A Study of Mathematical Determination through Bertrand's Paradox

Davide Rizza

*School of Politics, Philosophy, Languages and Communication, University of East Anglia,
Norwich, NR4 7TJ, U.K.*
E-mail: D.Rizza@uea.ac.uk

ABSTRACT

Certain mathematical problems prove very hard to solve because some of their intuitive features have not been assimilated or cannot be assimilated by the available mathematical resources. This state of affairs triggers an interesting dynamic whereby the introduction of novel conceptual resources converts the intuitive features into further mathematical determinations in light of which a solution to the original problem is made accessible. I illustrate this phenomenon through a study of Bertrand's paradox.

1. INTRODUCTION

Mathematical problems often call for the introduction of new concepts or methods because certain intuitive features involved in their formulation cannot be codified by the mathematical apparatus canonically available to study them. In such cases what looks like an inherent difficulty of a given problem is best regarded as an effect of the fact that its intuitive content has not yet been resolved into mathematical determinations that can be relied upon in order to obtain a solution.

This paper aims to explore and clarify this phenomenon with respect to one particular example, namely Bertrand's paradox. The reasons for this choice are threefold. First, Bertrand's paradox is an interesting mathematical problem that has aroused much discussion among both philosophers and mathematicians. Secondly, a recent exchange on the paradox contained in [Rowbottom, 2013] and [Klyve, 2013] can be fruitfully reconsidered in light of the phenomenon that this paper discusses.

Finally, in view of this discussion, it is possible to introduce an elementary approach to Bertrand's paradox itself, motivated by the need to convert certain intuitive features of its geometrical setting into numerical determinations (more plainly, it is necessary numerically to specify the size of certain infinite collections of geometrical entities). This move can be made once the canonical resources of probability theory are supplemented with new computational resources. The next section extracts from the analyses of Rowbottom and Klyve an interpretation of Bertrand's paradox that stresses the incongruity between canonical probability models and the character of this problem. It is because of this incongruity that alternative probability models, to be introduced in Section 3, are required.

2. TWO READINGS OF BERTRAND'S PARADOX

If one were to draw at random a chord in a circle, what is the probability of its being shorter than the side of the inscribed equilateral triangle? This question, originally posed in [Bertrand, 1889], gives rise to a puzzle, generally known as Bertrand's paradox, on account of the fact that it is possible to specify distinct, seemingly equivalent, drawing procedures, each of which determines a distinct value for the sought probability. Bertrand specified three distinct drawing procedures, leading respectively to the probability values 2/3, 1/2 and 3/4. The debate around the existence of a uniquely determined solution has lasted longer than a century and has occupied several authors, such as Borel [1909], Mosteller [1965] and Jaynes [1973].

Recently, the structure of Bertrand's paradox and its interpretation have been helpfully re-examined in [Rowbottom, 2013] and [Klyve, 2013].[1] In my view both papers shed important light on the character of the problem and, more precisely, reveal it to be a problem of mathematical determination, in the sense of Section 1. Rowbottom [2013] points out that Bertrand's proposed solutions are all inapplicable, since none of them takes all possible chords into account. By effectively restricting attention to a designated subcollection of chords, those authors who selected one particular drawing procedure as the correct one (*e.g.*, Jaynes [1973]) could not reach an acceptable conclusion. As Rowbottom points out:

> [. . .] in each case a chord was drawn [. . .] at random *from a proper subset of the possible chords that might be drawn.* [2013, p. 112]

This remark seems to me crucial because it indirectly suggests that Bertrand's question about the probability of drawing a chord at random is to be interpreted as a question concerning the selection of a single chord with specific features

---

[1] Two important contributions on Bertrand's paradox have followed these articles, namely [Aerts and Sassoli de Bianchi, 2014] and [Gyenis and Rédei, 2015]. Their discussion is deferred to the penultimate section, where it will be possible to offer a sufficiently precise appreciation of these works' significance, in view of the full study of Bertrand's paradox provided in Section 3.

from the totality of all chords. Various probability values arise because the type of selection specifiable under any one of Bertrand's drawing procedures may operate on a restrictive ensemble, which is not representative of the random selection of interest. Each procedure leads to the deployment of a continuous, uniform probability distribution and, thus, to a probability model that can be handled for the sake of computing numerical probability values. However, the serviceability of Bertrand's drawing procedures may be at variance with the character of the problem, because all of them come at the cost of focussing on a subcollection of the full collection of chords, which is not the ensemble one intended to study in the first place.

