

**Zalán Gyenis and Miklós Rédei**

## Defusing Bertrand's paradox

**Article (Accepted version)  
(Refereed)**

**Original citation:**

Gyenis, Zalán and Rédei, Miklós (2013) *Defusing Bertrand's paradox*. [British journal for the philosophy of science](#) . ISSN 1464-3537 (In Press)

DOI: [10.1007/s00407-011-0089-y](https://doi.org/10.1007/s00407-011-0089-y)

© 2013 [Oxford University Press](#)

This version available at: <http://eprints.lse.ac.uk/49717/>

Available in LSE Research Online: July 2013

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

# Defusing Bertrand's Paradox

Zalán Gyenis

Department of Mathematics and its Applications  
Central European University  
Nádor u. 9. H-1051 Budapest, Hungary  
gyz@renyi.hu

Miklós Rédei \*

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

Forthcoming in  
The British Journal for the Philosophy of Science

## Abstract

The classical interpretation of probability together with the Principle of Indifference are formulated in terms of probability measure spaces in which the probability is given by the Haar measure. A notion called Labeling Invariance is defined in the category of Haar probability spaces, it is shown that Labeling Invariance is violated and Bertrand's Paradox is interpreted as the very proof of violation of Labeling Invariance. It is shown that Bangu's attempt [2] to block the emergence of Bertrand's Paradox by requiring the re-labeling of random events to preserve randomness cannot succeed non-trivially. A non-trivial strategy to preserve Labeling Invariance is identified and it is argued that, under the interpretation of Bertrand's Paradox suggested in the paper, the paradox does not undermine either the Principle of Indifference or the classical interpretation and is in complete harmony with how mathematical probability theory is used in the sciences to model phenomena; it is shown in particular that violation of Labeling Invariance does not entail that labeling of random events affects the probabilities of random events. It also is argued however that the content of the Principle of Indifference cannot be specified in such a way that it can establish the classical interpretation of probability as descriptively accurate or predictively successful.

## 1 The main claims

Bertrand's Paradox, published first in [3], is regarded a classical problem in connection with the classical interpretation of probability based on the **Principle of Indifference**, and it continues to attract interest [17], [24], [2], [22] in spite of alleged resolutions that have been suggested in the large and still growing literature discussing the issue ([11] and [16] are perhaps the most well-known suggestions for resolutions; the Appendix in [16] contains a brief summary of a number of typical views of the Paradox).

It is not the aim of this paper to offer yet another "resolution" or criticize the ones available; rather, we suggest a new interpretation of Bertrand's Paradox and analyze its relation to the classical interpretation of probability. The interpretation proposed here should make clear that Bertrand's Paradox cannot be "resolved" – not because it is an unresolvable, genuine paradox but because there is nothing to be resolved: the "paradox" simply states a provable, non-trivial mathematical fact, a fact which is perfectly in line both with the correct intuition about how probability theory should be used to model phenomena and with how probability theory is applied in the sciences.

---

\*Acknowledgement: research supported in part by the Hungarian Scientific Research Found (OTKA). Contract number: K68043.

The key idea of the interpretation to be developed here is that the category of probability measure spaces with an *infinite* set of random events for which a classical interpretation of probability based on the **Principle of Indifference** can be meaningfully formulated is the one in which the set  $X$  of elementary events is a compact topological group, the Boolean algebra  $\mathcal{S}$  representing the set of random events is the set of Borel subsets of  $X$  and the probability measure  $p_H$  is the (normalized) Haar measure on  $\mathcal{S}$ . After stating the **General Classical Interpretation** in terms of the probability measure space  $(X, \mathcal{S}, p_H)$  together with the **Principle of Indifference**, we will define a notion called **Labeling Invariance** in this category of measure spaces: **Labeling Invariance** states that a re-labeling  $(X', \mathcal{S}')$  of random events  $(X, \mathcal{S})$  is an isomorphism between  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$ . **Labeling Invariance** can be interpreted as the expression (in the context of the classical interpretation) of the general intuition we call **Labeling Irrelevance**: that the specific way the random events are named is irrelevant from the perspective of the value of their probability. It will be shown that **Labeling Invariance** does *not* hold in the category of Haar probability measure spaces and we interpret Bertrand's Paradox as stating this provable mathematical fact. We will also argue however that **Labeling Invariance** is *not* the proper way to express **Labeling Irrelevance**: our freedom to choose measure theoretically isomorphic probability theories to describe the same random phenomenon manifests the conventionality of naming random events in probabilistic modeling; thus violation of **Labeling Invariance** is perfectly compatible with **Labeling Irrelevance**.

This interpretation of Bertrand's Paradox makes it possible to formulate precisely the extra condition on re-labelings that ensures that re-labelings *do* preserve the probabilities of events; the condition is an expression of the demand that re-labelings do not affect our epistemic status about the elementary events. We also will show that the recent attempt by Bangu [2] to block the emergence of Bertrand's Paradox by requiring re-labelings to preserve randomness cannot succeed non-trivially.

The interpretation will also make it clear that Bertrand's Paradox does *not* affect the **Principle of Indifference** and does *not*, in and by itself, undermine the classical interpretation of probability – the classical interpretation, the **Principle of Indifference** and **Labeling Invariance** are independent ideas. This is not to say that the classical interpretation is maintainable however; the main problem with it is that it gives the impression that it is possible to infer empirically correct probabilities from an abstract principle stating some sort of epistemic neutrality. It would be a mystery if this were possible, but we will argue in the final section that this is not possible and does not in fact happen in applications of probability theory.

## 2 The elementary classical interpretation of probability

Bertrand's Paradox appeared at a time when probability theory had already progressed from the purely combinatorial phase involving only a finite number of random events to the period when it got intertwined with calculus. This development began in the early 18th century with the appearance of limit theorems (theorem of large numbers, Bernoulli 1713, and central limit theorem, de Moivre 1733, [7]); yet, by the late 19th century the theory had not yet reached the maturity that would have made the mathematical foundations of the theory clear and transparent. This was clearly recognized by Hilbert, who, in his famous lecture in Paris in 1900, mentioned the need of establishing probability theory axiomatically as one of the important open problems (Hilbert's 6th problem [28], [27][p. 32-36]). Hilbert's call was answered only in 1933, when Kolmogorov firmly anchored probability theory within measure theory [13]. (See [6] for the history of some of the major steps leading to the Kolmogorovian axioms.)

In the measure theoretic approach probability theory is a triplet  $(X, \mathcal{S}, p)$ , where  $X$  is the set of elementary random events,  $\mathcal{S}$  (the set of general random events) is a Boolean  $\sigma$  algebra of certain subsets of  $X$  and  $p$  (the probability) is a countably additive measure from  $\mathcal{S}$  into the unit interval  $[0, 1]$ . Typically, one also needs random variables to describe certain features of the phenomenon to be described probabilistically: A (real valued) random variable  $f$  is a measurable function  $f$  from  $X$  into the set of real numbers  $\mathbb{R}$ ; measurability being the requirement that the inverse image  $f^{-1}(d)$  of any Borel set  $d$  in  $\mathbb{R}$  belongs to  $\mathcal{S}$ . The measurability requirement entails that the distribution of a random variable  $d \mapsto p(f^{-1}(d))$  is well-defined, the distribution of  $f$  is in fact the probability measure  $p \circ f^{-1}$  on  $\mathcal{B}(\mathbb{R})$  defined as  $(p \circ f^{-1})(d) = p(f^{-1}(d))$  for all Borel sets  $d \in \mathcal{B}(\mathbb{R})$ . The number  $p(f^{-1}(d))$  is the probability that  $f$  takes its value in  $d$ . Note that the events also can be regarded as random variables: an element  $A$  in  $\mathcal{S}$  can be identified with the characteristic (also called: indicator) function  $\chi_A$  of the set  $A$  (see e.g. [21] for the mathematical notions of measure theoretic probability).

