‘is in some way connected to metaphysics’:\textsuperscript{198}

‘Every theory, which the investigator puts forward for a group of observable appearances, embodies the bigger or smaller part of a scientific worldview. Only [everyday language] gives words like ‘world-conception’ [‘\textit{Weltauffassung}’] or ‘\textit{Weltanschauung}’\textsuperscript{199} a metaphysical meaning transgressing the purely scientific […] I find it important [to] dismiss any aspiration in this direction.\textsuperscript{200}

At the same time, Von Mises could not but approach \textit{philosophical} questions about the natural-scientific discipline of probability theory from the perspective of the \textit{natural scientist} searching for ‘the simplest systematic description

\begin{itemize}
\item \textsuperscript{194} See Von Mises, 1928, p. iii.
\item \textsuperscript{195} Von Mises, 1919d.
\item \textsuperscript{196} See Von Mises, 1928, p. 181.
\item \textsuperscript{197} It may be remarked that the fourth ‘lecture’ (‘\textit{Abschnitt}’) of the book had appeared, in a slightly different from, on June 17, 1927 as a paper entitled ‘Über das Gesetz der grossen Zahlen und die Häufigkeitstheorie der Wahrscheinlichkeit’ (see Von Mises, 1927).
\item \textsuperscript{198} Von Mises, 1936 [1939], p. viii.
\item \textsuperscript{199} Apparently, the word \textit{Weltauffassung} (‘world-conception’) chosen by the Vienna empiricists to replace ‘Weltanschauung’ was in Von Mises’s opinion still too close to the German metaphysical ‘Lebensphilosophie’ (‘philosophy of life’) of Wilhelm Dilthey (1833–1911) and others (see Holton, 1992, p. 47, f. 57; Siegmund-Schultze, 2004, p. 349, f. 35).
\item \textsuperscript{200} ‘[J]ede Theorie, die der Naturforscher für irgend eine Gruppe beobachtbarer Erscheinungen gibt, bildet das größere oder kleinere Stück eines wissenschaftlichen Weltbildes. Allein ein verbreiteter Sprachgebrauch leih Worten wie “Weltauffassung” […] eine über das rein Wissenschaftliche […] hinausreichende, metaphysische Bedeutung. Es liegt mir daran, […] jeden Anspruch in dieser Richtung abzulehnen’ (Von Mises, 1928, pp. iii-iv).
\end{itemize}
of perceptually observable matters of fact'. The fact that the book did not contain any mathematical formulae and was neither of a mathematical nor of a physical character was accounted for in terms of the view that in so far as mathematics can be understood, in the traditional sense (‘im alteren Sinne’), as the unity or totality (‘Gesamtheit’) of the exact sciences, it is written ‘entirely and exclusively’ (‘durchaus und ausschliesslich’) as a mathematician. This statement can be made sense of with reference to Von Mises’s description of the fundamental aim of Wahrscheinlichkeit, Statistik und Wahrheit as giving the definition of probability (‘Wahrscheinlichkeit’) by means of the synthetic method of reasoning which, at least in the Kantian understanding, is that of mathematics as opposed
to the *analytic* method of philosophy. Following Kant and Mach – with ‘the intellect’ and ‘a posteriori’ somehow replacing ‘reason’ and ‘a priori’ – this meant that out of the definition of probability found in everyday language, dictionaries, etc. a new rational or scientific concept of probability is constructed by elimination. Where the *status* of this concept is described as that of an ‘exact idealized conception’ (‘*exakten, idealisierten Begriff*’) or ‘concept-construction’ (‘*Begriffsbildung*’) not unlike the ‘straight line’ of pure geometry, its instrumental *value* is explained, in every edition of the book, as follows:

‘[A]ll the theoretical constructions […] which are used in the various branches of physics are only imperfect instruments to enable the world of empirical fact to be reconstructed in our minds […] In dealing with the theory of probability […] I do not hope to achieve more than the results already attained by geometry, mechanics, [etc]. That is to say, I aim at the construction of a rational theory, based on the simplest possible exact concepts, on which, although admittedly

204 Von Mises wrote that ‘[w]e may say, with Kant, that our aim is to give not an analytic definition of probability but a synthesis one. We may leave open the question of the general possibility of finding analytic definitions at all’ (Von Mises, 1951 [1981], p. 4) (‘Unter Verwendung einer von Kant eingeführten Bezeichnung konnte ich auch sagen: Es ist nicht unsere Absicht, eine analytische Definition der Wahrscheinlichkeit zu geben, sondern eine synthetische, wobei wir die grundsätzliche Möglichkeit analytischer Definitionen dahingestellt sein lassen können’ (Von Mises, 1928, p. 5)). In the passage of the *Kritik der Reinen Vernunft* to which Von Mises referred in the footnote to this statement Kant wrote the following about his somewhat puzzling distinction between the analytic method of philosophy and the synthetic method of mathematics: ‘Die philosophische Erkenntnis ist die Vernunftkenntniss aus Begriffen, die mathematische aus der Konstruction der Begriffe […] Die philosophische Erkenntnis betrachtet also das Besondere nur im Allgemeinen, die mathematische das Allgemeine im Besonderen, ja gar im Einzelnen […] In dieser *Form* besteht also der wesentliche Unterschied dieser beiden Arten der Vernunftkenntnisse, und beruht nicht auf dem Unterschied ihrer Materie oder Gegenstande […]’ (Kant, 1781 [1998], A713/B741-A717/B745).