To see how this state of affairs hints at a problem of mathematical determination, consider, by way of comparison, the trivial case of throwing a fair die: in order to specify the probability that the outcome of a throw will be a number strictly smaller than three, it is sufficient to consider the totality of six outcomes and the totality of two outcomes of interest. The probability model implicitly adopted in this case is a uniform, discrete distribution on the space of outcomes resulting from a throw. The totality of outcomes as well as the subset of relevant outcomes can be *numerically* specified and the numerical specifications can then be used to carry out computations of probability values. Bertrand's question about selecting a chord from the totality of all chords mirrors the character of the die problem in an infinite setting. Rowbottom simply points out that the question posed by Bertrand refers to the totality of chords, not a part thereof, and to certain distinctive subcollections of this totality. If probability values are to be computed, the infinite collections involved must be assigned numerical determinations. Following the template of the die model, such determinations should lead to the introduction of a uniform, discrete distribution on the numerically specifiable totality of chords. This approach is not viable if the canonical resources of probability theory are employed but, as will be shown in Section 3, it is accessible to supplementary computational resources.

To sum up the discussion so far, Rowbottom's analysis points to the need for a direct consideration of the totality of chords determined by a circle. If this totality is to be part of a workable probability model, a numerical estimate of its size, with which ordinary arithmetical computations can be carried out, must be available. In other words, an intuitive feature of Bertrand's geometrical setup, *i.e.*, the fact that a circle determines an infinite collection of chords, is to be assigned a mathematical determination, *i.e.*, a numerical specification, which cannot be offered in the canonical (*i.e.*, measure-theoretic) context of probability theory. The possibility of introducing the missing determination depends on an expansion of the mathematical resources at hand: since the resources of probability theory are being used as instruments to intervene on a given geometrical setup, I shall refer to them as a particular mathematical instrumentality, or simply an instrumentality.

Thus, Bertrand's paradox poses a problem of mathematical determination that cannot be solved in presence of the canonical instrumentality of probability theory but may well be solved through the appeal to a distinct instrumentality

(which does not have to be a replacement of the canonical instrumentality, but may be a modification or extension thereof). As will be shown in Section 3, the new instrumentality yields numerical specifications of the restrictions on the ensemble of chords qualitatively alluded to by Rowbottom, but does not rule out the possibility that some of these restrictions should offer adequate characterisations of the problem. Such a question cannot be decided upon in the absence of a suitable mathematical determination of the problem itself. Before offering mathematical support to these remarks, I wish to turn to Klyve's analysis of Bertrand's problem in order to show that it points to the dynamics of determination and instrumentality revealed by Rowbottom's own discussion, albeit from a different point of view and despite the fact that Klyve is critical of Rowbottom's conclusions. Against him, Klyve maintains that Bertrand's drawing procedures are adequate, *i.e.*, actually take all chords into account, in which case:

> [t]he only thing that changes is that the method of selecting one (class of) chord from this set may be biased. [Klyve, 2013, p. 368]

In my opinion Klyve's important contribution does not lie in his intended refutation of Rowbottom but in his focus on what he calls the bias of a procedure, which is best spelled out as lack of mathematical determination and whose source is not so much a selection of drawing method as the resort to a prescribed instrumentality. In short, it seems to me possible at once to vindicate the correctness of Rowbottom's analysis and to extract from [Klyve, 2013] an important lesson, which is independent of the rejection of [Rowbottom, 2013].

   Klyve's critique of Rowbottom is based on a close reading of Bertrand's manner of specifying his drawing procedures. For instance, with respect to the procedure of fixing a diameter and then restricting attention to the chords parallel to it (or, in fact, their intersections with the diameter), Bertrand observes that '[t]he symmetry of the circle means that this information will not affect the probability, either favourably or unfavourably' [Bertrand, 1889, p. 5]. By way of commentary, Klyve notes that, since 'every chord in a circle can be chosen by the expedient of first choosing a radius, and then choosing a perpendicular chord' [Klyve, 2013, p. 367], Bertrand's drawing method does not involve a distorting restriction to a proper subset of the chords in a circle. In fact, this drawing method involves first selecting a radius or, equivalently, a diameter, at random, and then selecting a point from the diameter. Bertrand does not introduce two random variables to model this selection process but assumes that a random selection of the kind he is interested in ultimately corresponds to the selection of a point from an arbitrarily fixed diameter. What one ends up with is a, seemingly viable, restriction to certain chords intersecting the diameter. In the absence of further mathematical determination, it is an open question whether this *de facto* restriction is viable, however plausible it might appear (this issue will be addressed in Section 3.3). Analogously, although Klyve is correct to point out that *any* chord can be determined by specifying a diameter first, he seems to underestimate the fact that Bertrand does not have an

instrumentality that allows him to model the selection of single diameters and to determine whether the possibility of selecting individual diameters may not already suffice to model the selection of every other chord as well, without any later restrictions to an array of points along a diameter superadded. A similar argument applies to Klyve's discussion of the drawing procedure determined by first picking an arbitrary point on a circumference and then drawing from the ensemble of chords through it (whose adequacy will be discussed in Section 3.2).

