

The Maxim of Probabilism – with special regard to Reichenbach

Miklós Rédei* Zalán Gyenis†

Forthcoming in *Synthese*

Abstract

It is shown that by realizing the isomorphism features of the frequency and geometric interpretations of probability, Reichenbach comes very close to the idea of identifying mathematical probability theory with measure theory in his 1949 work on foundations of probability. Some general features of Reichenbach's axiomatization of probability theory are pointed out as likely obstacles that prevented him making this conceptual move. The role of isomorphisms of Kolmogorovian probability measure spaces is specified in what we call the “Maxim of Probabilism”, which states that a necessary condition for a concept to be probabilistic is its invariance with respect to measure-theoretic isomorphisms. The functioning of the Maxim of Probabilism is illustrated by the example of conditioning via conditional expectations.

1 Introductory comments

Probability theory featured prominently in Reichenbach's thinking and work throughout his whole career, both as a conceptual tool and as a subject of philosophical investigation: Already in his doctoral thesis (1915) (published in English translation in 2008 [33], see [23] for a compact review of this translation) probability takes center stage in the form of a “principle of lawful distribution”. This principle states, roughly, that the empirical relative frequencies of occurrences of events converge to a true limit to be understood as probability. This principle has a transcendental status in Reichenbach's theory of knowledge, similar to Kant's principle of causality in Kant's epistemology. The principle is transcendental because it is not empirically testable – rather it forms the basis of empirical science [9].

In his subsequent works, Reichenbach both applied probability theory in the analysis of specific philosophical problems and investigated the foundations of probability theory itself. An example of the former is Reichenbach's concept of common cause used in his *The Direction of Time* [32]. This notion has had a lasting impact on the analysis of causality and has been the conceptual predecessor of the Causal Markov Condition in the modern theory of Bayes nets (see [16] for a review, and [17], [42] for detailed analyses of Reichenbach's notion and the related principle of the common cause).

Reichenbach's foundational work on probability theory culminated in the substantial, almost 500 page long monograph published in 1949 [31], which is a re-worked version of the one published in German in 1935 [30]. In his 1949 work, Reichenbach gave both a formal axiomatization of probability theory and attempted to provide a foundation for it in the sense of the frequency view of probabilities. Both ideas rely on his earlier work; in particular the axiomatization in [31] is based on the paper published in 1932 [29].

The general assessment [9], [8] of Reichenbach's axiomatization is that now it only has historical significance because Kolmogorov's axiomatization published in 1933 [20] overshadowed it and became the mainstream. This verdict is based in [8] on a detailed critical analysis of Reichenbach's axiomatization and of his related concept of probability logic. We agree with this assessment. In Section 2 we recall some general features of Reichenbach's axiomatization and provide some more critical comments. On the positive side, in Section 2, we also show that Reichenbach's analysis contains an important idea that in principle opens up the road to an axiomatization in the sense of Kolmogorov:

*Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, m.redei@lse.ac.uk.

†Department of Logic, Institute of Philosophy, Jagiellonian University, Cracow, Poland

The idea is to regard as mathematical probability theory what is *isomorphic* in different interpretations. But this avenue remains unexplored in Reichenbach's work, which we claim is mainly due to ambiguities in the Reichenbachian axiomatic system. In Sections 3 and 4 we explore the role of isomorphism from the perspective of foundations of probability theory. While it is not clear what the notion of isomorphism in the Reichenbachian axiomatization would be, there are very natural notions of isomorphism in the Kolmogorovian axiomatization. Kolmogorov himself did not make use of them in his foundational book, but they became standard. In Section 3 we recall these notions of isomorphisms and formulate what we call the **Maxim of Probabilism**: The idea that a concept, reasoning, argument is probabilistic only if it is invariant with respect to the isomorphisms of the mathematical structures that are models of the axioms. In sections 4 and 5 we illustrate the usefulness of the **Maxim of Probabilism** by using it to clarify some neuralgic points in connection with conditioning with respect to probability zero events, in particular in the context of the Borel-Kolmogorov Paradox.

2 Comments on Reichenbach's axiomatization of probability theory

Reichenbach distinguished three approaches to axiomatization of probability theory: One that aims at an

“[...] interpreted form of axiomatic construction [...] which regards probability, from the very beginning, as a frequency, and derives from this interpretation, by the possible inclusion of additional postulates, the rules of the theory.” [31][p. 121]

The second

“[...] formal conception introduces the concept of probability by the method of implicit definitions, and uses no properties of the concept other than those expressed in a set of formal relations placed as axioms in the beginning of the theory, leaving open various possibilities for its interpretation.” [31][p. 121]

The third approach

“[...] connects the treatment of probability with the methods of symbolic logic. [...] constructing probability as a relation between statements, which includes logical implication as a special case.” [31][p. 122]

Reichenbach classifies both Kolmogorov's approach and his 1932 axiomatization as belonging to the second group. While this is certainly true for the Kolmogorov approach, it is not unambiguously true for his 1932 axiomatization: While a formal axiomatization in the sense of the second approach is given in his 1932 axiomatization, Reichenbach also introduces a “coordinating definition α ” [p. 591] that relates the formal probabilistic formula to limits of relative frequencies. On this basis he distinguishes two notions of *mathematical* probability:

“We call the resulting notion of probability, i.e. the concept that is determined by the axiom system including the coordinating definition α , the *contentual mathematical concept of probability*, in contrast to the *formal mathematical concept of probability* determined exclusively by the axioms, i.e. without assigning content to it.”¹ [29][p. 592] (emphasis in original)

Thus, although a purely formal axiomatization describing the *formal mathematical concept of probability* is part of Reichenbach's treatment indeed, Reichenbach's analysis is coupled to a frequency view even when it comes to a *mathematical* specification of the concept. There is thus no sharp separation in the *mathematical sphere* of the concept of probability from the frequency view in his 1932 axiomatization – in contrast to what he claims about his own axiomatization. This ambiguity prohibits Reichenbach, we claim below, to develop an idea that potentially leads to an axiomatization based on measure theory.