**Remark 1.** One should note that the Kolmogorovian axioms are not the only axiomatization of probability theory: Rényi axiomatized probability theory by taking the concept of conditional probability as primitive [20]; whereas conditional probability is a defined concept in Kolmogorovian axiomatization. Popper's axiomatization [19] takes relative probability as primitive (see [15] for a recent analysis

of Renyi's and Popper's concept). Keynes' replacing the interval  $[0, 1]$  as range of probability by a more general partially ordered set [12] also can be considered as an axiomatization different from the Kolmogorovian one. These axiomatizations notwithstanding, the Kolmogorovian axiomatization is by far the most important and it has been widely accepted as the mathematical theory of probability: All the major mathematical results on probability theory are obtained in this framework and the successful probabilistic models in the sciences also are created in terms of measure theoretic probability theory (for instance: classical statistical mechanics, theory of stochastic processes). One also should note the following:

- Probability theory understood as the triplet  $(X, \mathcal{S}, p)$  specified by the Kolmogorovian axioms is part of pure mathematics; hence "measure theoretic probability" (understood as  $(X, \mathcal{S}, p)$ ) is *not* an interpretation of probability and must thus not be contrasted with any interpretation of probability. More will be said on the relation of  $(X, \mathcal{S}, p)$  to interpretations in section 7.
- "Measure theoretic probability" (the triplet  $(X, \mathcal{S}, p)$ ) also describes mathematically the elementary situations where the set of random events is finite. These include the classical, paradigm examples of random phenomena (all sorts of gambling such as coin flipping, die throwing, card games). To describe these was the main motivation to develop probability theory in the first place. On mild additional conditions, measure theoretic probability also describes the betting situations in which the events to bet on are identifiable with propositions stating that the events happened (or not) – the mild condition is that the propositions are closed under the elementary logical operations, i.e. that they form a Boolean algebra.

The significance of probability theory being part of measure theory is that foundational-conceptual problems of probability theory, such as Bertrand's Paradox, can best be analyzed in terms of measure theoretic concepts. With few exceptions, the papers on Bertrand's Paradox typically do not aim at providing an analysis on this level of abstraction however, and, as a result, the precise nature of the paradox remains less clear than it should be. One such exception is Shackel's paper [24], which raises the issue of "Getting the level of abstraction right" [24][p. 156] explicitly. But the level of abstraction suggested by Shackel is a bit too high. To see why, we recall first the classical interpretation of probability together with the **Principle of Indifference** in measure theoretic terms.

The elementary version of the classical interpretation of probability concerns the probability space  $(X_n, \mathcal{P}(X_n), p_u)$ , where  $X_n$  is a finite set containing  $n$  number of random events and the full power set  $\mathcal{P}(X_n)$  of  $X_n$  represents the set of all events. The probability measure  $p_u$  is determined by the requirement that the probability  $p_u(A)$  be equal to the ratio of the "number of favorable cases to the number of all cases":

$$p_u(A) = \frac{\text{number of elements in the set } \{x_i : x_i \in A\}}{n} \quad (1)$$

This is equivalent to saying that  $p_u$  is the probability measure that is uniform on the set of elementary events.

While it is not always stated and emphasized explicitly, it also is part of the classical interpretation what we call here the **Interpretive Link**: that the numbers  $p_u(A)$  are related to something non-mathematical. Without such an interpretive link, the classical interpretation is not an *interpretation* of probability at all: the numbers  $p_u(A)$  defined by (1) are just pure, simple mathematical relations. There are two standard **Interpretive Links**: The **Frequency Link** and the **Degree of Belief Link**. We restrict the discussion to the **Frequency Link** because the classical interpretation emerged historically and was formulated on the basis of this link:

**Elementary Classical Interpretation**: In case of a finite number of elementary events the probabilities of events are given by the measure  $p_u$  that is uniform on the set of elementary events, and (**Frequency Link**;) the numbers  $p_u(A)$  will be (approximately) equal to the relative frequency of  $A$  occurring in a series of trials producing elementary random events from  $X_n$ .

Notice the future tense in the above formulation: it is this reference for future random trials that distinguishes the classical interpretation (with the Frequency Link) from the frequency interpretation, in which the ensemble of elementary random events determining  $A$ 's relative frequency must be specified *before* one can talk about probabilities (cf. [26][p. 24]).

The classical interpretation so formulated is not maintainable however: simple examples (such as throwing a loaded die) show that it is only under special circumstances that  $p_u(A)$  is indicative of the frequencies with which  $A$  will occur in trials. This is what the **Principle of Indifference** is supposed to express. To state this principle we reformulate first the condition (1). Let  $\Pi_n$  be the group of permutations of the  $n$  element set  $\{1, 2, \dots, n\}$  and  $\pi \in \Pi_n$  be a permutation. Then the probability measure  $p_u$  on  $\mathcal{P}(X_n)$  which is uniform on  $X_n$  is determined uniquely by the condition

$$\text{for every } \pi \in \Pi_n \text{ one has: } p_u(\{x_i\}) = p_u(\{x_{\pi(i)}\}) \quad \text{for all } i \in \{1, 2, \dots, n\} \quad (2)$$

**Elementary Principle of Indifference:** *If the permutation group  $\Pi_n$  expresses epistemic indifference about the elementary random events in  $X_n$ , then the (Elementary) **Classical Interpretation** is correct.*

Thus the (Elementary) **Principle of Indifference** states that the (elementary version of the) classical interpretation of probability is maintainable only if one is epistemically neutral in some sense about the elementary events. For now, we leave open the question of how the content of the “epistemic neutrality” should be specified in order for the **Principle of Indifference** to hold; we will return to the issue of epistemic neutrality in section 7.

### 3 The general classical interpretation of probability in terms of Haar measures

Bertrand’s Paradox is typically regarded as an argument against the universal applicability of the **Principle of Indifference**: Bertrand’s Paradox type arguments are intended to show that applying the **Principle of Indifference** can lead to assigning different probabilities to the same event. Both the original version of the argument and the numerous simplified versions of it involve an (uncountably) *infinite* number of elementary random events however. But then it is not obvious at all how one can apply the **Principle of Indifference** because the formulation of it in the previous section loses its meaning if the set of elementary events is not finite: there is no permutation group in the infinite case with respect to which one could require invariance of the measure yielding the “right” probabilities; equivalently: there is no probability measure on an infinite  $\mathcal{S}$  that would be uniform on the infinite set  $X$  of elementary events. What is then the **Principle of Indifference** in connection with such infinite probability spaces? Without answering this question in suitable generality, Bertrand’s Paradox cannot be properly discussed in measure theoretic concepts.

Shackel’s paper [24], which aims at an analysis of Bertrand’s Paradox in abstract measure theoretic terms, realizes the importance of this question but does not offer a convincing specification of the **Principle of Indifference**: Shackel just *assumes* a measure  $\mu$  on  $\mathcal{S}$  and stipulates that the probabilities  $p(A)$  be given by  $\mu$  as  $p(A) = \mu(A)/\mu(X)$  (“Principle of indifference for continuum sized sets” [24][p. 159]). But there are infinitely many measures  $\mu$  on  $\mathcal{S}$  that could in principle be taken as ones that define a probability  $p$ . Which one should be singled out that yields a  $p$  that could in principle be interpreted as *expressing epistemic indifference* about elements in  $X$ ? This crucial question remains unanswered in [24].

It is clear that without some further structure on an infinite  $X$  it is not possible to single out any probability measure on  $\mathcal{S}$  and hence it is impossible to formulate an indifference principle on such a measurable space. The formulation of the **Elementary Principle of Indifference** in terms of the permutation group  $\Pi_n$  gives a hint about what kind of structure is needed in the more general case however: It is a natural idea to try to replace the permutation group  $\Pi_n$  by another group  $\mathcal{G}$  to be interpreted as expressing epistemic neutrality and hope that the elements  $g$  of  $\mathcal{G}$  determine a function  $\alpha_g: X \rightarrow X$  (an action on  $X$ ) in such a way that if one requires the analogue of (2) by postulating

$$\text{for all } g \in \mathcal{G} : p^*(A) = p^*(\alpha_g[A]) \quad \text{for all } A \in \mathcal{S} \tag{3}$$

then the above condition (3) determines a unique probability measure  $p^*$  on  $\mathcal{S}$ , just like in the case of a finite number of events. Problem is that for a general measurable space  $(X, \mathcal{S})$  with a continuum sized  $X$  there is no guarantee *in general* that a  $\mathcal{G}$  exist leading to a  $p^*$  – much less that it leads to a *unique*  $p^*$ . There is however such a guarantee under some additional assumptions: If  $X$  itself is a topological group satisfying certain conditions.