Klyve's argument does not therefore successfully undermine Rowbottom's insistence on the restrictiveness of the drawing procedures considered by Bertrand: restrictions are eventually imposed, without mathematical determinations that would enable a judgment on their adequacy. Despite the drawback in his argument, Klyve makes a very important observation when he qualifies Bertrand's original intention as follows:

> [h]e wished only to show that the command to choose something at random from an infinite set is too imprecise unless we specify the means of making the choice. [Klyve, 2013, p. 368]

This conclusion must suggest itself if one is an advocate of the absolute validity of the canonical instrumentality of probability theory, which does not afford numerical means to, *e.g.*, count alternatives over an infinite set or deploy a uniform, discrete distribution on it. If, however, one is not an absolute advocate of a prescribed instrumentality, the same conclusion can be read as a call for numerical resources that offer a more precise specification of the command to choose a chord at random. Precisely this call will be answered in Section 3.

Thus, if one accepts Klyve's interpretation of Bertrand's intention, it reveals, from an angle alternative to Rowbottom's analysis, that the canonical instrumentality of probability theory is too imprecise or, in the present terminology, lacks sufficient determination to tackle the problem of selecting a chord at random. Thus, it is best to dismiss Klyve's references to Bertrand's results as effects of biassed drawing procedures that are sufficiently well-determined mathematically. An independent reason for this can be offered by a brief discussion of the numerical example taken from Bertrand, upon which Klyve relies in order to illustrate what he means by bias. The example is presented as a solution to the problem of determining the probability of choosing a number greater than 50 by picking at random in the sample space $\{1, \ldots, 100\}$. Given a uniform, discrete distribution, the answer is trivial, but, since the numbers in the sample space are uniquely determined by their squares, one might also decide to choose over $\{1, \ldots, 10,000\}$, in which case the probability of drawing a number whose square root is greater than 50 (but possibly not an integer) is $3/4$ and not $1/2$, as in the original setup. Klyve qualifies the second problem as a variant of the first in which only the procedure for picking a number has changed, thus introducing a bias. As a matter of fact, the sample space has changed from one scenario to the other and the question being answered is no longer the same (in the second case one is picking at random a number whose square root is greater

than 50 and not a number greater than 50). It is certainly possible to exchange a move to a different sample space with a move to a different distribution over the same sample space $\{1, 2, 3, \ldots, 100\}$, but the non-uniform distribution that gives rise to the probability value 3/4 has been manufactured out of the explicit consideration of a different problem.

The problems in question here are easily distinguishable because sufficient numerical specifications are available to tell them apart. What Klyve calls bias reduces to their discriminability on numerical grounds. This reduction is less straightforward in the context of Bertrand's geometrical problem because there are insufficient numerical resources to identify restrictions and effect discriminations. The same reduction becomes apparent, though, when sharper numerical specifications can actually be used, as will be seen in the next section. What Klyve calls an effect of bias is in fact a problem of mathematical determination: his remarks point in the same direction as Rowbottom's. Under the canonical instrumentality of probability theory, Bertrand's paradox is intimately connected with a lack of mathematical determination. Supplying resources that provide a more sharply determined problem leads to a novel analysis of Bertrand's three drawing procedures and to some surprising conclusions about their agreement. This is the subject of Section 3.

3.   A STUDY OF BERTRAND'S PARADOX

The discussion from Section 2 has primarily served the purpose of identifying Bertrand's paradox as a determination problem: its root is the unavailability of numerical specifications for certain infinitely large collections of chords, from which probability values may be computed. The intuitive idea that a circle determines an infinitely large number of chords, which in turn is the sum of the numbers of chords longer, shorter, and equal to the side of the inscribed equilateral triangle cannot be canonically rendered within a probability model. The goal is then to introduce a new instrumentality under which the numerical specifications being sought can be supplied. This will prove sufficient to set up a probability model that describes the random selection of a chord in a manner free from inadequacies caused by lack of mathematical determination.

The new instrumentality is obtained by supplementing the existing apparatus of concepts and techniques in probability theory with the computational methodology recently introduced by Yaroslav Sergeyev (see in particular his [2003; 2009a; 2009b]). In other words, the fundamental notions of probability theory (*e.g.*, sample space, distribution, random variable, *etc.*) are not jettisoned, but made to interact with computational resources that extend their purview.