Furthermore, in his [29], Reichenbach sees the need for a further *axiom* to be added to the mathematical axioms: This is the *Axiom of Induction* [29][p. 614]. This axiom is precisely the

¹Our translation. Original German: “Wir nennen den dabei entstehenden Wahrscheinlichkeitsbegriff, also den durch das Axiomensystem einschließlich der Zuordnungsdefinition α bestimmten Begriff, den *inhaltlichen mathematischen Wahrscheinlichkeitsbegriff*, im Gegensatz zu dem *formalen mathematischen Wahrscheinlichkeitsbegriff*, der durch das Axiomensystem allein definiert wird, also ohne eine inhaltliche Belegung.”

principle of lawful distribution that appeared in his 1915 dissertation – as Reichenbach explicitly acknowledges, citing his PhD dissertation ([29][p. 614], especially footnote 24 in [31]). This axiom has a status that is conceptually different from those that specify the mathematical probability because it is not part of mathematics: it postulates the applicability of probability theory in the sense of the frequency view.

Coupling probability theory to the frequency interpretation, Reichenbach follows a well-established tradition, of which he is fully aware:

“In order to develop the frequency interpretation, we define probability as the *limit of a frequency* within an infinite sequence. The definition follows a path that was pointed out by S.D. Poisson in 1837. In 1854 it was used by George Boole, and in recent times it was brought to fore by Richard von Mises, who defended it successfully against critical objections.” [31][p. 68] (emphasis in original)

The key difference between Reichenbach’s frequentism and von Mises’ concept of probability as limit of relative frequency is that Reichenbach abandons von Mises’ requirement of randomness of the infinite sequence in which relative frequencies are supposed to be calculated. For von Mises it is not enough that the limits of frequencies in the infinite sequence exist: The infinite sequence, the “ensemble”(von Mises calls it “Kollektiv”) must also be disorderly, “random” [38], [39][p. 23]. Von Mises specified the content of randomness of an ensemble by requiring invariance of the limits of relative frequencies in the ensemble with respect to place selections: Selecting an infinite sub-ensemble of the original ensemble by a rule, the limits of relative frequencies *in the sub-ensemble* should be equal to the limits of the relative frequencies in the original ensemble [39][p. 23]. According to von Mises, this invariance should hold for *any* place selection determined by a rule that does not involve the random event whose frequency one calculates. Given this concept of randomness, the problem of its consistency arises: Do random ensembles exist at all? Von Mises [39][p. 88-89] recalls the reasoning that consistency cannot be proved in the strict sense of mathematical proof: An infinite sequence can only be specified by a mathematical rule, which can in principle be used to select a sub-ensemble in which the frequencies differ from the one in the original sequence. But he rejects the position that one should restrict the class of place selection to a class for which consistency of the corresponding restricted randomness concept is provable, saying that for any conceivable sub-class “... it will be possible to indicate place selections” that are not in the class, and, consequently

“It is not possible to build a theory of probability on the assumption that the limiting values of the relative frequencies should remain unchanged only for a certain group of place selections, predetermined once and for all.” [39][p. 90]

Ultimately, in the chapter “Consistency of the randomness axiom” in [39], von Mises seems to be content with what one could call “pragmatic consistency” of randomness, and which is based on results (due to Copeland and Wald) stating that for any *countable* set of place selections there exist random ensembles:

[...] from what we know so far, it is certain that the probability calculus, founded on the notion of the collective, will not lead to logical inconsistencies, *in any application of the theory known today.*” [39][p. 91] (our emphasis)

In §31 of [31] Reichenbach reviews the main results on the consistency of randomness (especially the works by Copeland, Wald and Ville), but draws a conclusion from them that is different from von Mises’ pragmatic consistency:

“The significance of the problem of the definition of random sequence should not be overestimated, however. Within the general calculus of probability, random sequences merely represent a special type [...] In actual applications, all kinds of probability sequences are encountered. Some show the features of randomness; others represent intermediate types between strictly ordered and random sequences. [...] It would constitute a rather narrow conception of probability if the name of probability sequences would be reserved for random sequences.” [31][p. 150-151]

Reichenbach intends to have his theory to be flexible enough to accommodate such “intermediate types” of infinite sequences, i.e. sequences which embody different degrees of order, not just the random ones:

“An essential feature of my theory of order is that it deals with all possible forms of probability sequences and is not restricted to sequences of one type of order [...]. In

this respect my probability theory differs from others – in particular from that developed by R. von Mises. Such theories regard randomness as an essential characteristic of the very concept of probability; and they contend that the meaning of probability cannot be exhaustively formulated without reference to randomness.” [31][p. 132]

Concerning the classification of his 1949 attempt, Reichenbach writes: “My own presentation undertakes to unite the axiomatic method with the construction of logico-mathematical calculus...” [31][p. 122]. Indeed Reichenbach’s axiomatization is a mixture of formal axiomatization in the sense of symbolic logic and of informal axiomatization in the sense of semi-formal mathematics – as axiomatization is done e.g. when groups are defined by the group axioms. The problem is that in Reichenbach’s treatment the syntax is not completely specified and no formal semantics is given [8][p. 371-373]; and, viewed as axiomatization in the semi-formal sense of mathematics

“[...] these axioms are not sufficient to provide an axiomatization of probability, since they do not ensure that the space the probabilities are applied to is closed under complementation and countable union, i.e. that it forms a sigma-field.” [8][p. 371].

Hence, because of the lack of an explicit semantics and its clear separation from syntax, for a logician, Reichenbach’s axiomatization was too much informal mathematics; for a practicing mathematician, the formal logic involved in the axiomatization separated it too much from mainstream mathematics to be useful; and for a physicist interested in applying probability theory, Reichenbach’s axiomatization was too much logic, mathematics and philosophy altogether. For philosophers the axiomatization offered a target for philosophical criticism, which it received indeed (see section 5 in Eberhardt and Glymour’s paper [8] for a review of the main philosophical criticisms, and [24] for a defense of Reichenbach against a specific objection raised by C.I. Lewis).