If  $X$  is a locally compact abelian topological group, or a not necessarily abelian but compact topological group, then there exists a unique (up to multiplication by a constant) measure (called: the Haar measure)  $p_H$  on (the Borel sets of)  $X$  which is invariant with respect to the group action. Furthermore, if  $X$  is compact then the measure  $p_H$  is normalized and  $p_H$  is then a probability measure. (The Appendix collects some elementary facts about the Haar measure; equation (29) in the Appendix formulates the invariance of the Haar measure precisely).

The canonical example of an unbounded Haar measure is the Lebesgue measure on the real line: the Lebesgue measure is the unique measure on the real line that is invariant with respect to the real numbers as an additive group – the group action is the shift on the real line. The same holds for the Lebesgue measure on  $\mathbb{R}^n$ . The normalized restrictions of the Lebesgue measure on  $\mathbb{R}^n$  to bounded, compact subsets of  $\mathbb{R}^n$  are thus distinguished by the feature that they originate from a shift-invariant measure; moreover, the Lebesgue measure on any interval  $[a, b]$  also can be regarded as Haar measure in its own right and the same holds for sets  $\times_i^n [a_i, b_i]$  in  $\mathbb{R}^n$  (cf. Appendix). Both the original Bertrand’s Paradox and all of the simplified versions of it take the normalized restriction of

the Lebesgue measure to some bounded, compact sets in  $\mathbb{R}^n$  ( $n = 1, 2$ ) as the measure that expresses the **Principle of Indifference**. This amounts to interpreting (more or less tacitly) the group that generates the Lebesgue measure as a symmetry expressing epistemic neutrality about the elementary random events.

Thus, in general, the group action on  $X$  determined by  $X$  itself as a group can play the role of the action of the permutation group on  $X_n$ , and the Haar measure  $p_H$  on a compact  $X$  is the analogue of the uniform distribution on  $X_n$  if a non-zero uniform distribution on the elements  $X$  does not exist, which is the case if  $X$  is an infinite set. Note that taking the Haar measure as the analogue of the uniform distribution is also justifiable using maximum entropy techniques (see [10]). In what follows,  $(X, \mathcal{S}, p_H)$  stands for a probability measure space in which  $X$  is a compact topological group with continuous group action,  $\mathcal{S}$  is the Borel  $\sigma$  algebra on  $X$  and  $p_H$  is the Haar measure on  $\mathcal{S}$ . In the terminology of these group and measure theoretic notions the general classical interpretation of probability and the related principle of indifference can be consistently formulated generally as follows:

**General Classical Interpretation:** If  $X$  is a compact topological group, then the probabilities of the events are given by the Haar measure  $p_H$  on (the Borel sets of)  $X$ , and (**Frequency Link:**) the numbers  $p_u(A)$  will be (approximately) equal to the relative frequency of  $A$  occurring in a series of trials producing elementary random events from  $X$ .

**General Principle of Indifference:** If  $X$  is a compact topological group and  $if$  the group action expresses epistemological indifference about the elementary random events in  $X$ , *then* the General Classical Interpretation is correct.

## 4 Labeling Invariance and Labeling Irrelevance

Probabilities understood in the spirit of the *elementary* classical interpretation based on the **Principle of Indifference** have the property we call here **Labeling Invariance**: the probability measure which is permutation invariant in one labeling assigns the same value of probability to the random events as the probability measure that is permutation invariant in a different labeling of the same random events. For instance, if one has a symmetric die sides of which are numbered by numbers  $1, 2, \dots, 6$ , then the permutation invariant measure on the elementary events is the uniform probability measure that assigns the value  $\frac{1}{6}$  to each side. If, instead of labeling the sides by the numbers  $1, 2, \dots, 6$ , we label them by painting them using the colours *red, blue, green, yellow, black, white*, say, then the measure which is invariant with respect to the permutation of colours will assign the value  $\frac{1}{6}$  to each colour, i.e. the same value that was assigned to the sides on the basis of the labeling using the numbers  $1, 2, \dots, 6$  as labels.

**Labeling Invariance** seems to be an important and attractive feature of probabilities interpreted according to the elementary classical interpretation because it entails that **Labeling Irrelevance** holds for the elementary classical interpretation: **Labeling Irrelevance** states that labeling of random events does not affect their probabilities, that a particular labeling of random events is a matter of convention. **Labeling Irrelevance** is an important condition in probabilistic modeling: its violation would entail a radical ambiguity and arbitrariness in assigning probabilities to random events. Without **Labeling Irrelevance** one could not collect statistical data, and violation of **Labeling Irrelevance** is obviously incompatible with any interpretation of probability that treats probability as an objective feature of the world; furthermore, violation of **Labeling Irrelevance** also would make subjective degrees of belief vulnerable to a Dutch book: if degrees of belief would change as a result of re-naming the random events involved, then the bookie could make money by just re-naming the events in the bet. Given this conceptual importance of **Labeling Irrelevance** and the fact that **Labeling Invariance** entails it, one expects **Labeling Invariance** to hold for the *general* classical interpretation as well. We will see shortly that Bertrand's Paradox can be viewed as proof that **Labeling Invariance** does *not* hold in the general classical interpretation. It should be emphasized however that this does *not* entail that **Labeling Irrelevance** cannot be maintained in general, nor does it follow from violation of **Labeling Invariance** that the classical interpretation based on the **Principle of Indifference** is inconsistent because **Labeling Invariance** and **Labeling Irrelevance** are not equivalent. We will say more on the relation of **Labeling Invariance**, **Labeling Irrelevance** and probabilistic modeling in section 7.

To formulate the idea of **Labeling Invariance** generally and precisely, we need the notion of re-labeling (re-naming) first: If  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  are two probability spaces describing the same phenomenon, then the map  $h: X \rightarrow X'$  is called a re-labeling if it is a bijection between  $X$  and

$X'$  and both  $h$  and its inverse  $h^{-1}$  are measurable, i.e. it holds that

$$h[A] \in \mathcal{S}' \quad \text{for all } A \in \mathcal{S} \quad (4)$$

$$h^{-1}[B] \in \mathcal{S} \quad \text{for all } B \in \mathcal{S}' \quad (5)$$

(Here  $h[A] = \{h(x) : x \in A\}$  and  $h^{-1}[A'] = \{h^{-1}(x') : x' \in A'\}$ .) Note that without the measurability condition required of  $h$  it can happen that a general event  $A \in \mathcal{S}$  has probability but its re-named version  $h[A]$  does not – in this case  $h$  cannot be called re-naming of random events (and similarly for  $h'$ ).

**Labeling Invariance** is the claim that the probabilities understood in the spirit of the classical interpretation are invariant with respect to re-naming; that is to say, if  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  are two probability spaces and  $h$  is a re-labeling between  $X$  and  $X'$  then it holds that

$$p'_H(h[A]) = p_H(A) \quad \text{for all } A \in \mathcal{S} \quad (6)$$

$$p_H(h^{-1}[A']) = p'_H(A') \quad \text{for all } A' \in \mathcal{S}' \quad (7)$$

Recall (see e.g. [1][p. 3]) that two probability measure spaces  $(X, \mathcal{S}, p)$  and  $(X', \mathcal{S}', p')$  are called isomorphic if there are sets  $Y \in \mathcal{S}$  and  $Y' \in \mathcal{S}'$  such that  $p(Y) = 0 = p'(Y')$  and there exists a bijection  $f: (X \setminus Y) \rightarrow (X' \setminus Y')$  such that both  $f$  and its inverse  $f^{-1}$  are measurable and such that both  $f$  and  $f^{-1}$  preserve the measure  $p$  and  $p'$ , respectively; i.e. (8)-(9) below hold:

$$p'(f[A]) = p(A) \quad \text{for all } A \in \mathcal{S} \quad (8)$$

$$p(f^{-1}[A']) = p'(A') \quad \text{for all } A' \in \mathcal{S}' \quad (9)$$

The function  $f$  is called then an isomorphism between the probability measure spaces. **Labeling Invariance** can therefore be expressed compactly by saying

**Labeling Invariance:** Any re-labeling between probability spaces  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  is an isomorphism between these probability spaces.