Sergeyev's approach may be regarded as an infinitary extension of numerical analysis, whereby it becomes possible to introduce numerical approximations of the sizes of infinite collections or the length of infinite processes. In the present context, a numerical estimate of the collection of all chords determined by a circle will be relied upon. Once this is available, the probability of the events that interested Bertrand can be computed to a degree of accuracy, which can be improved depending on the needed level of precision. In what

follows, probability models are accurate enough to fix the finite part of relevant probability values. The starting point used to introduce infinite numerical estimates is to employ numerical measures of infinite collections for which traditional ordered-field arithmetic holds. This is necessary in order to compute numerical probability values. Moreover, the required measures must be able to discriminate between an infinite collection and its infinite subcollections. This is necessary in order to keep track, in a computationally effective way, of the restrictions to infinite subcollections of chords involved in Bertrand's drawing methods.

It is important to realise that the two *desiderata* just listed call for measures alternative to Cantorian cardinals, which abrogate the principle that strict subsets always have smaller measure than the sets including them. Ordinals are unsuitable for the same reason. Moreover, in both cases ordinary arithmetical laws fail:[2] in other words, computational drawbacks and identification between part and whole make an appeal to Cantorian ideas unsuitable for supplying the kind of mathematical determination required by infinite probability models. It is mandatory to look for a 'counting' measure that is computationally effective and reinstates the general principle that the part should be smaller than the whole. These conditions are met by Sergeyev's approach.[3]

Sergeyev's informal approach consists in drawing a distinction between infinite collections, most notably $\mathbb{N}$, and the numerals that refer to their elements and to the sizes of their parts. In presence of this distinction, it is natural to think that a richer numeral system than one relying on a finite base should support size discriminations between infinite parts of a collection, not only between finite ones. The desired enrichment is obtained by introducing a suitable base for the richer numeral system, which, given the goal at hand, can only be infinitely large. Sergeyev's numeral system works with the infinite base ① (read: gross-one), which is intended to refer to the number of items in the infinite collection $\mathbb{N} = \{1, 2, 3, \ldots\}$. Then ① denotes an infinitely large integer, greater than the natural numbers representable in a finite base. The purpose of introducing ① is not merely to denote a specification of the 'level' of infinity attained by the set of natural numbers, but to increase the discriminability of 'levels' in a way that vindicates the principle that the whole should be greater than

---

[2] For an illuminating discussion of this fact in the context of a construction of 'counting systems' for infinite sets alternative to those proposed by Cantor, see [Benci and Di Nasso, 2003, pp. 50–53].

[3] There are other ways of introducing systems of measures that extend to infinite collections the whole-part relation typical of finite ones and, in addition, are supported by sufficiently rich algebraic structure. A remarkable instance is provided by the numerosities of [Benci and Di Nasso, 2003]. Their approach is not equivalent to Sergeyev's. To see this, it suffices to note that, as will be shown below, on Sergeyev's approach the sets of even natural numbers and of multiples of three are assigned different numerical measures. Benci and Di Nasso work with labelled sets and numerosity assignments are sensitive to the choice of labelling. In particular, there is a (non-canonical) labelling under which the last two sets can be assigned the same numerosity. This possibility is ruled out in Sergeyev's framework, which does not require choices of labellings.

the part. Thus, for instance, the set $\mathbb{N} \cup 0$ has a number of elements denoted by $① + 1 > ①$ and the set $\{2, 3, 4, \ldots\}$ has a number of elements denoted by $① - 1 < ①$. Moreover, as pointed out above, it is assumed that the familiar laws of field arithmetic extend to a notation for elements of the real field that includes terms expressible by means of the symbol $①$. In this setting, the terms $① + 1, ① + 2, ① + 3, \ldots, 2①, \ldots, 3①, \ldots, ①^2, \ldots$ all denote infinitely large reals not in $\mathbb{N}$, which can be summed and multiplied in the usual manner. Multiplicative inverses satisfy identities like $①^0 = ①^{1-1} = ① \cdot ①^{-1} = \frac{①}{①} = 1$ and $(① - 3) \cdot (① - 1)^{-1} = \frac{①-3}{①-1} = 1 - \frac{2}{①-1}$, which will be used in section 3.1 below. If one sought to develop arithmetic using, *e.g.*, ∞ or $\aleph_0$, the above expressions would inevitably contain indeterminate forms.