205 See Von Mises, 1928, pp. 2-6.
inadequate to represent the complexity of the real processes, is able to reproduce satisfactorily some of their essential properties.\textsuperscript{206}

There are two stages to this construction, corresponding to the direct definition of probability before the existence of a collective and the indirect definition of probability* after the existence of a collective – the two definitions being reconciled after its completion. Firstly, there is the explication of those phenomena in empirical reality with which probability theory must concern itself in order to become an exact natural science, namely mass phenomena or repetitive events in infinite sequences of random events (‘empirical collective’) with the laws of large numbers as statements about the fluctuations of averages of these sequences and their subsequences. Secondly, there is the rational concept-construction of a ‘mathematical collective’ that expresses a practically unlimited sequence (‘praktisch unbegrenzte Folge’) of elements with attributes (‘events’) that occur in a random order and have a relative frequency tending to a limit (‘probability*’), thereby allowing probability theory to establish itself as an exact natural science like geometry. These two ‘stages’ account for the ambiguous fact that if the rational or scientific concept of probability is ‘the only basis of probability calculus’,\textsuperscript{207} a collective must exist before it is possible to speak of probability as part of a rational or scientific theory.

The ‘infinite sequences’ of mathematical collectives are exact idealizations from ‘finite sequences’ in so far as the two ‘axioms’ are abstracted from the two empirical laws for empirical collectives (‘limiting relative frequency’, ‘ran-


\textsuperscript{207} Von Mises, 1981 [1951], p. 11 (‘die ausschliessliche Grundlage der Wahrscheinlichkeitsrechnung’ (Von Mises, 1928, p. 12)).
domness’), as obtained from theoretically-informed observation, and make mathematically precise their two theorems (‘addition’, ‘multiplication’). This process of idealization involved in the construction of probability theory is necessary, Von Mises claimed, in order to make the mathematical representation of empirical reality tractable, that is, to mathematically express probabilistic phenomena and their laws without mathematizing them. The ‘axiomatics is thus not an axiomatization transforming probability theory into an autonomous branch of pure mathematics, but an ‘informal’ system with an abstract-logical and pragmatic, rather than a mathematical, existence. The main reason for this ‘non-Hilbertian’ approach is that the empirical (not the mathematical) law or rule of the constancy of the relative frequencies with large numbers of trials (‘Versuchszahlen’) or observations is chosen as the foundation for the mathematical theory of probability*. After noting that if his ‘probability concept largely contradicts the [‘equiprobable’] one which the older textbooks of

---

208 As Gillies writes, ‘[a]ccording to Von Mises, an empirical law is obtained by observation, and a mathematical axiom of the theory is abstracted from it. Now a rough empirical law might indeed by obtained directly from observation, but, to make it more precise, it looks as if we should temporarily abandon observation in favour of mathematics. Mathematical calculations suggest more precise versions of the empirical law, e.g. that the frequency is likely to remain constant to one decimal place after 500 goes, and that, in general, the frequency is likely to converge to its limit at the rate of $1/\sqrt{n}$. These results could then be checked by further observations’ (Gillies, 2000 [2006], p. 94).

209 See Gillies, 2006 [2000], pp. 90-91 for this explanation of Von Mises’s approach at this point.

210 In a letter of December 12, 1919 to Pólya, Von Mises wrote that he would never speak of mathematical existence given that he did not even understand what it could mean (see Siegmund-Schultze, 2006, p. 470). Reflecting on this statement, Siegmund-Schultze notes that ‘in one respect von Mises’ notion of ‘logical existence’ seems to have differed from the Hilbertian one, that is “absolutized by set theory” […] This is what one could call the pragmatic, or operationalist connotation of von Mises’s [allusion] to the possibility “to operate with the concept of a collective without contradictions arising” and hinting to the “expediency” of his definition. The “pragmatic aspect” of von Mises’s notion of “existence” is underlined by von Mises’ remark with respect to the “novelty” of his axioms [and the claim that] he was not claiming more than “having made explicit and formulated what the theory of probability of today is already based on”’ (Siegmund-Schultze, 2006, p. 471).

211 Von Mises, 1928, p. 22.

212 It was Laplace who coined the traditional definition of probability – the quotient of the number of favorable ‘cases’ over the number of all possible ‘cases’, based on (what Von Mises, among others, called) the ‘symmetry principle’ which ascribes probability values on the basis of geometrical properties.
probability calculus put forward as the *formal definition of probability*, the axiomatics is in agreement with the *actual content* derived from the concept used by these authors, Von Mises wrote the following:

'Poisson [in Poisson, 1837] describes how in the most varied areas of human experience a phenomenon has been found with which we have become familiar as the constancy of relative frequency with a large number of repetitions of observations. Poisson calls [this] matter of fact [...] the Law of Large Numbers and considers [it] the *very foundation for the possibility of probability calculus*. In his actual investigations he assumes the *formal definition* of probability introduced by Laplace [...] from which he derives, by *analytical methods*, a mathematical proposition which he then also calls the Law of Large Numbers. [But] [t]his mathematical proposition asserts something very different from that which enters the beginning of the theory as a *general empirical foundation*

---

213 ‘[Der Wahrscheinlichkeitsbegriff] widerspricht wohl im großen ganzen dem, was in den älteren Lehrbüchern der Wahrscheinlichkeitsrechnung als formelle “Definition” der Wahrscheinlichkeit an die Spitze gestellt wird’ (Von Mises, 1928, p. 21, my emphasis).