## 5 General Bertrand's Paradox

**Labeling Invariance** is obviously a very strong claim and Bertrand's paradox can be interpreted as the proof that it cannot be maintained in general (see below). But why would one think that **Labeling Invariance** holds in the first place? The answer is: because **Labeling Invariance** *does* hold for an *infinite* number of probability spaces: for probability spaces with any *finite* number elementary random events probabilities of which are given by the uniform probability measure. A bijection  $h$  between two finite sets  $X_n$  and  $X' = X_m$  of elementary events exists if and only if the sets  $X_n$  and  $X_m$  have the same number of elements,  $n = m$ , and this entails that the two uniform distributions on those equivalent sets will assign the same probability to  $A$  and  $h[A]$  (and to  $A'$  and  $h^{-1}[A']$ ) – no Bertrand's Paradox can arise in this case. Since the intuition about probability theory was shaped historically by situations involving only a finite number of random events, it is not surprising that **Labeling Invariance** became part of the intuition about probability. It turns out however that this intuition is a poor guide if the set of elementary events is not finite: This is precisely what Bertrand's Paradox shows, general form of which is the following statement:

**General Bertrand Paradox:** Let  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  be probability spaces with compact topological groups  $X$  and  $X'$  having an infinite number of elements and  $p_H, p'_H$  being the respective Haar measures on the Borel  $\sigma$  algebras  $\mathcal{S}$  and  $\mathcal{S}'$  of  $X$  and  $X'$ . Then **Labeling Invariance** does not hold for  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  in the sense that

- either there is no re-labeling between  $X$  and  $X'$ ;
- or, if there is a re-labeling between  $X$  and  $X'$ , then there also exists a re-labeling that violates **Labeling Invariance**.

The **General Bertrand's Paradox** is a trivial consequence of the following non-trivial theorem in measure theory:

**Proposition 1** ([25], [23]). *If  $X$  is an infinite, compact topological group with the Haar measure  $p_H$  on the Borel  $\sigma$  algebra  $\mathcal{S}$  of  $X$ , then there exists an autohomeomorphism  $\theta$  of  $X$  and an open set  $E$  in  $\mathcal{S}$  such that  $p_H(\theta[E]) \neq p_H(E)$ .*

By definition an autohomeomorphism  $\theta$  of  $X$  is a bijection from  $X$  into  $X$  such that both  $\theta$  and its inverse  $\theta^{-1}$  are continuous. Since continuous functions are Borel measurable, an autohomeomorphism is a re-labeling: a re-labeling of  $X$  in terms of its own elements. Assume now that  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  are two probability spaces with infinite, compact topological groups  $X$  and  $X'$  and Haar measures  $p_H$  and  $p'_H$ . If  $h: X \rightarrow X'$  is a re-labeling between  $X$  and  $X'$  then either  $h$  is an isomorphism between the probability spaces (i.e. preserves the probability in the sense of (6)-(7)) or it is not. If it is not, then **Labeling Invariance** is violated by  $h$ . If  $h$  *does* preserve the probability (and is thus an isomorphism between  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$ ) then by Proposition 1 there exists an autohomeomorphism  $\theta$  on  $X$  and there exists an open set  $E \in \mathcal{S}$  such that  $p_H(\theta[E]) \neq p_H(E)$ . This means that for the re-labeling given by the composition  $h \circ \theta$  we have

$$p'_H((h \circ \theta)[E]) = p'_H(h[\theta[E]]) = p_H(\theta[E]) \neq p_H(E) \quad (10)$$

so the re-labeling  $h \circ \theta$  violates (6) and thus  $h \circ \theta$  violates **Labeling Invariance**. In either case **Labeling Invariance** is violated. Furthermore, the autohomeomorphism ensured by Proposition 1 provides a re-labeling of the elementary set of events of any infinite compact group in terms of its own elementary events in such a way that the Haar measure yielding the probabilities of the events in the spirit of the classical interpretation are not preserved under the re-labeling.

The General Bertrand's Paradox is thus a general feature of infinite probability measure spaces with the Haar measure yielding the probabilities, and note that it says more than the original Bertrand's Paradox, which only claimed that there exist Haar measures and re-labelings that violate **Labeling Invariance**: The General Bertrand's Paradox says that *no two* infinite Haar probability spaces can satisfy **Labeling Invariance**; i.e. if there is at all a re-labeling between two probability spaces  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  with infinite  $X$  and  $X'$  then there is also a re-labeling between these spaces that violates **Labeling Invariance**, and for *any* space  $(X, \mathcal{S}, p_H)$  with an infinite  $X$  there exists a space (namely itself) and a self-re-labeling of  $(X, \mathcal{S}, p_H)$  that violates **Labeling Invariance**. Thus Bertrand's 1888 Paradox can be viewed as the specific "Lebesgue measure case" of a mathematical theorem that was proved in full generality in 1993 only.

We close this section by giving an explicit, elementary example of violation of **Labeling Invariance**; this example will be referred to in the next section. In a well-defined sense (explained in Remark 2) the example is general.

**Example** Let  $[a, b]$  and  $[c, d]$  be two closed intervals of the real numbers and

$$([a, b], \mathcal{S}_{[a, b]}, p_{[a, b]}) \text{ and } ([c, d], \mathcal{S}_{[c, d]}, p_{[c, d]})$$

be the two probability spaces with  $p_{[a, b]}$  and  $p_{[c, d]}$  being the normalized Lebesgue measures on the intervals  $[a, b]$  and  $[c, d]$ , with  $\mathcal{S}_{[a, b]}$  and  $\mathcal{S}_{[c, d]}$  being the Borel measurable sets of the respective intervals. Elementary algebraic calculation and reasoning show that one can choose the parameters  $\alpha, \beta$  and  $\gamma$  in the definition of the simple quadratic map  $h$  defined on the real line by

$$h(x) = \alpha x^2 + \beta x + \gamma \quad (11)$$

in such a way that  $h$  maps  $[a, b]$  to  $[c, d]$  bijectively and both  $h$  and its inverse are continuous hence (Borel) measurable. Thus (the restriction to  $[a, b]$  of)  $h$  is a re-labeling between  $([a, b], \mathcal{S}_{[a, b]}, p_{[a, b]})$  and  $([c, d], \mathcal{S}_{[c, d]}, p_{[c, d]})$ . Specifically, the parameters below have this feature

$$\alpha = \frac{d-c}{(b-a)^2} \quad (12)$$

$$\beta = -2a \frac{d-c}{(b-a)^2} \quad (13)$$

$$\gamma = a^2 \frac{d-c}{(b-a)^2} + c \quad (14)$$

Furthermore, if  $\epsilon$  is a real number such that  $[a, a + \epsilon] \subseteq [a, b]$  then

$$p_{[a, b]}([a, a + \epsilon]) = \frac{\epsilon}{b-a}$$

and since  $h$  takes  $[a, a + \epsilon]$  into  $[c, c + \frac{d-c}{(b-a)^2} \epsilon^2]$  one has

$$p_{[c, d]}(h[[a, a + \epsilon]]) = \frac{1}{d-c} \left( c + \frac{d-c}{(b-a)^2} \epsilon^2 \right)$$

It is clear then that for many  $\epsilon$

$$p_{[a, b]}([a, a + \epsilon]) = \frac{\epsilon}{b-a} \neq \frac{1}{d-c} \left( c + \frac{d-c}{(b-a)^2} \epsilon^2 \right) = p_{[c, d]}(h[[a, a + \epsilon]]) \quad (15)$$

which is a violation of **Labeling Invariance**.



**Remark 2.** Note that the above example is typical in the following sense: A probability measure space is called a *standard probability space* if  $X$  is a complete, separable metric space and  $\mathcal{S}$  is the Borel  $\sigma$  algebra of  $X$ . Standard, non-atomic probability measure spaces are isomorphic to  $([a, b], \mathcal{L}_{[a,b]}, p_{[a,b]})$  with some interval  $[a, b]$  where  $\mathcal{L}_{[a,b]}$  is the algebra of Lebesgue measurable sets in  $[a, b]$  (see [1][Chapter 1, p. 3]). Hence the above example gives a large number of re-labelings that violate **Labeling Invariance** in the category of spaces  $(X, \mathcal{S}, p_H)$  with  $X$  being a complete, separable metric space. This covers all the spaces that occur in connection with Bertrand’s Paradox.