But at a certain point Reichenbach comes very close to the idea of identifying mathematical probability theory with measure theory in the spirit of Kolmogorov: In Chapter 6 of [31], Reichenbach discusses an “admissible interpretation” [31][p. 203] of the purely mathematical part of the axioms that is different from the frequency view: The geometrical interpretation. Reichenbach demonstrates in this chapter that (with one exception ! – see below) his axioms are satisfied by subsets of the two dimensional plane with probability identified with the normalized area measure. This idea is present already in the 1932 paper [29][§ 5], but it is more systematically developed and more explicitly stated in [31]:

“The possibility of a geometrical representation of probabilities results from the considerations given in § 40. By showing that both the frequency interpretation and the geometrical interpretation satisfy the axioms of the formal system of probability, that is, are interpretations of this system, we have demonstrated the *isomorphism*, or structural identity, of the two interpretations. Every operation carried out in terms of probability formulas entails analogous operations in the frequency interpretation and the geometrical interpretation. Any derived probability relation is, therefore symbolized in the geometrical interpretation by those geometrical relations that have been specified above for the geometrical interpretation of the probability concept.” [31][p. 207-208] (emphasis in original)

Reichenbach even sees that the isomorphism holds if the two dimensional plane is replaced by a higher dimensional Euclidean space with its Lebesgue measure: “The foregoing considerations can easily be generalized for an attribute space of more than two dimensions.” [31][p. 208]

So in his 1949 monograph Reichenbach is just one intellectual step away from saying that what is common in the frequency and geometrical interpretations is the measure-theoretic structure and the axioms should express exactly this. But this step is not taken and we see several reasons for this. One is that, as Reichenbach himself emphasizes [31][p. 205], one group of axioms (called “group v”, the “Axioms of the theory of order” [31][p. 137]), are *not* satisfied by the two-dimensional Lebesgue measure. This group is precisely the one that connects probability to the frequency view: “. . . the axioms v, like the previous axioms, are valid for all probability sequences, for they could be derived from the frequency interpretation” [31][p. 139]. The two axioms that form this group express that *if* probabilities are limits of relative frequencies in *infinite* sequences then the sequences possess randomness in a limited sense of place selection. Reichenbach is aware that the presence of these axioms distinguishes his axiomatization from those – including Kolmogorov’s – that “. . . omit the development of the theory of the order of the probability sequences.” [31][p. 121] Reichenbach clearly regards this as a virtue of his axiomatization. But the conceptually not entirely sharp separation in Reichenbach’s axiomatization of the frequency interpretation from the purely formal axiomatization becomes a conceptual obstacle to draw the consequences of the isomorphism he recognized.

Another difficulty standing in the way of drawing the consequence of the described isomorphism concerns the unavoidable measure (hence probability) zero sets in the geometrical interpretation: Reichenbach thinks that this would need an additional axiom in probability theory [31][p. 207].

More generally, taking the step of isolating measure theory as the isomorphism-invariant structure presupposes being aware of the development of abstract measure theory, especially of the possibility of moving from the theory of Lebesgue's measure towards abstract measures. Kolmogorov explicitly mentions in the introduction of his book this as a prerequisite for the conceptual move:

“[...] if probability theory was to be based on the above analogies [*involving Lebesgue measure and integral*] it still was necessary to make the theories of measure and integration independent of the geometric elements which were in the foreground with Lebesgue” [20][p. v]

Doob [6] mentions the following crucial steps in the creation of abstract measure theory that were needed for the Kolmogorovian axiomatization:

- Lebesgue's extension of volume in \mathbb{R}^n to the Borel sets in \mathbb{R}^n (1902).
- Radon's definition of a general measure on the Borel sets in \mathbb{R}^n (1913).
- Fréchet's realization that one needs only a σ -algebra of subsets of a set for a meaningful measure theory with a σ -additive measure (1915).

It is perhaps understandable that a mathematician like Kolmogorov was more familiar with these developments in measure theory than the physicist-philosopher Reichenbach. Thus the full ramifications of the isomorphism seen by Reichenbach remain unexplored by him. But the idea of relating what is probability theory to what is isomorphism-invariant is a deep thought. The next section makes this idea explicit in the context of the Kolmogorovian axiomatization.

3 Isomorphism of probability measure spaces and the Maxim of Probabilism

In the Kolmogorovian specification, mathematical probability theory is a *probability measure space* (X, \mathcal{S}, p) , where \mathcal{S} is a Boolean σ -algebra of subsets of the set X (with respect to the standard set theoretical operations \cap, \cup and complement A^\perp), and the probability p is a countably additive map from \mathcal{S} into $[0, 1]$. Accepting this measure-theoretic specification of probability theory leads naturally to both the notion of isomorphism of probability measure spaces and the methodological ramification we call below the **Maxim of Probabilism**.

Since a probability measure space consists of three components, the set of elementary events, a Boolean algebra of general events and a probability measure, the notion of isomorphism is supposed to respect all these three components. Moreover, when it comes to defining isomorphism of probability measure spaces, the possible presence of probability zero events has to be taken into account; accordingly, there are two (inequivalent) notions of isomorphism: (i) *strict isomorphism* (also called *point isomorphism*) and (ii) *isomorphism up to probability zero* (also called *isomorphism mod 0* [2]; the terminology *almost isomorphism* also is used). We define first the isomorphism of measurable spaces:

Definition 3.1. Given two measurable spaces (X, \mathcal{S}) and (Y, \mathcal{Z}) , a bijection $f: X \rightarrow Y$ is called an *isomorphism between (X, \mathcal{S}) and (Y, \mathcal{Z})* , if both f and its inverse f^{-1} are measurable (establishing a Boolean-algebra isomorphism between \mathcal{S} and \mathcal{Z}). In this case (X, \mathcal{S}) and (Y, \mathcal{Z}) are called *isomorphic via f* .

Definition 3.2. Two probability measure spaces (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) are called *measure-theoretically strictly isomorphic* [2][p. 275], if the measurable spaces (X, \mathcal{S}) and (Y, \mathcal{Z}) are isomorphic via some bijection $f: X \rightarrow Y$, and the isomorphism between the Boolean σ -algebras \mathcal{S} and \mathcal{Z} determined by f preserves the probability measures p and q :

$$\begin{aligned} q(B) &= p(f^{-1}[B]) && \text{for all } B \in \mathcal{Z} && (1) \\ p(A) &= q(f[A]) && \text{for all } A \in \mathcal{S} && (2) \end{aligned}$$

For the definition of isomorphism mod0 of probability spaces, we need the following simple notion of *reduction* of probability spaces (X, \mathcal{S}, p) : Let $M \in \mathcal{S}$ be such that $p(M) = 1$. Let \mathcal{S}_M be defined by

$$\mathcal{S}_M \doteq \{A \cap M : A \in \mathcal{S}\} \quad (3)$$

then \mathcal{S}_M is a Boolean σ -algebra, and one can define p on \mathcal{S}_M to obtain a probability measure p_M :

$$p_M(A \cap M) \doteq p(A \cap M) \quad A \in \mathcal{S} \quad (4)$$

(M, \mathcal{S}_M, p_M) is then a probability measure space.