214 In Poisson’s own words: ‘The phenomena of any kind are subject to a general law, which one can call the Law of Large Numbers. It consists in the fact, that, if one observes very large numbers of phenomena of the same kind depending on constant or irregularly changeable causes, however not progressively changeable, but one moment in the one sense, the other moment in the other sense; one finds ratios of these numbers which are almost constant’ (Poisson, 1837, p. 7).

215 It may be remarked that in the second German edition Von Mises added an additional sentence in which he expressed his agreement with Poisson with regard to the empirical law of large numbers being the basis of probability theory (see Von Mises, 1981 [1951], p. 22).
Von Mises recognized the law of large numbers, which says that the ratios of numbers derived from the observation of a large number of similar events remain practically constant – or, for that matter, that relative frequencies of events eventually stabilize or become constant – not only as the foundational notion for a rational mathematical theory of probability, but also as being ‘identical with’ (‘vollig gleichbedeutend mit’) the first axiom (L) of its ‘axiomatics’ (collective): it is the empirical basis of the definition of probability as the limiting relative frequency that is made mathematically precise by axiom (L). Where Poisson himself had considered the mathematical proposition or theorem (‘Poisson’s law’) as a confirmation of the empirical law, and others had ‘wavered between two positions: the [first] is alleged either to imply [the latter] or to contradict it’, Von Mises wanted to show that the two ‘laws’ are unconnected.


218 In modern terminology, Poisson ‘proved’ his two-sided law of large numbers (as the approximate stability of arithmetic means or relative frequencies) by means of ‘a special two-stage model of causation for the occurrence of an event [and] he established two auxiliary theorems on stochastic convergence: the first concerning the arithmetic means of non-identically distributed random variables, the second concerning relative frequencies of an event which generally does not occur with constant probability. He based these theorems, which are equivalent to the now so-called ‘laws of large numbers’, on his general [central limit theorem]’ (Fischer, 2011, p. 36).

219 For example, Irénée-Jules Bienaymé (1796-1879) and Chebyshev were among the critics of Poisson’s work on the law of large numbers.

It is true that Poisson employed the term ‘law of large numbers’ for the approximate stability of both relative frequencies (of events) as well as arithmetic means (of so-called ‘non-identically distributed random variables’). Von Mises wrote that the first ‘part’ says that ‘if a certain result occurs \(m\) times in \(n\) trials, we call \(m/n\) its ‘relative frequency’ and the second (Bernoul- lian) ‘part’, in a generalized form found in Poisson and Chebyshev, that ‘if an experiment, whose results are simple alternatives with the probability \(p\) for the positive result, is repeated \(n\) times, and if \(\varepsilon\) is an arbitrary small number, the probability that the number of positive results will be not smaller than \(n(p - \varepsilon)\), and not larger than \(n(p + \varepsilon)\), tends to 1 as \(n\) tends to infinity’. The difference is that if the empirical law is a generalization from empirical results expressed in terms of ratios interpretable as relative frequencies, the ‘Bernoulli-Poisson-Chebyshev’ proposition (henceforth, ‘BPC-theorem’) is an arithmetical statement about the properties and proportion of certain numbers which, in classical probability theory, is expressed in terms of probabilities. For example, the fact that in the case of the set of natural numbers represented by 0’s and 1’s the concentration of their frequency around the value \(\frac{1}{2}\) becomes increasingly more pronounced

‘is expressed […] by saying: In the first sequence the ‘probability’ of the results 49 to 51 zeros is [e.g.] 0.16; in the second sequence the ‘probability’ of the results 490 to 510 is [e.g.] 0.47 [etc.]. By assuming \(\varepsilon = 0.01\) and \(p = \frac{1}{2}\), the theorem [is]: Let us write down […] all \(2^n\) numbers which can be written by means of 0’s and 1’s containing up to \(n\) figures. The

---

221 This much becomes clear, for instance, in his ‘proof’ of his law of large numbers in Poisson, 1837, pp. 139-143.
222 ‘Tritt ein Ereignis in \(n\) Versuchen \(m\)-mal ein, so nennen wir den Quotienten \(m/n\) die ‘relative Häufigkeit’ (Von Mises, 1928, p. 79).
223 Von Mises referred to ‘proposition II’ of Jakob Bernoulli’s (1654-1705) Aris Conjectandi post-humously published in 1713 (Bernoulli, 1713 [2006]).
224 ‘Wenn […] einen einfachen Alternativ-Versuch, dessen positives Ergebnis die Wahr- scheinlichkeit \(p\) besitzt, \(n\)-mal wiederholt (wird) und mit \(\varepsilon\) eine beliebig kleine Zahl bezeichnet, so geht die Wahrscheinlichkeit dafür, dass der Versuch mindestens \((pn - \varepsilon n)\)-mal und höchstens \((pn + \varepsilon n)\)-mal positive ausfällt, mit wachsendem \(n\) gegen Eins. Konkreter’ (Von Mises, 1928, p. 86).
proportion of numbers containing from $0.49n$ to $0.51n$ zeros increases steadily as $n$ grows'.