## 6 Attempts to save Labeling Invariance

One may attempt to defend **Labeling Invariance** by trying to block the emergence of the general Bertrand’s Paradox. The previous section makes it clear what the possible strategies are to achieve this: One can impose some extra condition on re-labelings that entails either that re-labelings satisfying the extra conditions do not exist (Strategy A) or that the re-labelings satisfying the additional conditions force the re-labelings to be isomorphisms of the probability spaces (Strategy B). Although not formulated in this terminology, Bangu’s recent attempt [2] is an example of Strategy A. We show below that Bangu’s suggestion for Strategy A is ambiguous however and that resolving the ambiguity makes it either a trivial case of Strategy B or is unsuccessful. A *successful* implementation of Strategy B is to say that it is unreasonable to expect a re-labeling to preserve probabilities unless the re-labeling also preserves our epistemic status with respect to the elementary events: after all, the **Principle of Indifference** states that  $p_H$  is the empirically correct probability *only if* the group structure of  $X$  expresses epistemic neutrality. So the following stipulation is in the spirit of the **Principle of Indifference**:

**Definition:** The re-labeling  $h$  between probability spaces  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  preserves the epistemic status if it is a group isomorphism between  $X$  and  $X'$ .

Since the probability measures  $p_H$  and  $p'_H$  are completely determined by the respective group actions, re-labelings that preserve the epistemic status are isomorphisms between the measure spaces, hence no Bertrand’s Paradox can arise with respect to such re-labelings; furthermore, not every re-labeling is a group isomorphism – thus this strategy works in a non-trivial way.

Bangu’s suggestion is that one should only expect **Labeling Invariance** to hold for bijections that “preserve randomness” – this is his **Assumption R** – Bertrand’s paradox is only a paradox in his view if **Labeling Invariance** is violated by re-labelings satisfying the randomness condition, which, he claims, has *not* been shown and the burden of proof is on those who claim such re-labelings exist. It is clear from the wording of his paper that he conjectures that no such proof can be given, i.e. that no randomness preserving re-labelings exist that violate **Labeling Invariance** (i.e. that he is following Strategy A).

As Bangu also points out, the notion of randomness is notoriously both vague and rich: the adjective “random” can be applied to different entities (events, processes, dynamics, ensembles etc.), it can come in the form of a pre-theoretical informal intuition, in the form of precise mathematical definitions, and it also can come in degrees. Thus one has to be very careful and specific when it comes to the problem of whether “randomness is preserved” under a re-labeling of the elementary random events. Bangu leaves it deliberately open in what sense precisely “randomness” might not be invariant under re-labeling of the random events; hence his suggestion remains vague. No matter what kind of notion of randomness one has in mind, if it is to be relevant for probabilistic modeling of a phenomenon, then it must be expressible in terms of probabilities, since the basic principle guiding the modeling of phenomena by probability theory is the maxim:

**Distribution Relevance:** “A property is probability theoretical if, and only if, it is describable in terms of a distribution” [14][p. 171].

In the spirit of **Distribution Relevance** one can take the position that randomness of a phenomenon expressed by “randomness” of the random variables that describe the phenomenon are encoded in the *distribution* of the random variables. Consequently, under this interpretation of randomness, if one is given two probability models  $(X, \mathcal{S}, p)$  and  $(X', \mathcal{S}', p')$  of a given phenomenon and  $h: X \rightarrow X'$  is a re-labeling between  $(X, \mathcal{S}, p)$  and  $(X', \mathcal{S}', p')$ , then  $h$  preserves the randomness of the two probabilistic descriptions if and only if it holds that if  $f: X \rightarrow \mathbb{R}$  is any random variable in  $(X, \mathcal{S}, p)$  with distribution  $p \circ f^{-1}$  then the distribution  $p' \circ f'^{-1}$  in  $(X', \mathcal{S}', p')$  of the re-named random variable  $f' = f \circ h^{-1}$  coincides with  $p \circ f^{-1}$ :

$$(p' \circ (f \circ h^{-1})^{-1})(d) = (p \circ f^{-1})(d) \quad \text{for all } d \in \mathcal{B}(\mathbb{R}) \quad (16)$$

and conversely: for every random variable  $g': X' \rightarrow \mathbb{R}$  which is the re-named version of a random variable  $g = g' \circ h$  in  $(X, \mathcal{S}, p)$  it holds that the distribution  $p' \circ g'^{-1}$  in  $(X', \mathcal{S}', p')$  of  $g'$  and the

distribution  $p \circ g^{-1}$  of  $g = g' \circ h$  in  $(X, \mathcal{S}, p)$  coincide:

$$(p \circ (g' \circ h)^{-1})(d) = (p' \circ g'^{-1})(d) \quad \text{for all } d \in \mathcal{B}(\mathbb{R}) \quad (17)$$

Since the random events themselves are random variables, the two equations (16)-(17) must hold for every characteristic function  $\chi_A$  ( $A \in \mathcal{S}$ ) in place of  $f$  and every characteristic function  $\chi_{A'}$  ( $A' \in \mathcal{S}'$ ) in place of  $g'$  as well, so this requirement of preserving randomness amounts to the demand that the following two equations hold:

$$p'(h[A]) = p(A) \quad \text{for all } A \in \mathcal{S} \quad (18)$$

$$p(h^{-1}[A']) = p'(A') \quad \text{for all } A' \in \mathcal{S}' \quad (19)$$

which is precisely **Labeling Invariance** (eqs. (6)-(7)). So, if “preserving randomness by re-labeling” in **Assumption R** is understood in the spirit of **Distribution Relevance** as conditions (16)-(17) then the only randomness-preserving re-labelings are the isomorphisms and no Bertrand paradox can arise indeed – requiring preserving randomness in this sense is equivalent to the requirement that the re-labelings are isomorphism. Strategy A, so interpreted, is trivial.

One can try to argue that this is an extremely strong interpretation of “preserving randomness” and that randomness also can be interpreted differently as expressed by some other property  $\Phi(p)$  of the probability measure  $p$ . For instance, one has the intuition that a probability measure sharply concentrated on a single point in  $X$  is far less “random”, it represents much more certainty by having zero variance than a probability distribution that has a large variance. The usual (Shannon) entropy of a probability measure also can be taken as a measure of “randomness” of the phenomenon that the probability model describes [4][p. 61-62]. Thus one can interpret the requirement of “preserving randomness under re-labeling” in **Assumption R** in different ways depending on what property  $\Phi$  one chooses:

**Assumption R** $[\Phi]$ : If  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  are two probability spaces and  $h$  is a re-labeling between  $X$  and  $X'$  then we say that **Assumption R** $[\Phi]$  is satisfied if both  $\Phi(p_H)$  and  $\Phi(p'_H)$  hold.

It is clear then that if there is a property  $\Phi$  of randomness of a probability measure and there exists probability spaces  $(X, \mathcal{S}, p_H)$  and  $(X', \mathcal{S}', p'_H)$  with a re-labeling  $h: X \rightarrow X'$  such that **Assumption R** $[\Phi]$  is satisfied but **Labeling Invariance** is violated by  $h$  then Bertrand’s paradox re-emerges.

The variance and the entropy are such properties: Consider the probability spaces  $([a, b], \mathcal{S}_{[a,b]}, p_{[a,b]})$  and  $([c, d], \mathcal{S}_{[c,d]}, p_{[c,d]})$  described in the **Example** in section 5. The variance  $\sigma(p_{[a,b]})$  of the normalized Lebesgue measure  $p_{[a,b]}$  on any interval  $[a, b]$  is by definition equal to

$$\sigma(p_{[a,b]}) = \int_a^b \frac{1}{b-a} x^2 dx - \left[ \int_a^b \frac{1}{b-a} x dx \right]^2 = \frac{(b-a)^2}{12} \quad (20)$$

and the entropy  $E(p_{[a,b]})$  of  $p_{[a,b]}$  is by definition

$$E(p_{[a,b]}) = - \int_a^b \frac{1}{b-a} \log\left(\frac{1}{b-a}\right) dx = \log(b-a) \quad (21)$$

It follows then that if  $b-a = d-c = t$  then

$$\sigma(p_{[a,b]}) = \sigma(p_{[c,d]}) = \frac{t^2}{12} \quad (22)$$

$$E(p_{[a,b]}) = E(p_{[c,d]}) = \log(t) \quad (23)$$

On the other hand, the map  $h$  defined in the **Example** remains a re-labeling even if  $b-a = d-c$  and **Labeling Invariance** is violated by this map because for  $b-a = d-c = t$  eq. (15) entails that for many  $\epsilon$  we have

$$p_{[a,b]}([a, a+\epsilon]) = \frac{\epsilon}{t} \neq \frac{1}{t} \left( c + \frac{1}{t} \epsilon^2 \right) = p_{[c,d]}(h([a, a+\epsilon])) \quad (24)$$

Thus Bertrand’s paradox re-emerges: The probability space  $([c, d], \mathcal{S}_{[c,d]}, p_{[c,d]})$  can be regarded as a re-named version of the probability space  $([a, b], \mathcal{S}_{[a,b]}, p_{[a,b]})$  via the re-labeling  $h$  defined by (11) and (12)-(14), furthermore, if  $b-a = d-c$  then this re-labeling satisfies **Assumption R** $[\Phi]$  with  $\Phi$  being the variance or entropy, and because of (15)  $h$  violates **Labeling Invariance** (6)-(7).