Definition 3.3. Two probability measure spaces (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) are called measure-theoretically *isomorphic mod0*, if there are sets $M \in \mathcal{S}$ and $N \in \mathcal{Z}$ with $p(M) = q(N) = 1$ such that (M, \mathcal{S}_M, p_M) and (N, \mathcal{Z}_N, q_N) are strictly isomorphic. We call the strict isomorphism between (M, \mathcal{S}_M, p_M) and (N, \mathcal{Z}_N, q_N) a *mod0 isomorphism* between (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) .

Embracing the Kolmogorovian specification of probability theory as a triplet (X, \mathcal{S}, p) leads naturally to what we call:

Maxim of Probabilism: *A concept/claim/property/reasoning/argument is probabilistic only if it is invariant with respect to measure-theoretic isomorphisms between probability measure spaces.*

To be more precise, one can distinguish two senses of the adjective “probabilistic” in connection with the **Maxim of Probabilism**: A *weak* and a *strong*, depending on the notion of isomorphism involved:

A concept/claim/property/reasoning/argument is

- *weakly* probabilistic only if it is invariant with respect to *strict* measure-theoretic isomorphisms between probability measure spaces;
- *strongly* probabilistic only if it is invariant with respect to *mod0 isomorphisms*.

The **Maxim of Probabilism** provides a necessary condition for what probability is. In applications of probability theory the set of elementary random events are frequently modeled by a set X in which structures are defined in addition to the σ -field \mathcal{S} (for instance: metric, topological, or order structures). As a consequence, reasonings in the context of (X, \mathcal{S}, p) might involve features of these – from a measure-theoretic viewpoint “surplus” – structures and thus the probabilistic reasonings get intertwined with considerations that are *not* in fact probabilistic. This molding of probabilistic and non-probabilistic elements in reasonings is potentially misleading because it might make invisible where precisely the probabilistic content lies. This can, in turn, lead to misguided questions and puzzles. The **Maxim of Probabilism** can in such situations be used to disambiguate the probabilistic and non-probabilistic components of reasonings and concepts: This **Maxim** tells us that an argument or a concept that is formulated in the context of a probability measure space (X, \mathcal{S}, p) *cannot* be regarded even as weakly probabilistic if it is such that it *cannot* be formulated also in a probability measure space with which (X, \mathcal{S}, p) is isomorphic via a strict isomorphism.

We will illustrate the usefulness of the **Maxim of Probabilism** in the next two sections on the example of clarifying certain conceptual perplexities concerning conditioning involving probability zero events; the illustration involves *violation* of the **Maxim of Probabilism**. Below we give two examples of concepts that are invariant with respect to isomorphisms; hence these examples illustrate how concepts can satisfy the **Maxim of Probabilism**. The two notions are: the correlation function and the feature of pure measure-theoretic non-atomicity of probability spaces. In both examples below it is assumed that (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) are probability spaces and f is a mod0 isomorphism between (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) , i.e. f is a strict isomorphism between (M, \mathcal{S}_M, p_M) and (N, \mathcal{Z}_N, q_N) .

Example: correlation function

Each probability space (X, \mathcal{S}, p) determines a real-valued function $\text{Corr}_{(X, \mathcal{S}, p)} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ defined by

$$\text{Corr}_{(X, \mathcal{S}, p)}(A, B) \doteq p(A \cap B) - p(A) \cdot p(B) \quad A, B \in \mathcal{S}$$

So we also have

$$\text{Corr}_{(Y, \mathcal{Z}, q)}(A, B) \doteq q(A \cap B) - q(A) \cdot q(B) \quad A, B \in \mathcal{Z}$$

Let (C_X, C_Y) be a pair of events with $C_X \in \mathcal{S}$ and $C_Y \in \mathcal{Z}$. Call this pair *f*-related if

$$f[(C_X \cap M)] = C_Y \cap N \quad (5)$$

Then, since M^\perp and N^\perp are p -measure (respectively q -measure) zero sets and f is a strict isomorphism between (M, \mathcal{S}_M, p_M) and (N, \mathcal{Z}_N, q_N) , we have

$$p(C_X) = p(C_X \cap M) \quad (6)$$

$$= p_M(C_X \cap M) = q_N(f[C_X \cap M]) = q_N(C_Y \cap N) \quad (7)$$

$$= q(C_Y) \quad (8)$$

If (A_X, A_Y) and (B_X, B_Y) are both f -related, then applying (6)-(8) to $C_X = A_X, B_X, (A_X \cap B_X)$ and $C_Y = A_Y, B_Y, (A_Y \cap B_Y)$ we obtain

$$\text{Corr}_{(X, \mathcal{S}, p)}(A_X, B_X) = \text{Corr}_{(Y, \mathcal{Z}, q)}(A_Y, B_Y) \quad (9)$$

Equation (9) means that the notion of a correlation function is invariant under mod0 isomorphisms; hence it satisfies the necessary condition to be *strongly* probabilistic in the spirit of the **Maxim of Probabilism**. Note that if $f: X \rightarrow Y$ is a *strict* isomorphism between (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) , then (9) is simply

$$\text{Corr}_{(X, \mathcal{S}, p)}(A, B) = \text{Corr}_{(Y, \mathcal{Z}, q)}(f[A], f[B]) \quad (10)$$

for all $A, B \in \mathcal{S}$. The content of (10) is that taking the same definition of a correlation function in strictly isomorphic spaces yields the same function, and this means in particular that the notion of a correlation function is invariant under strict isomorphisms.