For present purposes, field arithmetic based on $①$ is not enough, because it does not, on its own, allow enough numerical discriminations of size. In order to compute the number of, say, chords in a circle that are as long as the side of an inscribed equilateral triangle, it will prove necessary to rely on a divisibility property that Sergeyev also postulates. Divisibility amounts to the assumption that any partition of $\mathbb{N}$ into n disjoint arithmetic progressions, with n finite, should have cells containing the same number of elements, denoted by $①/n$.[4] Note that $①/n$, as the evaluation of the size of an infinite aggregate, denotes an infinitely large natural number since $①/n < ①$. It follows that the partition of $\mathbb{N}$ into the two disjoint progressions of odd and even numbers determines two cells containing the same infinitely large number of items, denoted by $①/2$. In a similar vein, the numerical specification of the collection of all multiples of three is $①/3 < ①/2$.

It is worth remarking that these ideas have been formalised in [Lolli, 2015], within the context of a conservative extension of second-order, predicative Peano arithmetic.[5] Lolli's idea is to work with models of arithmetic that contain infinitely large elements and to fix an infinitely large 'cut-off' point, denoted by $①$, intuitively intended to single out $\mathbb{N}$ within a larger model. Axioms governing a suitable measure guarantee that, given an initial segment of a model, *e.g.*, the set of all items satisfying $x < ①$, every subset thereof has a measure. Measures so defined identify (bounded) sets in one-to-one correspondence and enforce the principle that the whole should be greater than the part. Divisibility axioms guarantee that there is a sufficiently rich family of measures that are actually computable and can be expressed using Sergeyev's numeral system. Computability of measures guaranteed by divisibility will play a crucial role in the discussion of Bertrand's paradox to follow. In order to apply Sergeyev's computational methodology to it, it is necessary to decide how the chords in a circle should be parametrised and what probability distribution is to be imposed upon

---

[4] More precisely, $①/n$ denotes the number of elements of any arithmetical progression of the form $k, k + n, k + 2n, \ldots$, with $1 \leq k \leq n$ and k, n finite. Once n is fixed, letting k increase from 1 to n, one obtains a partition of $\mathbb{N}$ into n progressions.

[5] The same treatment is possible on the basis of first-order Peano arithmetic, at the cost of cumbersome numerical coding. A discussion of this matter can be found in [Lolli, 2015, p. 9].

them. In the next section the choice of an adequate parametrisation and distribution will be first motivated and then used to compute probability estimates.

3.1.   A Counting Argument

Let $\mathcal{C}$ be a circle of unit radius in $\mathbb{R}^2$. In order to describe the random selection of a chord from the collection of all chords in $\mathcal{C}$, I shall adopt a strategy that preserves the spirit of Bertrand's treatment without requiring the restrictions imposed by his drawing methods.

Since each of Bertrand's methods relies on a uniform distribution, I shall set up a probability model based on a uniform distribution. Because Sergeyev's methodology makes it possible to count chords, I am going to employ a uniform discrete distribution, as opposed to Bertrand's uniform continuous distributions. The choice of a uniform discrete distribution is also motivated by the parallel between Bertrand's setup and the throw of a fair die discussed in Section 2.

As for the parametrisation of chords, one of Bertrand's drawing methods describes them uniquely by the pairs of their endpoints, one of which is a fixed point on the boundary of $\mathcal{C}$. As will become clear in the following subsections, this is the only parametrisation, among the three proposed by Bertrand, that can be retained without having to restrict attention to a proper subcollection of the full collection of chords in $\mathcal{C}$. In the presence of Sergeyev's methodology, which presupposes the standpoint of numerical analysis, the parametrisation of chords by pairs of distinct points on the boundary of $\mathcal{C}$ depends on a preliminary specification of the number of discriminable points. It is of the essence to realise that, when handling the computational instrumentality proposed by Sergeyev, there is no question, in general, of obtaining exact numerical results: in what follows only approximate probability values or probability estimates are computed. This is, however, enough to restrict the range of inaccuracy to an infinitely small order of magnitude.

The degree of accuracy selected to deal with Bertrand's problem is fixed as soon as it is declared, by means of a numerical specification, how many points on the boundary of the circle $\mathcal{C}$ can be discriminated. In general, if the numeral system adopted includes the symbol n, denoting a natural number, then the partition of $\mathcal{C}$ into equal arcs of length $2\pi/n$ makes it possible to discriminate distinct points on the boundary of $\mathcal{C}$ by assigning them distinct labels from the list $\{1, 2, \ldots, n\}$. This may not be very helpful if one can only end up with finitely many discriminable points, but it becomes a fruitful approach if an infinitely large number of discriminations can be effected. An obvious, but fruitful, choice is to set n equal to ① (greater, infinitely large, numbers could also be chosen, depending on the required level of accuracy[6]). A numerical