Von Mises’s problem with this statement was that if, in so far as its definition of probability merely says something about the proportion between the number of favourable and unfavourable cases, it cannot reach a conclusion as to the actual result or end (‘Ablauf’) of a sequence, it has nothing to do with the distribution of 1’s and 0’s resulting from such sequences of observed empirical events. The reason that Poisson considered the mathematical theorem as a confirmation of the empirical law was that he employed a different meaning for probability at the beginning and at the end of the calculation: ‘At the beginning […] the probability $1/2$ of ‘heads’ shall be the ratio of favourable cases to that of all equally possible cases, but the probability ‘nearly 1’ that follows from the calculation […] shall mean that the corresponding event, the occurrence of between $0.49n$ and $0.51n$ heads in a series of $n$ throws, is always, and in any series of trials, observable’. Von Mises drew several conclusions from this observation. Firstly, the (‘deus ex machina’) auxiliary hypothesis reconciling the classical definition with Poisson’s conclusion (‘If a calculation gives a value not very different from 1 for the probability of an event, then this event takes place in nearly all repetitions of the corresponding experiment’) is nothing

---

226 ‘Es wird […] beschrieben, dass man sagt: die “Wahrscheinlichkeit” der Ergebniszahlen 49 bis 51 sei im ersten Fall 0,16, die analoge im zweiten Fall 0.47 [usw.]. Der Inhalt des Theorems […] lasst sich, wenn $E = 0.01$ wahlen, wie folgt aussprechen: Schreibt man alle aus Nullen und Einsen bestehenden Zahlen der Größe nach geordnet bis einschließlich der $n$-stelligen auf, so bilden diejenigen unter ihnen, bei denen die Anzahl der Einser mindestens $0.49n$ und höchstens $0.51n$ beträgt, eine mit wachsendem $n$ immer stärker werdende Majorität’ (Von Mises, 1928, p. 83, my emphasis).

227 Von Mises, 1928, p. 83.


229 ‘Die Wahrscheinlichkeit $1/2$ eines “Kopf”-Wurfes, die in die Rechnung eingeht, soll nur der Quotient der “günstigen” durch die “gleichmöglichen” Fälle sein, aber die Wahrscheinlichkeit nahe 1, die aus der Rechnung hervorgeht […] die soll bedeuten, dass das betreffende Ereignis, nämlich das Auftreten von $0.49n$ bis 0.51n “Kopf”-Würfen in einer Serie von $n$ Versuchen, fast immer, bei fast jedem Serienversuch, zu beobachten ist’ (Von Mises, 1928, p. 84).

230 ‘Sobald eine Rechnung für ein Ereignis einen Wahrscheinlichkeitswert, der nur wenig kleiner als 1 ist, ergeben hat, tritt dieses Ereignis bei fortgesetzten Versuchen fast jedesmal ein’ (Von Mises, 1928, pp. 84-85).
but ‘the frequency definition in a restricted form’\textsuperscript{231} and ‘adds nothing to what [is] assumed’\textsuperscript{232} in the case of the empirical law. More precisely, given that that the assumption of the existence of a limiting value for relative frequency, that is, of the validity of the empirical law is implied in the definition of probability as limiting relative frequency ‘what could be the point of obtaining, by tiresome calculation (‘\textit{durch mühsame Rechnung}’), a theorem that merely asserts what has already been assumed?’. The answer to this question is that, secondly, ‘the proposition, which arises as the result of a mathematical deduction from [BPC-theorem], implies [...] much more than the mere existence of a limiting value’\textsuperscript{233}, namely a ‘definite statement’ (‘\textit{bestimmte Aussage}’) about the result (‘\textit{Ablauf}’\textsuperscript{234}) of events (e.g. ‘heads’ and ‘tails’) in an indefinitely long sequence. This suggests, thirdly, that if ‘Bernoulli-‘ or ‘Poisson-‘sequences are sequences that can be described by a mathematical formula or law in so far as their limiting values ‘belong to higher mathematics’\textsuperscript{235} such that they do not satisfy the requirement of complete ‘lawlessness’ or ‘randomness’ (‘\textit{Regellösigkeit}’),\textsuperscript{236} it ‘does not make sense’ (‘\textit{es hat wenig Sinn}’\textsuperscript{237}) to speak of probability proper in the case of the Bernoulli- or Poisson-theorem. For example, Von Mises dismissed as a ‘mistake’ (‘\textit{Verirrung}’\textsuperscript{238}) Evgeny Evgenievich Slutskii’s (1880-1948) St.-Petersburgian\textsuperscript{239} attempt to combine (‘\textit{zusammenzubrauen}’\textsuperscript{240}) the existence of limit values and the BPC-theorem into a theorem about the ‘(random) stochas-
tic limit’ or ‘asymptote’ (‘Slutskii’s theorem’). Similar to the ‘mistakes’ of Hermann Weyl (1885-1955) and Georg Pólya (1887-1985), Slutskii allegedly assumed for his mathematical theorem ‘the classical definition of probability, which has nothing to do with the [actual] course of phenomena’, and then (‘nachträglich’) employed ‘a way of phrasing that does concern this course (‘it is to be expected with necessity [‘mit Sicherheit’], that...’).