Example: measure-theoretic non-atomicity

By definition, (X, \mathcal{S}, p) is measure-theoretically purely non-atomic if for any $A \in \mathcal{S}$ with $p(A) > 0$ there is $B \subset A$ such that $p(A) > p(B) > 0$. We show that if (X, \mathcal{S}, p) is measure-theoretically purely non-atomic, and (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) are isomorphic mod0, then (Y, \mathcal{Z}, q) is also measure-theoretically purely non-atomic. The proof relies on the fact that for any event $A \in \mathcal{S}$, $p(A) = p(A \cap M) = p_M(A \cap M)$ and similarly, for $B \in \mathcal{Z}$ we have $q(B) = q_N(B \cap N)$. It follows that (X, \mathcal{S}, p) is purely non-atomic if and only if (M, \mathcal{S}_M, p_M) is purely non-atomic. Suppose (X, \mathcal{S}, p) (and thus (M, \mathcal{S}_M, p_M)) is purely non-atomic. The calculation below shows that the isomorphism f between (M, \mathcal{S}_M, p_M) and (N, \mathcal{Z}_N, q_N) preserves non-atomicity, and therefore (Y, \mathcal{Z}, q) is purely non-atomic as well: Take an $A \in \mathcal{Z}$ with $q(A) > 0$. Then $q_N(A \cap N) > 0$, hence $p_M(f^{-1}[A \cap N]) > 0$. Using non-atomicity of (M, \mathcal{S}_M, p_M) there is $B \cap M \in \mathcal{S}_M$ such that $B \cap M \subset f^{-1}[A \cap N]$ and $p_M(f^{-1}[A \cap N]) > p_M(B \cap M) > 0$. Now, f being an isomorphism ensures $f[B \cap M] \subset A \cap N$ and $q_N(A \cap N) > q_N(f[B \cap M]) > 0$, which completes the proof.

Measure theoretically purely non-atomic spaces are not rare: The Lebesgue measure on $[0, 1]$ defines a purely non-atomic probability measure space. Moreover, we have

Proposition 3.4 ([41][p. 55].) Every probability measure space (X, \mathcal{S}, p) , where X is a complete metric space and \mathcal{S} is the Borel σ -algebra, is isomorphic mod0 to the probability space $[0, 1]$ with the Lebesgue measure – if (X, \mathcal{S}, p) is purely non-atomic.

It is noteworthy that purely non-atomic probability spaces also have philosophically relevant features: they are common cause complete in the sense that they contain a common cause of every correlation $\text{Corr}_{(X, \mathcal{S}, p)}(A, B) > 0$, see [11], [21], [12], [17]. Thus common cause completeness also satisfies the strong necessary condition to be probabilistic.

4 Conditioning and the Maxim of Probabilism

The general concept of conditioning in the measure-theoretic formalism is based on the notion of conditional expectation, which was introduced into probability theory by Kolmogorov in [20] together with his axiomatization. Given (X, \mathcal{S}, p) , and a σ -subalgebra \mathcal{A} of \mathcal{S} , a map

$$\mathcal{E}(\cdot | \mathcal{A}): \mathcal{L}^1(X, \mathcal{S}, p) \rightarrow \mathcal{L}^1(X, \mathcal{S}, p) \quad (11)$$

is an \mathcal{A} -conditional expectation on the set of integrable real-valued random variables $\mathcal{L}^1(X, \mathcal{S}, p)$ if (i) for all $f \in \mathcal{L}^1(X, \mathcal{S}, p)$, the function $\mathcal{E}(f | \mathcal{A})$ is \mathcal{A} -measurable; and (ii) it preserves the integral: $\int_Z \mathcal{E}(f | \mathcal{A}) dp = \int_Z f dp$ for all $Z \in \mathcal{A}$. It is important that the conditional expectation exists as a consequence of the Radon–Nikodym theorem but it is unique only up to p -probability zero. Conditional expectations that differ only on a p -probability zero set are called *versions*. $\mathcal{P}(\cdot | \mathcal{A})$

denotes the restriction of $\mathcal{E}(\cdot | \mathcal{A})$ to (the characteristic functions of) \mathcal{S} . Conditional probabilities of random events $B \in \mathcal{S}$ as real numbers are defined in this framework of conditioning in the following manner:

Let $q_{\mathcal{A}}$ be a probability measure on the Boolean sub- σ -algebra \mathcal{A} that is absolutely continuous with respect to the restriction of p to \mathcal{A} . Then $q_{\mathcal{A}}$ yields a density function g (the Radon–Nikodym derivative), and the conditional probability $q(B | \mathcal{A})$ of $B \in \mathcal{S}$ on condition that the probabilities of events in \mathcal{A} are given by the probability measure $q_{\mathcal{A}}$ is, by definition

$$q(B | \mathcal{A}) \doteq \int_X g \mathcal{P}(\chi_B | \mathcal{A}) dp \tag{12}$$

where χ_B is the characteristic (indicator) function of B . It can be shown (see e.g. [13]) that formula (12) reduces to the Jeffrey rule, if \mathcal{A} is generated by a countable (measurable) partition of X , and that the formula (12) yields Bayes' rule, if \mathcal{A} is generated by one single event A on which $q_{\mathcal{A}}$ takes value 1, provided $q_{\mathcal{A}}$ is absolutely continuous with respect to p . Thus conditionalization using the notion of conditional expectation is a general form of Bayesian conditionalization.

One also finds in the literature a somewhat controversial interpretation of this kind of conditioning however: the value of the function $\mathcal{P}(\chi_B | \mathcal{A})$ on $x \in X$ is sometimes viewed as the “conditional probability of B on condition $\{x\}$ ”:

$$\underbrace{\mathcal{P}(\chi_B | \mathcal{A})(x)}_{\text{conditional probability of } B \text{ on condition } \{x\}} \qquad B \in \mathcal{S} \tag{13}$$

Since it can happen that the p -probability of $\{x\}$ is zero, $p(\{x\}) = 0$, the formula (13) would yield then a conditional probability of B on the probability zero event $\{x\}$; which is regarded as a major virtue of this “Kolmogorovian conditioning”.

It has been recognized in the mainstream literature on probability theory that this concept of conditional probability with respect to probability zero conditioning events is not unproblematic ([27][p. 62]; [35][p. 153, 156]; “Difficulties and Curiosities” in [1][p. 437-439]). All the problems are related to the fact that, since $\mathcal{P}(\cdot | \mathcal{A})$ is the restriction of a *version* of the \mathcal{A} -conditional expectation, $\mathcal{P}(\cdot | \mathcal{A})$ also is only a *version*: Different versions yield different values for $\mathcal{P}(\chi_B | \mathcal{A})(x)$; in fact, if $p(\{x\}) = 0$, then for *any* real number r (in particular any real number r in $[0, 1]$) there is a version such that this version yields r as the value of the “conditional probability of A on condition $\{x\}$ ”. Which values are then the “real” conditional probabilities on condition $\{x\}$?