---

[6] One could, *e.g.*, pick ①^2 or ①^3, both of which are evenly divided by 3, by divisibility, a fact on which the argument that follows relies. One could even consider $3 \cdot 10^{①}$ discriminable points, which measures the continuum $[0, 3)$, if one deploys a numeral system based on decimal expansions with ① places, each of which is filled by one digit from the list

$$①$$ 1   2

$$① - 1$$ 3

$$\cdots$$ $$\cdots$$

$$\frac{3①}{4} + 1$$ $$\frac{①}{4} + 1$$

$$\cdots$$ $$\cdots$$

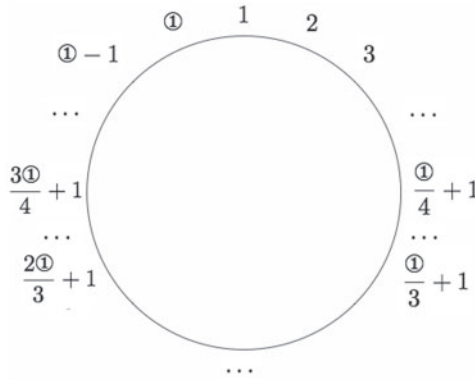$$\frac{2①}{3} + 1$$ $$\frac{①}{3} + 1$$

$$\cdots$$

Fig. 1.   Labelled points around $\mathcal{C}$.

specification of discriminable points leads to a direct computation of the number of discriminable chords.

As Figure 1 shows, this computation is based on the subdivision of $\mathcal{C}$'s boundary into least discriminable arcs marked by ① equally spaced, labelled points. A discriminable chord is uniquely determined by a pair of labelled points on the circumference. Once a labelled endpoint is fixed, $① - 1$ discriminable chords through it may be counted. As one ranges through the ① labelled endpoints, $①(① - 1)$ discriminable chords are counted, but each chord is counted twice, since the two distinct orderings of its labelled endpoints are counted as distinct chords. As a consequence, the total number of discriminable chords is the infinitely large integer denoted by the term:

$$\frac{①^2 - ①}{2}.$$

The last numerical specification, inexpressible in a traditional numeral system and, thus, within the canonical instrumentality of probability theory, goes some way towards addressing the determination problem identified in Section 2. Whereas it was not possible, under any of Bertrand's drawing methods, to rely on a numerical specification of the full collection of chords in $\mathcal{C}$, an infinite estimate of the number of discriminable chords is now available. In its presence, it is possible to introduce a uniform, discrete distribution on the sample space of discriminable chords. Such a distribution assigns each chord the infinitely small probability $2/(①^2 - ①)$. A simple probability model is now in place, which leads to the computation of probability estimates for the events we are interested in.

$\{0, 1, 2, \ldots, 9\}$. In this case, $10^{①}$ points are discriminable on $[0, 1)$ and three times this number on $[0, 3)$.

In particular, let $P(e), P(s), P(l)$ be, respectively, the probability of selecting a discriminable side of some equilateral triangle inscribed in $\mathcal{C}$, the probability of selecting a shorter chord, and the probability of selecting a longer chord. The problem set by Bertrand is to evaluate $P(s)$. Since $1 = P(e) + P(s) + P(l)$ once $P(e), P(s)$ are computed, a value for $P(l)$ can also be determined.

In order to compute $P(e)$, it is convenient to count first the number of points that lie on any arc subtended by the side of some equilateral triangle inscribed into $\mathcal{C}$. Since the whole arc has length $2\pi/3$ and two consecutive, discriminable points are separated by an arc of infinitesimal width $2\pi/①$, there are $①/3$ least discriminable arcs covering one third of the circumference. Note that $①/3$ denotes a natural number, by divisibility. It now follows that an arc of length $2\pi/3$ contains $①/3 + 1$ discriminable points. It is convenient to work with the assignment of labels $1, 2, \ldots, ①-1, ①$ illustrated in Figure 1. Any discriminable side of an inscribed equilateral triangle is uniquely determined when one of its discriminable endpoints is fixed. The other endpoint is identified by summing $①/3$ to the label on the endpoint that has been fixed. The discriminable sides of equilateral triangles inscribed in $\mathcal{C}$ are thus systematically identified by the following pairs of labels:[7]

$$\left(1, \frac{①}{3} + 1\right), \left(2, \frac{①}{3} + 2\right), \ldots, \left(\frac{2①}{3} + 1, 1\right), \left(\frac{2①}{3} + 2, 2\right), \ldots, \left(①, \frac{①}{3}\right).$$

Along this sequence, which has $①$ elements, no pair is counted twice. It is clear that all discriminable pairs are counted, since their endpoints are only assigned labels from $\{1, 2, \ldots, ①-1, ①\}$. It follows that there are $①$ discriminable sides of inscribed equilateral triangles, *i.e.*:

$$P(e) = \frac{2}{①-1}.$$

The value of $P(e)$ just computed is a positive infinitesimal. It would have been inexpressible under the canonical instrumentality of probability theory, which assimilates the selection of the side of an inscribed equilateral triangle to the impossible event (its probability is zero in each of the three scenarios considered by Bertrand).