Von Mises emphasized that once the empirical notion of limiting relative frequency is adopted as the (axiomatic) starting point for probability theory it is possible to maintain that mathematical theorems related to the (‘direct’) BPC-theorem (‘First Law of Large Numbers) and the (‘indirect’ or ‘inverse’) Bayesian-theorem (‘Second Law of Large Numbers’) contain ‘a valuable assertion [‘Aussage’] providing ‘some information’ (‘einen Aufschluß’) for the prediction of frequencies from given probabilities and the estimation of probabilities from observed frequencies, respectively. The reason for this was that ‘a proposition that is to assert [‘aussagen’] something about reality is only derivable when one puts at the beginning of the derivation certain well-defined [‘bestimmte [...] sogenannte’] axioms [(L) and (IR)], [these being] initial prop-

---

241 See Von Mises, 1928, p. 89.
242 See below (‘An afterword’).
243 ‘[D]er klassischen Wahrscheinlichkeitsdefinition, die nichts mit dem Erscheinungsablauf zu tun hat’ (Von Mises, 1928, p. 89).
244 Von Mises, 1928, p. 89.
246 Von Mises, 1928, p. 89.
247 Von Mises, 1928, p. 89.
248 This distinction of probability theory into two parts corresponds to the formulation, in the Fundamentalsätze of 1919, of two ‘fundamental theorems’ that are to ‘master, from a unified and general analytical viewpoint’ (‘von einem einheitlichen und allgemeinen analytischen Gesichtspunkt aus zu bewältigen’) (Von Mises, 1919a, p. 2) several individual problems of the theory; on the one hand, ‘direct’ problems (‘the Bernoullian and Poissonian problem and its generalization, the Gaussian error law, the law of large number’ (‘das Bernoullische und Poissonsche Problem und seine Verallgemeinerung, das Gaussche Fehlergesetz, das Gesetz der grossen Zahl’)) (Von Mises, 1919a, p. 2) and, on the other hand, ‘indirect’ problems (‘the Bayesian and Laplace-Bienaymean problem and its generalization, a second proposition of error theory and a second law of large number’ (‘das Bayessche und Laplace-Bienaymesche Problem und seine Verallgemeinerung, einen zweiten Satz der Fehlertheorie und ein zweites Gesetz der grossen Zahl’)) (Von Mises, 1919a, p. 2).
ositions ['Ausgangssätze'] obtained from experience'.\(^\text{249}\) Von Mises strongly denied the idea of the separation into an (applied) extra-mathematical part and a (pure) mathematical part of probability theory;\(^\text{250}\) where the two axioms for collectives ‘are a necessary consequence of the frequency [definition], in the sense that if one interprets probability as limiting relative frequency then [series] will exhibit [collective-]like properties’\(^\text{251}\) the mathematical theorems for probability only make sense, probabilistically, as mathematical deductions from these axioms.

An afterword: Von Mises and Cournot’s principle: the status of the law of large numbers

The ‘lecture’ on the laws of large numbers introduced as a counter-example (‘Gegenbeispiel’) against Poisson’s identification of the empirical law of large numbers (or the axiom (L)) with the BPC-theorem a certain non-random sequence of 1’s and 0’s – with ‘1’ standing for the (‘positive’) result that either 5, 6, 7, 8 or 9 is the sixth figure after the decimal point resulting from the calculation of the square root of positive integers and ‘0’ standing for the (‘negative’) result that this figure is either 0, 1, 2, 3 or 4. Von Mises showed that the beginning of this table confirms the BPC-theorem, in that ‘every series of \(n\) numbers consist of about [50\%] 0’s and about [50\%] 1’s if one chooses a sufficiently large \([n]\),\(^\text{252}\) but that this is no longer the case ‘when the table [of square roots] is imagined as proceeding beyond the actually printed [‘tatsächlich Gedruckte’].\(^\text{253}\) He then concluded the following: on the one hand, there are mathematical

\(^{249}\) ‘[E]in Satz, der etwas über die Wirklichkeit aussagen soll, nur dann [...] ableitbar ist, wenn man an die Spitze der Ableitung bestimmte, der Erfahrung entnommene Ausgangssätze, sogenannte Axiome stellt’ (Von Mises, 1928, p. 90).

\(^{250}\) For example, in a letter to Pólya of December 12, 1919 Von Mises remarked that ‘[m]y decided tendency – which cannot be separated into a mathematical and an extra-mathematical – is precisely to put an end to the current situation, where a presentation of the theory of probability begins with the words [...] that one does not know really what probability is’ (Von Mises quoted in Siegmund-Schultze, 2006, p. 472).


\(^{252}\) ‘[J]ede Serie von \(n\) Zifffern [besteht] ungefähr zur Hälfte aus Nullen und ungefähr zur Hälfte aus Einsern [...] wenn man [...] eine genugend große [...] \(n\) wählt’ (Von Mises, 1928, p. 87).

\(^{253}\) ‘[W]enn man sich die Tafel über das tatsächlich Gedruckte hinaus fortgesetzt denkt’ (Von Mises, 1928, p. 87).
'non-Bernoullian' and non-random) sequences with relative frequencies tending toward limiting values which do not obey the BPC-theorem. On the other hand, there are ('Bernoullian') sequences, with an initial randomness, obeying the BPC-theorem which can also be defined by mathematical formulae. Taken together, when the improvable ('nicht beweisbare') empirical generalization of the existence of limits is assumed as an axiom (L), a theorem like the BPC-theorem can be derived, in a logically valid way, from the frequency definition of probability as a special case of the additional assumption ('unter Hinzunahme') of randomness (IR).