Another problem is that for a fixed $x \in X$ the map

$$\mathcal{S} \ni A \mapsto \mathcal{P}(\chi_A | \mathcal{A})(x) \in \mathbb{R} \tag{14}$$

is *not* a countably additive map on \mathcal{S} in general [27][p. 47], [1][p. 438-439]. So for a fixed x the map (14) is *not* a probability measure on \mathcal{S} ; hence the values given by equation (13) are *not* probabilities – if one takes seriously the Kolmogorovian specification of what probabilities are: They are given by a probability measure (which is countably additive). Saying that “... conditional probabilities behave ‘essentially’ like ordinary probabilities” [35][p. 156] is just acknowledging that they are *not* probabilities.

One might want to say: The “real” conditional probabilities are the ones that are given by a version for which the map in equation (14) *is* countably additive. A conditional probability $\mathcal{P}(\cdot | \mathcal{A})$ is called *regular* if the map in (14) is countably additive for p -almost all x [27][p. 46]. The problem is that such a version might not exist: there exist probability spaces for which this happens [1][p. 438-439; 443].

These difficulties are well known. The reactions to the difficulties are mixed. Billingsley disregards the difficulty, saying “... it does not matter that conditional probabilities may not, in fact, be measures.” [1][p.439]. One reason why he sees this dismissal as being justified is that one can show that $\mathcal{P}(\cdot | \mathcal{A})(x)$ is additive for an x that has positive p -measure [1][p. 439]. But this is not helpful if one wishes to maintain that

“The whole point of this section [*on conditional expectations*] is the systematic development of a notion of conditional probability that covers conditioning with respect to events of probability 0. This is accomplished by conditioning with respect to collections of events – that is, with respect to σ -fields...” [1][p. 432]

Rosenthal's assessment [35][p. 153] of this “accomplishment” amounts to acknowledging that the goal has not been achieved.

Rao's reaction:

“All these studies show that conditioning in the general case is not simple, and the occasional counterexamples served only to deepen the mystery of the subject.” [27][p. 62]

The alleged mystery involved in conditioning via conditional expectations disappears naturally however if one keeps in mind the **Maxim of Probabilism**: That a concept is genuinely probabilistic only if it is invariant with respect to measure-theoretic isomorphisms:

Assume that (X, \mathcal{S}, p) and (Y, \mathcal{Z}, q) are *strictly* isomorphic via a strict isomorphism $f: X \rightarrow Y$. Let \mathcal{A} be a sub- σ -field of \mathcal{S} . Then \mathcal{A} is taken by f into a sub- σ -field $f[\mathcal{A}] \doteq \{f[A] : A \in \mathcal{A}\}$. Consider versions $\mathcal{E}(\cdot | \mathcal{A})$ and $\mathcal{E}(\cdot | f[\mathcal{A}])$ of the conditional expectations

$$\mathcal{E}(\cdot | \mathcal{A}) : \mathcal{L}^1(X, \mathcal{S}, p) \rightarrow \mathcal{L}^1(X, \mathcal{S}, p) \quad (15)$$

$$\mathcal{E}(\cdot | f[\mathcal{A}]) : \mathcal{L}^1(Y, \mathcal{Z}, q) \rightarrow \mathcal{L}^1(Y, \mathcal{Z}, q) \quad (16)$$

Assume that there is $x_0 \in X$ such that $p(\{x_0\}) = q(\{f(x_0)\}) = 0$. Then for a $g \in \mathcal{L}^1(X, \mathcal{S}, p)$ either

$$\mathcal{E}(g | \mathcal{A})(x_0) \neq \mathcal{E}(f \circ g | f[\mathcal{A}])(x_0)$$

or, if

$$\mathcal{E}(g | \mathcal{A})(x_0) = \mathcal{E}(f \circ g | f[\mathcal{A}])(x_0)$$

then we can take another version $\mathcal{E}'(\cdot | \mathcal{A})$ such that

$$\mathcal{E}'(g | \mathcal{A})(x_0) \doteq \mathcal{E}(g | \mathcal{A})(x_0) + r \quad \text{for } r \neq 0$$

and then

$$\mathcal{E}'(g | \mathcal{A})(x_0) \neq \mathcal{E}(f \circ g | f[\mathcal{A}])(x_0)$$

This means that the concept of a *particular version* of the conditional expectation (or rather the value of a version at a given point) is not invariant with respect to strict isomorphisms.

But definition (12) *does* yield a unique probability value: whichever version of $\mathcal{P}(\cdot | \mathcal{A})$ one takes in (12), since the p -integral is insensitive to p -measure zero differences, the conditional probability defined by (12) is the same. This can be expressed formally by stating that the conditional expectation is unique if considered as a map on the space of equivalence classes of integrable random variables, where the equivalence relation is “equal except on p -probability zero set”. Consequently, the conditional probability values provided by *the* (unique) conditional expectation lacks the ambiguity involved in versions. And this can also be expressed in terms of the **Maxim of Probabilism**: Using the fact that a mod0 isomorphism generates an isomorphism of the spaces of equivalence classes of integrable random variables, one can show [14] that *the* conditional expectation is invariant with respect to mod0 isomorphisms – and so are the conditional probabilities defined by it in the manner of (12).

So the situation is the following:

- (i) The concept of conditional expectation viewed on the space of equivalence classes of functions satisfies the *necessary* condition for a concept to be *strongly* probabilistic.
- (ii) The notion of *version* of a conditional expectation does *not* satisfy the necessary condition for a concept to be even *weakly* probabilistic.

So the “mystery” Rao mentions is explained by the fact that *specific versions* of conditional expectations are not determined probabilistically, they are not purely probabilistic. There is no “canonical version” of a conditional expectation in general, choosing a particular version can only be motivated by considerations that involve non-probabilistic elements. This consequence of the **Maxim of Probabilism** also helps in understanding certain features of the Borel-Kolmogorov Paradox. We discuss this in the next section on the basis of [10].