In order to find $P(s)$, it now suffices to consider the discriminable chords containing $①/3$ or fewer points. There are $①$ times $①/3$ discriminable chords shorter than the side of the inscribed equilateral triangle. When the degenerate ones, consisting of one point, are excluded, only $(①^2 - 3①)/3$ chords are left. Then:

$$P(s) = \frac{①^2 - 3①}{3} \cdot \frac{2}{①^2 - ①} = \frac{①(① - 3)}{3} \cdot \frac{2}{①(① - 1)}$$
$$= \frac{① - 3}{3} \cdot \frac{2}{① - 1} = \frac{2}{3} \cdot \frac{① - 3}{① - 1} = \frac{2}{3} \cdot \left(1 - \frac{2}{① - 1}\right) = \frac{2}{3} - \frac{4}{3(① - 1)}.$$

---

[7] It is perhaps worth noting that no appeals to symmetry, of the kind required in [Jaynes, 1973], are needed in this argument. It suffices to have deployed only a convenient reference frame, in which consecutive numeral labels are attached to consecutive points.
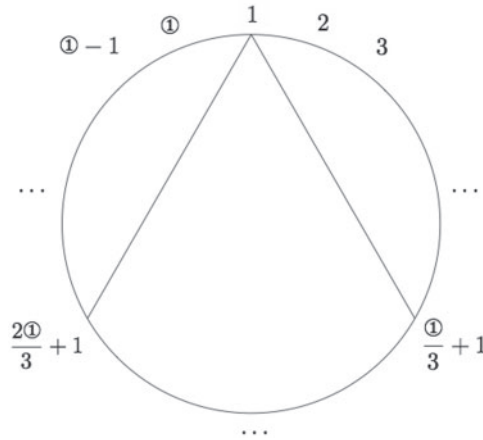
Fig. 2. Chords through a fixed point.

This is an evaluation of the probability requested in the problem set up by Bertrand. It is infinitely close to $2/3$. It is finally possible to compute:

$$P(l) = 1 - [P(s) + P(e)] = 1 - \left(\frac{2}{3} - \frac{4}{3(① - 1)} + \frac{2}{① - 1} \right)$$

$$= 1 - \left(\frac{2}{3} + \frac{2}{3(① - 1)} \right) = \frac{1}{3} - \frac{2}{3(① - 1)},$$

whose finite part is $1/3$. Although the analysis carried out in this section can be refined by more accurate numerical estimates (see fn 5), information about the finite part of the estimates $P(s)$ or $P(l)$ is already available. Furthermore, it is possible to reconsider Bertrand's drawing methods as giving rise to approximations of the probability model introduced in this subsection and assess their adequacy against it. The next subsections are devoted to carrying out precisely this task.

3.2. Selecting Chords Through a Fixed Point

Among the three drawing procedures considered by Bertrand, let us consider first the one that represents the random selection of a chord's endpoint, followed by another endpoint selection, as a *de facto* selection from the collection of chords in $\mathcal{C}$ through an arbitrary fixed point on the circumference. Using the system of ① labels illustrated in Figure 1, we may conveniently fix the point whose numeral label is 1 (see Figure 2).

In this case only $① - 1$ out of $(①^2 - ①)/2$ discriminable chords are being taken into account. Rowbottom's observation to the effect that only a proper subcollection of chords is taken into account can be numerically vindicated. Now let us call $P_1(e), P_1(s), P_1(l)$ the probabilities of selecting a chord respectively equal, shorter, or longer, than the side of an equilateral triangle inscribed in $\mathcal{C}$,

from the ensemble of ① − 1 chords through the point marked by the numeral label 1. As in Section 3.1, the probabilities sought can be computed from a uniform, discrete distribution imposed on the given collection of chords.

Since only the chords determined by the endpoints $(1, ①/3 + 1)$ and $(2①/3+1, 1)$ are sides of inscribed equilateral triangles, we immediately have $P_1(e) = 2/(①−1) = P(e)$. The fact that the last probability is the same as that obtained by the unrestricted method from the previous subsection depends on the fact that, under the given drawing procedure, the relative proportions of types of chords are preserved, although their numbers are scaled by the infinitesimal factor $2/①$.