Von Mises introduced the counter-example and the conclusion drawn from it in order to expose the uncritical ('kritiklos') treatment of probability theory by mathematicians such as Pólya, Weyl and Slutskii. His main criticism was that these all of them remained committed, in a surprisingly unreflective way, to the classical fallacy of employing the BPC-theorem to make statements about empirical reality from the notion of 'equiprobability' – an 'a priori' and/or 'subjective' notion which assumes uniform distributions of 'non-geomet-

---

254 Von Mises, 1928, p. 100.
255 That is to say, without changing the meaning of probability in the course of the calculation.
256 Von Mises, 1928, p. 100.
257 Von Mises wrote that the BPC-theorem 'turns out to be a special consequence of the complete randomness of a sequence' ('erweist sich eben als eine spezielle Konsequenz aus der vollständigen Regellosigkeit einer Folge') (Von Mises, 1928, p. 90).
258 It could thus be said that if the notion of randomness follows from the frequency definition of probability it does so in so far as it is to protect probability theory against the abovementioned 'non-Bernoullian' sequences.
259 Von Mises, 1928, p. 88.
260 See Von Mises, 1928, pp. 68-70. Von Mises explained the idea of the a priori establishment of 'equally likely cases' as an attempt to derive knowledge about probabilities from the physical characteristics, that is, the homogeneity and/or symmetry, of, for example, a die.
261 See Von Mises, 1928, pp. 72-74. Von Mises explained the idea of the 'subjective' establishment of 'equally likely cases' as an attempt to derive something from nothing or, for that matter, knowledge from ignorance on the basis of the 'principle of indifference'. This principle holds that if there is no known reason to assign differing probabilities to two events, they are to be assigned the same probability.
equal probabilities in the original collectives to the exclusion of all (applications with) other distributions and probabilities. Although Antoine-Augustin Cournot (1801-1877), Johannes von Kries (1853-1928), Ladislaus von Bortkiewicz (né Vladislav Bortkewitsch) (1868-1931), Bruns and Czuber, among others probabilists, had criticized the Laplacian definition of probability on philosophical and practical grounds, none of them had fundamentally moved beyond it (‘kein Author erhebt sich wesentlich über [...]’). But given that, for instance, Slutskii had explicitly dismissed this very definition as a basis for the mathematization of probability in his assessment of Markov’s 1900 textbook, it may be said that Von Mises’s criticism concerned not solely the early-twentieth-century renovation of classical probability theory (e.g. incorporation of geometric probability, introduction as theorems of total, compound and relative probability), but also, more specifically, ‘Cournot’s principle’.

The history of this principle can be traced back to Bernoulli’s 1713 treatment of events with very small probability being ‘morally impossible’ (i.e. they will not happen) and events with very high probability being ‘morally certain’ (i.e.}

---

262 The negative term ‘non-geometric’ is to suggest that the assumption of ‘equally likely’ cum ‘equally possible’ events in the initial collectives, that is the assumption of a uniform distribution cannot be upheld in the case of geometric probabilities where the geometric extension (area, volume) of cases is measured.

263 The neo-Kantian philosopher Kries had developed a so-called ‘logico-objective’ definition of probability in his *Principien der Wahrscheinlichkeitsrechnung* of 1886 – and this with the explicit aim of eliminating the ‘subjective’ elements from the notion of probability. Kries’s ‘*Spielraum*’ or ‘range’ theory of probability can be explained, in modern terminology, as follows: ‘Take the disjunctive set $E$ of ontological conditions $E_1, E_2, [...] E_n$ which constitute the whole *Spielraum* of a state of affairs $H$ relative to certain [‘nomological’] laws $L$. Together with this set of laws some of these alternatives imply the occurrence of $H$, and the remaining ones imply the occurrence of not-$H$. [T]he *Spielraum* is [thus] divided into two parts [...] If we call the first part [...] favorable to $H$, we can say that the probability of $H$ relative to $E$ and $L$ is the ratio of the number of favorable alternatives of the *Spielraum* to the number of all alternatives’ (Heidelberger, 2001, pp. 38-39). Given this presentation and the attempt to determine the alternatives via ‘equipossibility’, Von Mises felt able to dismiss Kries’s theory as a specific version of classical probability theory (see Von Mises, 1928, pp. 74-75).