5 The Maxim of Probabilism and the Borel-Kolmogorov Paradox

The Borel-Kolmogorov Paradox arises from the question “*What is the conditional probability on a great circle on a sphere in 3 dimension if on the sphere one assumes the uniform probability measure?*” One

might have the intuition that the conditional probability in question is determined and is the uniform probability. But the usual definition of conditional probability by the ratio formula (on which Bayes' rule is based) does not yield any conditional distribution on the great circle because any great circle has probability zero in the uniform measure on the sphere. This tension between the intuition and the definition of conditional probability by the ratio formula is the Borel-Kolmogorov Paradox. It has been extensively discussed both in probability theory proper (see [20][p. 50-51], [1][p. 441], [4], [5][p. 203], [25], [26], [27][p. 65], [37]), and in the literature on philosophy of probability (see [3][p. 100-104], [7], [15], [19][p. 470], [18], [22], [10]², [34], [36]).

Kolmogorov [20][p. 50-51] argued that the paradox is resolved if one conditionalizes using the concept of conditional expectation: He specified a σ -field \mathcal{A} on the sphere containing the great circle and calculated a version of the corresponding \mathcal{A} -conditional expectation determined by the σ -field \mathcal{A} and by the uniform probability on the sphere. This version yields a conditional probability on the great circle, but this conditional probability is *not* the uniform probability on the great circle. Although this is counterintuitive, it is a consequence of how Kolmogorov chose the σ -field \mathcal{A} ; and one can show [10] that choosing a different σ -field \mathcal{B} one obtains a version of the corresponding \mathcal{B} -conditional expectation that does yield the uniform conditional probability on the great circle. It was argued in [10] that obtaining both a non-uniform and the uniform conditional probability on the great circle is not a contradiction because the respective two σ -fields \mathcal{A} and \mathcal{B} are not isomorphic, hence they represent different conditioning conditions. More importantly: both the uniform and the non-uniform conditional probability on the great circle are given by specific *versions* of the respective \mathcal{A} - and \mathcal{B} -conditional expectations. And since versions of the conditional expectations are not determined probabilistically (as shown in section 4), neither the Kolmogorovian non-uniform conditional probability, nor the uniform conditional probability on the great circle are determined probabilistically by the facts that (i) one has the uniform probability on the sphere and (ii) one fixes as conditioning σ -fields \mathcal{A} or \mathcal{B} .

This undeterminateness of the conditional probability on the great circle, even in the framework of conditioning via conditional expectations, is concealed by the deceptive determinateness of the versions of the \mathcal{A} - and \mathcal{B} -conditional expectations. This seeming determinateness of the versions is due to the fact that the sphere is a two-dimensional surface, and this allows integrating two-place functions on the sphere with respect to one variable. It is this particular structure of the probability space on the sphere that leads to a natural selection of a version of the \mathcal{A} -conditional expectations featuring in Kolmogorov's resolution – and also of the \mathcal{B} -conditional expectations yielding the uniform conditional probability on the great circle (see [10] for details). But linear dimension is not a property that is invariant with respect to measure-theoretic isomorphisms mod0: The probability measure space consisting of the two-dimensional sphere with the uniform probability on its Lebesgue measurable sets is a purely non-atomic probability space, with the sphere being a complete metric space; hence Proposition 3.4 applies, and so the sphere with its uniform probability is isomorphic mod0 with the unit interval with the Lebesgue measure on it. In this latter probability space there is no natural selection of a version of the conditional expectation that corresponds to the version of the \mathcal{A} -conditional expectation Kolmogorov chose, nor is there a natural choice of a version of the conditional expectation that corresponds to the version of the \mathcal{B} -conditional expectation that yields the uniform conditional probability on the great circle. The **Maxim of Probabilism** tells us then that selecting either the version in the Kolmogorov resolution or in the resolution yielding the uniform conditional probability on the great circle does not satisfy the necessary condition to be even weakly probabilistic: the selections involve non-probabilistic features of the situation.

The upshot is that the **Maxim of Probabilism** tells us that the conditional probability on any given great circle is probabilistically genuinely undetermined by the assumption of the uniform probability on the sphere. Tacit, non-probabilistic reasonings (e.g. symmetry considerations [10]) play a role in influencing our intuition that the conditional probability on a great circle is determined probabilistically by the uniform probability on the sphere. But the group-theoretic structure of the sphere on which the symmetry considerations are based is also not invariant with respect to measure-theoretic isomorphisms.

²The paper [40] points out an important technical error in [10].

6 Concluding comments

The **Maxim of Probabilism** only gives a necessary condition to be satisfied by a concept in order to qualify as probabilistic. Why not sufficient as well? If one views mathematical concepts as originating in the attempts to describe natural and social phenomena (as for instance von Neumann saw mathematics [28]), then no sufficient and necessary conditions are feasible that relate a mathematical structure exclusively to a specific circle of phenomena because typically there is a large variety of phenomena whose main features are described by the *same* mathematical structure. This is so with (bounded) measure theory as well: A lot of diverse phenomena can be described mathematically in terms of bounded measure theory in addition to those that can be regarded as probabilistic in an intuitive sense.

Viewed from this empiricist perspective, Reichenbach's attempt at axiomatizing probability theory aims at specifying a mathematical structure that is richer than measure theory, embodying extra content, the extra content being the frequency interpretation. This leads to the difficulty that a finite frequency interpretation is too constraining and one has to allow that probabilities are limits of relative frequencies in infinite ensembles – but this latter view does not have a direct empirical basis. So Reichenbach creates one artificially by formulating the transcendental (non-empirical) *principle of lawful distribution* (Axiom of Induction). By assigning a major function to this principle in foundations of probability theory Reichenbach moves away from empiricism; on the other hand, insisting on expressing the frequency content in the mathematical axioms of probability theory he tried to remain very close to an empiricist position. We regard this tension the fundamental reason for the difficulties in his foundational work on probability, which however is still a rich source of inspiration – as we hope the idea of utilizing the notion of isomorphism in foundations of probability shows.

Acknowledgement

Research supported in part by the Hungarian National Research, Development and Innovation Office, contract number: K-134275. Zalán Gyenis' work was partially supported by the project no. 2019/34/E/HS1/00044 financed by the National Science Centre, Poland.