One also can try to question **Distribution Relevance**. But if one gives up **Distribution Relevance** and interprets “randomness” in a way that makes randomness not expressible exclusively in terms of the distributions involved, then the appropriately modified **Assumption R** constrains even less the emergence of Bertrand’s Paradox. Rowbottom and Schackle [22] take **Assumption R** to be (a technically undefined) “unpredictability” and argue (informally) that there are re-labelings that

preserve “unpredictability” and which are not isomorphisms, contrary to what Bangu [2] seems to conjecture. As a technically more explicit example, assume that a dynamic  $\{\alpha_t : t \in \mathbb{R}\}$  is given on  $([a, b], \mathcal{S}_{[a,b]}, p_{[a,b]})$  and a dynamic  $\{\alpha'_t : t \in \mathbb{R}\}$  is given on  $([c, d], \mathcal{S}_{[c,d]}, p_{[c,d]})$ , where  $\alpha_t$  and  $\alpha'_t$  are one parameter groups of measure preserving maps on  $[a, b]$  and  $[c, d]$  respectively. As randomness of the dynamical systems  $([a, b], \mathcal{S}_{[a,b]}, p_{[a,b]}, \{\alpha_t\})$  and  $([c, d], \mathcal{S}_{[c,d]}, p_{[c,d]}, \{\alpha'_t\})$  one can take the randomness of the respective *dynamics* such as ergodicity, or mixing, which are not expressible in terms of  $p_{[a,b]}$  and  $p_{[c,d]}$  only. Given the re-labeling  $h$  between  $([a, b], \mathcal{S}_{[a,b]}, p_{[a,b]})$  and  $([c, d], \mathcal{S}_{[c,d]}, p_{[c,d]})$  described in the **Example** in section 5 that violates **Labeling Invariance** one can then specify the dynamics  $\{\alpha_t\}$  and  $\{\alpha'_t\}$  in such a way that they are both ergodic, [4][p. 34], generating a Bertrand’s Paradox, or in such a way that  $\{\alpha_t\}$  is ergodic whereas  $\{\alpha'_t\}$  is not, which would be a violation of preserving randomness (**Assumption R**) hence not a case of Bertrand’s Paradox (according to Bangu’s requirement) – anything is possible under such a dynamical interpretation of randomness.

Thus the emergence of Bertrand’s paradox cannot be blocked in a non-trivial way by requiring the paradoxical examples to satisfy the randomness test and showing that they cannot pass this test: unless one requires in effect that the re-labeling be an isomorphism, Bertrand’s Paradox emerges: If **Distribution Relevance** is accepted and randomness is interpreted as measured by the variance or entropy of the probability measures, then elementary examples can be given that show violation of **Labeling Invariance**. If **Distribution Relevance** is abandoned, then the randomness requirement can be satisfied even more easily.

## 7 Comments on the classical interpretation of probability

Probability theory, as specified by the Kolmogorovian axioms, is part of pure mathematics. Just like other branches of pure mathematics, probability theory also can be used to describe certain phenomena however. One has to distinguish such *applications* of probability theory both from pure mathematics and from *interpretations* of probability as this latter term is used in philosophy of science.

In an application of probability theory one relates the mathematical elements in a triplet  $(X, \mathcal{S}, p)$  to non-mathematical entities. This involves two tasks:

**Event Interpretation** To specify what the elements in  $X$  and  $\mathcal{S}$  stand for.

**Truth Interpretation** To clarify when the proposition “ $p(A)=r$ ” is true/false.

In an application, probability theory thus becomes a mathematical *model* of a certain phenomenon that is external to mathematics. A probability measure space is a good model of the phenomenon if it has two features: descriptive accuracy and predictive success. Descriptive accuracy means that under the fixed specification of the Event and Truth Interpretations propositions such as  $p(A) = r$  are true about events that have been observed in the past. Predictive success means that the probabilistic propositions  $p(A) = r$  will be true in future observations. It is clear that both descriptive correctness and predictive success are *robustly empirical features* and that descriptive accuracy up until time  $T$  does *not* entail predictive success for times after  $T$  – this is just a particular formulation of the problem of induction; hence, whether a probability space is a good model is a question that can be answered only on the basis of empirical considerations. This is of course not new, there is nothing peculiar or mysterious about probabilistic modeling, probabilistic scientific theories are just like any scientific theory from this perspective.

The mathematical notion of isomorphism between probability measure spaces is in complete harmony with the application of probability theory – and so is the General Bertrand Paradox: The Event Interpretation and Truth Interpretation are conceptually different issues, the former does not determine the latter, and, accordingly, two probability spaces are defined to be isomorphic if *two* conditions are satisfied: the random events in the two spaces are connected by a re-labeling *and* the re-labeling preserves the probabilities. From the perspective of the notion of isomorphisms of probability spaces finite probability spaces with the uniform probability measure just happen to have the “contingent” feature that in this category re-labelings *are* isomorphisms; in this case the re-labelings contain enough information to make them isomorphisms.

This contingent feature is very deceptive however because it gives the impression that **Labeling Invariance** is the proper way to ensure **Labeling Irrelevance**. It is because of this conflation of **Labeling Invariance** and **Labeling Irrelevance** that violation of **Labeling Invariance** in the category of Haar measure spaces (i.e. General Bertrand’s Paradox) appears paradoxical. But there is nothing paradoxical about this, **Labeling Irrelevance** is respected in probabilistic modeling perfectly well – but it is respected not by **Labeling Invariance** holding true: If the elements in

the pair of sets  $(X, \mathcal{S})$  label the (elementary, respectively, general) random events of some random phenomenon, then one is free to use another pair of sets  $(X', \mathcal{S}')$  to label the events as long as no random events are lost in  $X'$  and  $\mathcal{S}'$ , i.e. as long as there is a re-labeling  $h$  between  $X$  and  $X'$ . **Labeling Irrelevance** says that the choice of  $(X, \mathcal{S})$  or  $(X', \mathcal{S}')$  does not affect the probabilities of the random events and this is in harmony with the fact that fixing either  $(X, \mathcal{S})$  or  $(X', \mathcal{S}')$  does not determine any probability measure on either  $(X, \mathcal{S})$  or  $(X', \mathcal{S}')$ : any of the (continuum number of) mathematically possible probability measures can be defined on both  $(X, \mathcal{S})$  and  $(X', \mathcal{S}')$ . So, if the probability measure  $p$  is such that  $(X, \mathcal{S}, p)$  is a descriptively accurate probabilistic model of the phenomenon in question, then  $(X', \mathcal{S}', p')$  with  $p'$  defined by  $p' \equiv p \circ h$  also is a descriptively accurate model of the same phenomenon and  $(X, \mathcal{S}, p)$  and  $(X', \mathcal{S}', p')$  are isomorphic with respect to  $h$ . Conversely, if  $(X, \mathcal{S}, p)$  and  $(X', \mathcal{S}', p')$  are isomorphic probability spaces with respect to  $h$ , then (up to a measure zero set)  $h$  is a re-labeling of the events preserving probabilities, and either  $(X, \mathcal{S}, p)$  or  $(X', \mathcal{S}', p')$  can be used to describe the phenomenon, choosing any of them – choosing any of the two labelings  $(X, \mathcal{S})$  or  $(X', \mathcal{S}')$  in particular – is a matter of convention. In short: **Labeling Irrelevance** is encoded in the notion of isomorphism of probability spaces and in the claim that isomorphic probability spaces can be used to describe the same phenomenon. Note that this interpretation of how **Labeling Irrelevance** is ensured in probabilistic modeling does not depend on any particular features of the probability spaces used in modeling; in particular it is not assumed that the probability measure is the uniform probability or a Haar measure. Indeed, it would be unacceptable if **Labeling Irrelevance** would only hold for situations in which the probabilities of the events are given by such special measures.