264 Von Mises, 1919b, p. 52.
they will happen). Cournot, who ‘gave the discussion a nineteenth century cast’, suggested that it is the fact that events with vanishingly (infinitely) small probability are physically impossible that connects mathematical probability to the world and gives it ‘substance’ or empirical meaning. The principle entered the Russian literature via the work on mathematical statistics of Markov, the disciple, together with Lyapunov, of Chebyshev in St. Petersburg and Chuprov, who was the professor of political economy and statistics at Moscow University (where he had graduated, in 1896, with a dissertation on probability as a basis for theoretical statistics under the (partial) supervision of Nekrasov) before becoming the director of the Economics Section of the St. Petersburg Polytechnic Institute in 1902. It would be adopted in its so-called ‘strong form’ (with the meaning, that is, the connection with the empirical world of application, of mathematical probability being found in ‘the non-happening of small [or zero] probability events singled out in advance’ by, among others, Maurice René Fréchet (1878-1973), Paul Lévy (1886-1971), Borel, Khinchin, Kolmogorov and Slutskii. For them, Cournot’s principle combined with the BPC-theorem ‘to produce the unequivocal conclusion that an event’s probability will be approximated by its frequency in [a] sufficiently

265 In his *Ars Conjectandi*, Bernoulli related mathematical probability to moral certainty – or, for that matter, quantified moral certainty by writing that ‘something is morally certain if its probability is so close to certainty that the shortfall is imperceptible’ and ‘something is morally impossible if its probability is no more than the amount by which moral certainty falls short of complete certainty’. He added that it ‘would be useful, accordingly, if definite limits for moral certainty were established by authority of the magistracy. For instance, it might be determined whether 99/100 of certainty suffices or whether 999/1000 is required’ (Bernoulli, 1713 [2006], p. 321).

266 Schafer & Vovk, 2006, p. 72.

267 In his *Exposition de la théorie des chances et des probabilités*, Cournot wrote the following: ‘The physical impossible event is [...] the one that has infinitely small probability, and only this remark gives substance – objective and phenomenal value [in the Kantian sense] – to the theory of mathematical probability’ (‘L’événement physiquement impossible est donc celui dont la probabilité mathématique est infiniment petite; et cette seule Remarque donne une consistence, une valeur objective et phénoménale à la théorie de la probabilité mathématique’) (Cournot, 1843, p. 78). See Martin, 1996, chapter 5 for a detailed account of Cournot’s views on the establishment of ‘objective values’ for probabilities. See Schafer, 2005, pp. 1-12; Schafer, 2006; Shafer & Vovk, 2006, pp. 74-76 for a brief history of ‘Cournot’s principle’.

268 It is by now well known that it was the (happenstance) correspondence between Markov and Chuprov that initiated the coming together into mathematical statistics of probability (Markov) and statistics (Chuprov) (see Sheynin, 2006 [2011]).

269 Schafer & Vovk, 2006, p. 74, my emphasis.
long sequence of [trials]'. The ‘strong form’ could be upheld either with (e.g. Lévy) or without (e.g. Kolmogorov, Slutskii) the principle of equally likely case as the foundation of the mathematics for probability, but did not complement any of the axiomatizations prior to Kolmogorov – be this the (‘abstract’) measure-theoretic system of axioms of, among others, Rudolf Laemmel (1879-1962), Ugo Broggi (1880-1965) and Antoni Lomnicki (1881-1941) or the (‘non-abstract’) system of axioms of Von Mises for (idealized) limiting frequencies or of Sergei Bernshtein’s (1880-1968) and Bruno de Finetti’s (1906-1985) system of axioms for qualitative probabilities. The Russian ‘theoretical statistician’ Chuprov, thereby reflecting the skepticism of the German probabilists (Von Kries, Czuber, Bortkiewicz) vis-à-vis the identification of zero probability (‘sets with measure zero’) with impossibility by German mathematicians like Hausdorff and Felix Bernstein (1878-1956), referred to the ‘weak’ form of Cournot’s principle in his work. This ‘weak form’ (with events with very small probabilities happening very rarely in repeated trials) combined with the BPC-theorem ‘to produce the conclusion that an event’s probability will usually be approximated by its frequency in a sufficiently long sequence of trials’.

---

270 Shafer & Vovk, 2006, p. 74, my emphasis.
271 Bernstein, whose studies at Gottingen (in 1902-1903) had been supervised by Hilbert and whose doctoral dissertation for the Sorbonne had been examined by Jacques Hadamard (1865-1963) and Poincaré, published a long paper on the axiomatization of (qualitative) probability as early as in 1917. Von Plato remarks that ‘[h]ardly anyone outside Russia and the Soviet Union can be expected to have read the article [which appeared in the proceedings of the mathematical society of Kharkov, Ukraine]. In his Grundbegriffe, Kolmogorov mentioned it as giving an axiomatization with a set of basic concepts different from his (see Kolmogorov, 1933, p. 2, f. 1).
272 Chuprov regularly used the term ‘theoretical statistics’, ‘obviously not favoring the expression ‘mathematical statistics’ (Sheynin, 2006 [2011], p. 101).
273 See, for example, Chuprov, 1905, p. 443.
274 Chuprov defined the ‘weak form’ of Cournot’s principle as follows: ‘Events whose possibility is very small, occur extremely seldom’ (Chuprov quoted and translated into English in Sheynin, 2006 [2011], p. 95).
sequence of [...] trials’. What is important is that Chuprov, not unlike other Russian statisticians, considered the intuitive version of the ‘weak law of large numbers’ (convergence to a limit) as an independent logical, rather than a mathematical principle that could be proved in a non-mathematical way from Cournot’s ‘general notion’ and that somehow grounded both the theorem of Bernoulli and the (generalized) theorem of Poisson. Chuprov argued that the ‘search for [any] grounds for [Poisson’s and Chebyshev’s generalization of] the [‘weak’] law of large numbers in the Bernoulli theorem [‘the weak law of large numbers’] is always a petitio principii because if the Bernoulli itself ‘says nothing about concrete phenomena’ it cannot function as the basis of its own generalization. In other words, Chuprov ‘understood that the ordinary law of large numbers [‘Bernoulli’s theorem’] was not sufficient for practical purposes’ and he believed that the connection between (‘the abstract notion of’) probability and (the) frequency (of ‘concrete phenomena’) ‘should come from logic rather than mathematics’.