References

- [1] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, Third edition, 1995.
- [2] V.I. Bogachev. *Measure Theory*, volume II. Springer, Berlin, Heidelberg, New York, 2007.
- [3] E. Borel. *Éléments de la Théorie des Probabilités*. Librairie Scientifique A. Herman & Fils, Paris, 1909. English translation: by J. Freund, “Elements of the Theory of Probability”, Englewood Cliffs, 1956, Prentice-Hall.
- [4] L. Bungert and P. Wacker. The lion in the attic – A resolution of the Borel–Kolmogorov Paradox. arXiv:2009.04778 [math.PR], 2020.
- [5] B. de Finetti. *Probability, Induction, and Statistics*. Wiley, New York, 1972.
- [6] J. Doob. The development of rigor in mathematical probability theory (1900-1950). *American Mathematical Monthly*, pages 586–595, 1996.
- [7] K. Easwaran. *The Foundations of Conditional Probability*. PhD thesis, University of California at Berkeley, 2008.
- [8] F. Eberhardt and C. Glymour. Hans Reichenbach's probability logic. In D.M. Gabbay, S. Hartmann, and J. Woods, editors, *Handbook of the History of Logic*, volume 10. Inductive Logic, pages 357–388. Elsevier, 2009.
- [9] C. Glymour and F. Eberhardt. Hans Reichenbach. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.
- [10] Z. Gyenis, G. Hofer-Szabó, and M. Rédei. Conditioning using conditional expectations: The Borel-Kolmogorov Paradox. *Synthese*, 194:2595–2630, 2017.

- [11] Z. Gyenis and M. Rédei. Characterizing common cause closed probability spaces. *Philosophy of Science*, 78:393–409, 2011.
- [12] Z. Gyenis and M. Rédei. Atomicity and causal completeness. *Erkenntnis*, 79:437–451, 2014. doi: 10.1007/s10670-013-9456-1.
- [13] Z. Gyenis and M. Rédei. General properties of Bayesian learning as statistical inference determined by conditional expectations. *The Review of Symbolic Logic*, 10:719–755, 2017.
- [14] Z. Gyenis and M. Rédei. Topics in foundations of probability. Manuscript in preparation, 2020.
- [15] A. Hájek. What conditional probability could not be. *Synthese*, 137:273–333, 2003.
- [16] C. Hitchcock and M. Rédei. Reichenbach’s Common Cause Principle. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2020 edition, 2020.
- [17] G. Hofer-Szabó, M. Rédei, and L.E. Szabó. *The Principle of the Common Cause*. Cambridge University Press, Cambridge, 2013.
- [18] C. Howson. Finite additivity, another lottery paradox, and conditionalization. *Synthese*, 191:989–1012, 2014.
- [19] E.T. Jaynes. *Probability Theory. The Logic of Science*. Cambridge University Press, Cambridge, 2003. edited by G. Larry Bretthorst.
- [20] A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. English translation: *Foundations of the Theory of Probability*, (Chelsea, New York, 1956).
- [21] M. Marczyk and L. Wronski. Completion of the causal completeness problem. *The British Journal for the Philosophy of Science*, 66:307–326, 2015.
- [22] W. Myrvold. You can’t always get what you want: Some considerations regarding conditional probabilities. *Erkenntnis*, 80:573–603, 2015. doi:10.1007/s10670-014-9656-3.
- [23] F. Padovani. The Concept of Probability in the Mathematical Representation of Reality by Reichenbach. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 1:344–347, 2011. Book review.
- [24] J. Peijnenburg and D. Atkinson. Grounds and limits: Reichenbach and foundationalist epistemology. *Synthese*, 181:113–124, 2011.
- [25] M.A. Proschan and B. Presnell. Expect the unexpected from conditional expectation. *The American Statistician*, 52(3):248–252, 1998.
- [26] M.M. Rao. Paradoxes in conditional probability. *Journal of Multivariate Analysis*, 27:434–446, 1988.
- [27] M.M. Rao. *Conditional Measures and Applications*. Chapman & Hall/CRC, Boca Raton, London, New York, Singapore, 2nd, revised and expanded edition, 2005.
- [28] M. Rédei. On the tension between physics and mathematics. *Journal for General Philosophy of Science*, 51:411–425, 2020.
- [29] H. Reichenbach. Axiomatik der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 34:568–619, 1932.
- [30] H. Reichenbach. *Wahrscheinlichkeitslehre. Eine Untersuchung über die Logischen und Mathematischen Grundlagen der Wahrscheinlichkeitsrechnung*. Sijthoff, Leiden, 1935.
- [31] H. Reichenbach. *The Theory of Probability*. University of California Press, Berkeley, CA, 1949. Translation with additions of the first, German edition: [30].
- [32] H. Reichenbach. *The Direction of Time*. University of California Press, Los Angeles, 1956.
- [33] H. Reichenbach. *The Concept of Probability in the Mathematical Representation of Reality*. Open Court, Chicago, 2008. English translation by F. Eberhardt and C. Glymour of German original “Der Begriff der Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit”, PhD thesis, 1915.
- [34] M. Rescorla. Some epistemological ramifications of the Borel-Kolmogorov Paradox. *Synthese*, 192:735–767, 2015.
- [35] J.S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, Singapore, 2006.

- [36] T. Seidenfeld. Remarks on the theory of conditional probability: Some issues of finite versus countable additivity. In V. F. Hendricks, editor, *Probability Theory*, pages 167–178. Kluwer Academic Publishers, 2001.
- [37] T. Seidenfeld, M.J. Schervish, and J.B. Kadane. Improper regular conditional distributions. *The Annals of Probability*, 29:1612–1624, 2001.
- [38] R. von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919.
- [39] R. von Mises. *Probability, Statistics and Truth*. Dover Publications, New York, 2nd edition, 1981. Originally published as “Wahrscheinlichkeit, Statistik und Wahrheit” (Springer, 1928).
- [40] P. Wacker. Errata to “Z. Gyenis and G. Hofer-Szabó and M. Rédei: Conditioning using conditional expectations: The Borel-Kolmogorov Paradox. *Synthese*, 194:2595-2630, 2017”. <http://phil.elte.hu/gyz/counterexample.pdf>, 2019. Manuscript.
- [41] P. Walters. *An Introduction to Ergodic Theory*, volume 79 of *Graduate Texts in Mathematics*. Springer, New York, 1982.
- [42] L. Wronski. *Reichenbach’s Paradise. Constructing the Realm of Probabilistic Common “Causes”*. De Gruyter, Warsaw, Berlin, 2014.