*Interpretations* of probability are typical *classes* of applications of probability theory, classes consisting of applications that possess some common features, which the interpretation isolates and analyzes. The elementary classical interpretation concerns the application of the particular probability spaces  $(X_n, \mathcal{S}, p_u)$ , where the set  $X_n$  of elementary events is finite and the probability measure  $p_u$  is the uniform probability measure on  $X_n$ . The problem with the classical interpretation (understood with the amendment of the **Principle of Indifference**) is *not* Bertrand's Paradox, i.e. not that the **Principle of Indifference** cannot be consistently generalized from  $(X_n, \mathcal{S}, p_u)$  to the case of an infinite number of elementary random events: We have seen in section 3 that a **General Principle of Indifference** and a **General Classical Interpretation** *can* be consistently formulated in terms of Haar measure spaces  $(X, \mathcal{S}, p_H)$  and one *can* in principle maintain the **General Classical Interpretation**. What one has to abandon is the general version of **Labeling Invariance**, which is not maintainable in view of the General Bertrand Paradox. But the **Principle of Indifference** and **Labeling Invariance** are *independent* ideas: One can reject **Labeling Invariance** completely, and one can do so without abandoning the conceptually important **Labeling Irrelevance** because, as we have seen above, **Labeling Invariance** is different from **Labeling Irrelevance**. One also can in principle restrict **Labeling Invariance** to the domain in which it holds: to the category of probability measure spaces with a finite number of random events or to re-labelings that preserve the epistemic status. All these options are logically consistent with the logically consistent formulation of the **General Principle of Indifference** and with the General Bertrand's paradox – Bertrand's Paradox is defused.

This is not to say that the classical interpretation of probability is acceptable however. The main problem with the classical interpretation is that it disregards the empirical character of the applications of probability theory and gives the impression that descriptive accuracy and predictive success in applications are based on (and can be ensured by referring to) an priori-flavored principle that expresses some sort of epistemic indifference about random events. But this is not possible, which is shown by the difficulty (often pointed out in connection with the **Principle of Indifference** [8]) that it is unclear how to specify the precise content of "epistemic neutrality" in such a way that the **Principle of Indifference** does not become circular and holds nevertheless: The **Principle of Indifference** holds only if epistemic neutrality *does* entail that the probabilities of the events given by the uniform probability measure *will* be equal to the frequencies of events in actual trials producing elementary random events, and such a conclusion cannot be validly based on a priori considerations – if it could, the **Principle of Indifference** would have solved the problem of induction.

Note in this connection that Jaynes' [11] resolution of Bertrand's Paradox cannot be interpreted as an instance of deducing empirically correct probabilities from abstract principles in the sense which is relevant for the claims we made above about non-deducibility of empirical frequencies from abstract principles. The logic of Jaynes' treatment of Bertrand's problem is the following:

- A Jaynes shows that certain mathematical conditions entail a unique probability distribution on the two-dimensional plane.
- B He argues that those mathematical conditions should be interpreted as expressing symmetry tacit in the formulation of Bertrand's problem.

- C He shows that there exists an empirical arrangement, a particular way of actually generating an ensemble of events, in which the relative frequencies approximate the probability of the event. (Jaynes was dropping straws on a circle from a ladder.)

A+B+C cannot be regarded as derivation of empirically correct probabilities from non-empirical (a priori) conditions in the relevant sense of “derivation” because Jaynes did not show that the ensemble of events *has to be* generated the way C claims is *possible*. That the ensemble in which the frequencies are equal to the probabilities given by the probability measure Jaynes deduced from symmetries is produced in a particular way (by dropping straws from a ladder) cannot be deduced in a logically valid manner from anything a priori because the actual production of the ensemble in this particular way is a contingent, empirical fact – no conceivable abstract **Principle of Indifference** can validly entail that Jaynes dropped straws from a ladder in a specific way. Jaynes just showed that the mathematical probability theory with the particular probability measure he derived from mathematical symmetry assumptions *can* be a descriptively accurate model of a segment of reality under the frequency interpretation of probability, i.e. that there exist experimentally executable conditions that generate random events in such a way that the symmetry mentioned in B is in fact a feature of the experimental arrangement and that frequencies of events in the ensemble are described by the probability measure he derived. That Jaynes did not *derive* that this *has to be* so is also confirmed by the fact that, as Marinoff [16] shows, one can distinguish different types of random generators representing different types of randomness and, depending on which random generator actually produces the random events featuring in a Bertrand Paradox type situation, one obtains different empirical probability distributions – none of these can be derived from a priori considerations.

Note finally that it also is possible to interpret the General Bertrand’s Paradox presented in this paper as an argument in favor of the claim that probability measure spaces with an uncountably infinite number of random events are too general and abstract to be free of conceptually controversial features. This conclusion might also be supported by referring to the impossibility of a frequency interpretation of probabilities in certain probability measure spaces with “too large” sets of random events and such an interpretation would be in line with the well-known position that rejects countable additivity as a reasonable feature of probability measures interpreted as representing degrees of belief. Given that a wide range of probability measure spaces with uncountably many random events have found very successful applications in the sciences, especially in physics, we would not draw such a radical conclusion however. We prefer the modest interpretation of the General Bertrand’ Paradox: Bertrand’s Paradox shows violation of **Labeling Invariance**, i.e. that re-labelings are not necessarily isomorphisms in the category of Haar probability measure spaces with infinite random events. But the violation of **Labeling Invariance** does not undermine the classical interpretation of probability understood with the **Principle of Indifference**, and violation of **Labeling Invariance** also is in complete harmony with how mathematical probability theory is used in the sciences to model phenomena. In particular, violation of **Labeling Invariance** does not entail violation of **Labeling Irrelevance**; yet, irrespective of Bertrand’s Paradox, the content of the **Principle of Indifference** cannot be specified in such a way that it can establish the classical interpretation of probability (with the frequency link) as descriptively accurate or predictively successful.

## Appendix

This Appendix recalls some elementary facts about the Haar measure. Standard references for the Haar measure are [18] and [9][Chapter XI.], for a more recent presentation see [5].

$X$  is called a topological group with multiplication  $(x, y) \mapsto x \cdot y$  and inverse  $x \mapsto x^{-1}$  if the map  $(x, y) \mapsto x^{-1} \cdot y$  is continuous ( $x, y \in X$ ). A measure  $p$  on the Borel algebra  $\mathcal{S}$  of the group  $X$  is called *left* invariant (respectively *right* invariant) with respect to the group action if eq. (25) (respectively eq. (26)) below hold

$$p(A) = p(xA) \quad \text{for all } x \in X \quad A \in \mathcal{S} \quad (25)$$

$$p(A) = p(Ax) \quad \text{for all } x \in X \quad A \in \mathcal{S} \quad (26)$$

where for an  $x \in X$ , the sets  $xA$  and  $Ax$  are defined by

$$xA = \{x \cdot y : y \in A\} \quad (27)$$

$$Ax = \{y \cdot x : y \in A\} \quad (28)$$

The measure  $p$  is called invariant if it is *both* left *and* right invariant, i.e. if

$$p(A) = p(xA) = p(Ax) \quad \text{for all } x \in X \quad A \in \mathcal{S} \quad (29)$$

On any locally compact topological group there exists both a left  $p_H^L$  and a right  $p_H^R$  invariant Haar measure and they are unique up to multiplication by a constant. The left and right invariant Haar measures are in general different. Since both the left and Haar measure is unique up to constant multiplication, and since for any  $x \in X$  the measure  $p_x(A) \doteq p_H^L(Ax)$  is again a left invariant measure, there exists a real number  $\Delta(x)$  such that  $p_x(A) = \Delta(x)p_H^L(A)$ . The map  $x \mapsto \Delta(x)$  is called the modular function of the group. If  $\Delta(x) = 1$  for all  $x$ , then the groups are called *unimodular*; for unimodular groups the left and right invariant Haar measures coincide and yield an invariant measure. Compact and locally compact abelian groups are unimodular. The Haar measure is bounded if and only if  $X$  is compact – the Haar measure is then a probability measure.