276 Chuprov arrived at this conclusion as follows: ‘When certain general conditions characterize a very long series of events [of trials] all the phenomena [the events] possible under these conditions occur roughly in proportion to the possibilities of their occurrence. This proposition can be proved in the following way: 1) Events whose possibility is very small, occur extremely seldom, 2) The possibility that in a long series of trials the possible phenomena occur in a number of times not essentially proportional to the respect. possibilities is very small. Ergo, 3) It is extremely seldom for the phenomena to occur in long series of trials in a number of trials not roughly proport. to their possibilities. Proposition 1) is derived from general notions (Cournot), proposition 2) is ein Theorem der Kombinationslehre (Kries). Proposition 3) can be obtained from 2) [...] only by adding i). Therefore, proposition 3) cannot be regarded as a purely mathematical theorem. According to the opinion of [Czuber] proposition 3) can be arrived at from the main notions without mathematical derivations in the same way as proposition 1)’ (Czuber quoted and translated into English in Sheynin, 2006 [2011], p. 95).


279 Sheynin, 1993, p. 248.


Taken together, in the pre-Grundbegriffe era the ‘strong form’ of Cournot’s principle gave expression to the ‘non-modern’ emphasis on extra-mathematical (‘physical’) meaning by those mathematical probabilists working on the analytical subfields whose integration would mark the establishment of modern probability as a branch of modern mathematics (axiomatics (measure theory), strong laws of large numbers, stochastic processes, limit theorems for distributions of sums of random variables). When compared to the proponents of the ‘weak form’ who, at any rate, introduced the principle mainly as a way to account for probability theory’s practical value without having recourse to the ‘weak law of large numbers’ (‘Von Mises’s BPC-theorem), the ‘counter-modernism’ of the proponents of the ‘strong form’ is to be found, for instance, in their acceptance of the mathematical ‘strong law of large numbers’ as the basis of applicability of the ‘weak law’. The non-conformist Von Mises contributed to the ‘modernization’ of probability theory (axiomatization, sums of random variables etc.), but, indeed, ‘tended to be ‘anti-modern’ even in those of his activities which belonged to ‘modernism’.

For example, Von Mises could be said to have shown his awareness of the future role of sets of probability zero (‘0’) within an axiomatized probability theory when he wrote that $W_A = 0$ is not always equivalent to impossibility, $W_A = 1$ not always to ‘certainty’. But the fact that he justified this principle with reference to the practical import of

---

283 For example Lévy argued that the notion of equally likely events ‘suffices as a foundation for the mathematics of probability, but so long as we base our reasoning only on this notion, our probabilities are merely subjective’ (Schafer & Vovk, 2006, p. 73, see also Lévy, 1925, p. 21, p. 34).


285 See, for instance, Khinchin, 1928.

Gauss’s error theory suggests that Von Mises did not introduce it in order to be able to interpret or to account for the applicability of mathematical collectives, but to re-emphasize that mathematical collectives (and their infinite sequences) are to express all empirically generalized statistical properties of empirical collectives (and their finite sequences) and that the only criterion for accepting properties of the former was their use in solving problems in the case of the latter. There was no place for Cournot’s ‘Bernoulliesque’ principle in the Von Misean axiomatic framework because the very problem of bridging either the ‘crack’ between ‘frequency’ and ‘probability’ or the (‘modern’) gap between ‘real world’ and ‘mathematical’ probability was absent from it. The fundamental reason for which was, of course, that limiting relative frequencies in the Kollektivs did not owe their existence to the (mathematical) law of large numbers and did not obtain their (empirical) meaning through interpretation by means of a certain principle.

287 Reflecting on the principle (see footnote 578), Von Mises wrote that ‘[t]his seemingly not unimportant deviation from the ordinary ['classical'] theory is indeed quite necessary [...] if one considers any kind of continuous ['geometrical'] distribution. For instance in the Gaussian error theory each error has probability zero, but each error is possible’ (‘Diese scheinbar nicht unwesentliche Abweichung von der üblichen Theorie ist tatsächlich ganz unerlässlich [...] wenn man irgendwelche Fälle stetiger Verteilungen in Betracht zieht. Z.B. hat in der Gaußschen Fehlertheorie jede Fehlergröße die Wahrscheinlichkeit null, aber jede Fehlergröße ist möglich’) (Von Mises, 1919b, p. 56, f. 8).

288 For detailed explanations of these points see Van Lambalgen, 1996, pp. 363-364.

289 It may here be insightful to remark that neither the proponents of the ‘strong form’ of Cournot’s principle nor those of the ‘weak form’ supported Von Mises’s probability – with Slutskii, on the one hand, and Chuprov, on the other hand, being cases in point.

Acknowledgments

I would like to thank Annette Vogt for sharing with me some of her articles on Richard von Mises and her vast knowledge of the history of German mathematics, statistics and probability theory.

References