The canonical examples of unbounded Haar measures are the Lebesgue measure on the real line and the Lebesgue measure on  $\mathbb{R}^n$ . It is shown below that the normalized restrictions of the Lebesgue measure on  $\mathbb{R}^n$  to subsets of the form  $\times_i^n [a_i, b_i]$  in  $\mathbb{R}^n$  also can be regarded as Haar measures in their own right with respect to a compact group  $\mathcal{G}$ . This entails that the Lebesgue measure on the *closed* set  $\times_i^n [a_i, b_i]$  also can be viewed as a Haar measure with respect to  $\mathcal{G}$  because the Lebesgue measure space over  $\times_i^n [a_i, b_i]$  and over  $\times_i^n [a_i, b_i]$  are isomorphic. (Note that  $\mathcal{G}$  is *not* the shift; it cannot be since shifted subsets of  $[0, 1)$  are not necessarily subsets of  $[0, 1)$  and the group of “shifts modulo 1” do not form a topological group due to discontinuity of the “shift modulo 1” operation.) Since  $[0, 1)$  can be mapped onto  $[a, b)$  by a continuous linear bijection connecting the (normalized) Lebesgue measures on the intervals  $[0, 1)$  and  $[a, b)$ , to see how the Lebesgue measure on  $[a, b)$  is a Haar measure in its own right, it is enough to see how the (normalized) Lebesgue measure  $p_{[0,1)}$  on the interval  $[0, 1)$  emerges as a Haar measure. Let

$$S^1 = \{z \in \mathbb{C} : |z| = 1\}$$

be the unit circle on the complex plane. As  $S^1$  is a compact topological subgroup of  $\mathbb{C}$  with the multiplication of complex numbers as the group operation, there exists a normalized Haar measure  $p_H$  on  $S^1$ . The exponential function  $f$  defined by

$$f : [0, 1) \rightarrow S^1, \quad f(t) = e^{2\pi it}$$

is a continuous and continuously invertible bijection between the unit interval  $[0, 1)$  and the unit circle  $S^1$ ; hence both  $f$  and its inverse are measurable. We claim that  $f$  is a measure theoretic isomorphism between the interval  $[0, 1)$  with the Lebesgue measure on it and  $S^1$  with the measure  $p_H$  on it; i.e. that

$$p_H = p_{[0,1)} \circ f^{-1} \tag{30}$$

To verify (30), by the uniqueness of Haar measures, it is enough to show that  $p_{[0,1)} \circ f^{-1}$  is a Haar measure, i.e. that  $p_{[0,1)} \circ f^{-1}$  is invariant with respect to the group operation in  $S^1$ , which is the multiplication of complex numbers. Since the exponential function  $f$  turns addition of real numbers into multiplication of complex numbers, for  $B \subset S^1$  and  $z \in \mathbb{C}$  we have

$$f^{-1}(B \cdot z) = f[B] + t \bmod 1 \tag{31}$$

where the translation

$$Y \mapsto Y + t \bmod 1 \tag{32}$$

is the standard shift of set  $Y \subset [0, 1)$  by  $t$  followed by “pulling back” into  $[0, 1)$  the part of  $Y$  that is shifted out of the bounds of  $[0, 1)$ ; formally:

$$Y + t \bmod 1 = (Y \cap [0, 1 - t) + t) \cup (Y \cap [1 - t, 1) - (1 - t))$$

$p_{[0,1)}$  is translation invariant on  $[0, 1)$  in the sense that for any measurable set  $A \subseteq [0, 1)$  and  $0 \leq t < 1$  we have

$$p_{[0,1)}(A) = p_{[0,1)}(A + t \bmod 1),$$

so we have

$$p_H(B \cdot z) = p_{[0,1)}(f^{-1}(B \cdot z)) = p_{[0,1)}(f^{-1}(B) + t \bmod 1) = p_{[0,1)}(f^{-1}(B)) = p_H(B)$$

The Lebesgue measure  $p_{[0,1)}^n$  on the  $n$ -dimensional cube  $[0, 1)^n$  also can be regarded as a Haar measure: one can consider the Haar measure  $p_H^n$  on the  $n$ -dimensional torus

$$T^n = S^1 \times S^1 \times \dots \times S^1 \text{ (} n \text{ times)}$$

which is a compact topological subgroup of  $\mathbb{C}^n$  with the coordinate-wise multiplication of complex numbers as group operation. Put

$$f : [0, 1)^n \rightarrow T^n, \quad f(t_0, \dots, t_n) = (e^{2\pi it_0}, \dots, e^{2\pi it_n})$$

Then  $f$  is a continuous and continuously invertible bijection and, applying the previous argument in each coordinates, one concludes

$$p_H^n = p_{[0,1]}^n \circ f^{-1}$$

## References

- [1] J. Aaronson. *An Introduction to Infinite Ergodic Theory*, volume 50 of *Mathematical Surveys and Monographs*. American Mathematical Society, Rhode Island, 1997.
- [2] S. Bangu. On Bertrand's Paradox. *Analysis*, 70:30–35, 2010.
- [3] J.L.F. Bertrand. *Calcul de Probabilités*. Gauthier-Vilars, Paris, 1888.
- [4] P. Billingsley. *Ergodic Theory and Information*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1965.
- [5] A. Deitmar and S. Echterhoff. *Principles of Harmonic Analysis*. Universitext. Springer, New York, 2009.
- [6] J. Doob. The development of rigor in mathematical probability theory (1900-1950). *American Mathematical Monthly*, pages 586–595, 1996.
- [7] H. Fischer. *A History of the Central Limit Theorem: From Classical to Modern Probability Theory*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer, New York, Dordrecht, Heidelberg, London, 2011.
- [8] A. Hájek. Interpretations of probability. The Stanford Encyclopedia of Philosophy (Summer 2012 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/sum2012/entries/probability-interpret/>, 2012. accessed May 29, 2012.
- [9] P. Halmos. *Measure Theory*. D. Van Nostrand, New York, 1950.
- [10] P. Harremoës. Maximum entropy on compact groups. *Entropy*, 11:222–237, 2009.
- [11] E. Jaynes. The Well Posed Problem. *Foundations of Physics*, 4:477–492, 1973.
- [12] J.M. Keynes. *A Treaties on Probability*. McMillan, London, 1921.
- [13] A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. English translation: *Foundations of the Theory of Probability*, (Chelsea, New York, 1956).
- [14] M. Loève. *Probability Theory*. D. Van Nostrand, Princeton, Toronto, London, Melbourne, 3rd edition, 1963.
- [15] D. Makinson. Conditional probability in the light of qualitative belief change. *Journal of Philosophical Logic*, 40:121–153, 2011.
- [16] L. Marinoff. A resolution of Bertrand's Paradox. *Philosophy of Science*, 61:1–24, 1994.
- [17] J.M. Mikkelsen. A resolution of the wine/water paradox. *The British Journal for the Philosophy of Science*, 55:137–145, 2004.
- [18] L. Nachbin. *The Haar Integral*. D. Van Nostrand, Princeton, NJ, 1965.
- [19] K. Popper. *The Logic of Scientific Discovery*. Routledge, London and New York, 1995. First published in English in 1959 by Hutchinson Education.
- [20] A. Rényi. On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiae Hungaricae*, 6:268–335, 1955.
- [21] J.S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, Singapore, 2006.
- [22] D.W. Rowbottom and N. Shackel. Bangu's random thoughts on Bertrand's Paradox. *Analysis*, 70:689–692, 2010.
- [23] W. Rudin. Autohomeomorphisms of compact groups. *Topology and its Applications*, 52:69–70, 1993.
- [24] N. Shackel. Bertrand's Paradox and the Principle of Indifference. *Philosophy of Science*, 74:150–175, 2007.
- [25] E. K. van Douwen. A compact space with a measure that knows which sets are homeomorphic. *Advances in Mathematics*, 52:1–33, 1984.
- [26] R. von Mises. *Probability, Statistics and Truth*. Dover Publications, New York, 2nd edition, 1981. Originally published as 'Wahrscheinlichkeit, Statistik und Wahrheit' (Springer, 1928).

- [27] J. von Plato. *Creating Modern Probability*. Cambridge Studies in Probability , Induction and Decision Theory. Cambridge University Press, Cambridge, 1994.
- [28] A. Wightman. Hilbert's 6th problem. In F.E Browder, editor, *Mathematical Developments Arising from Hilbert Problems: Proceedings*, volume 28 of *Proceedings of Symposia in Pure Mathematics*, pages 147–240. American Mathematical Society, 1983.