This is confirmed by the computation of $P_1(s)$, which can be determined by noting that there are $①/3 − 1$ discriminable chords from 1 to the consecutive vertices labelled by $2, 3, 4, \ldots ①/3 − 1, ①/3$ respectively: all of these chords are, among those singled out by the drawing procedure, shorter than the side of an equilateral triangle inscribed in $\mathcal{C}$. Since the chords are symmetrically distributed relative to the diameter through 1, the total number of chords shorter than the side of an equilateral triangle inscribed in $\mathcal{C}$ is $2(①/3 − 1)$, *i.e.*, the number for the chords shorter than the side of an equilateral triangle obtained in Section 3.1, scaled by the factor $2/①$. As a result:

$$P_1(s) = \frac{2(① − 3)}{3} \frac{1}{① − 1} = \frac{2}{3} \frac{① − 3}{① − 1} = \frac{2}{3} - \frac{4}{3(① − 1)} = P(s).$$

It follows that $P_1(l) = P(l)$. The drawing method just examined is a scaled version of the full model constructed in the previous subsection and thus exhibits no discrepancy relative to it. Compared with the same model, Bertrand's original treatment of this drawing method leads to the finite part of $P(s)$ and underestimates the infinitesimal part by setting it equal to zero.

In light of the new instrumentality applied so far, one is led to the interesting conclusion that the restriction of a full model for Bertrand's problem obtained by focussing on a proper part of the collection of all discriminable chords does not *per se* lead to an inadequate model: this is because scaled models are as good as the full model. In this respect, one may differ from Rowbottom's conclusion that Bertrand's drawing methods are all inapplicabile because they restrict attention to a sample space that does not include all chords. In a similar manner, one may agree with Klyve's remark to the effect that considering chords through a fixed endpoint does not undermine the validity of a model for Bertrand's problem. The last conclusions, however, become possible and meaningful only once the new instrumentality used here affords a subtler assessment of modelling choices.

3.3. Parallel Chords

Let us now turn to the drawing procedure that corresponds to the selection of a diameter and then a chord perpendicular to it, which is represented by Bertrand as the *de facto* selection of a chord from the ensemble of those perpendicular
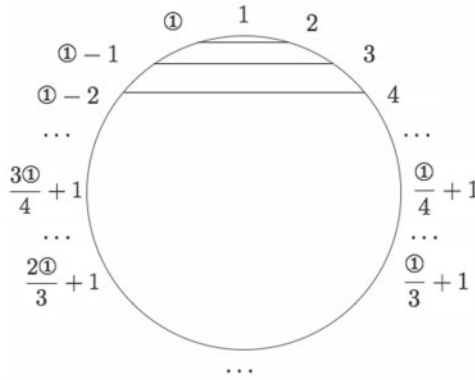
Fig. 3. Parallel chords.

to a fixed diameter. In the present context, this selection restricts the independently given ensemble of discriminable chords to those perpendicular to a fixed diameter and, thus, parallel to one another.

Let $P_2(e), P_2(s), P_2(l)$ be the probabilities of selecting a chord respectively equal, shorter, or longer than the side of an equilateral triangle inscribed in $\mathcal{C}$ from the restricted ensemble. In order to study this case as an approximation of the model from Section 3.1 — based on a suitable uniform, discrete distribution — it is convenient to focus on the chords perpendicular to the diameter through the point labelled by 1 (see Figure 3).

Among these chords, the single one that is also a diameter has endpoints marked by the numerals $①/4 + 1$ and $3①/4 + 1$. The first of these numerals is identified by observing that the arc of length $\pi/2$ traced clockwise from the point labelled by 1 is covered by $①/4$ least arcs and must therefore contain $①/4+1$ points. The second numeral is obtained by a similar argument. Elementary geometry shows that the chords with endpoints labelled by the following pairs: $(2, ①), (3, ① - 1), (4, ① - 2), \ldots, (①/4 + 1, 3①/4 + 1)$, are all perpendicular to the diameter that has been fixed. Because these chords are counted by the consecutive labels from 2 to $①/4 + 1$, there are $①/4$ of them, including the diameter between $①/4 + 1$ and $3①/4 + 1$. The same situation arises in the lower semicircle from Figure 3. As a consequence, the total number of chords determined by the drawing procedure is $①/2 - 1$ (the diameter in this ensemble being counted only once).

It is clear that, among the parallel chords, only two can be sides of an inscribed equilateral triangle (one of them has endpoints labelled by $5①/6 + 1$ and $①/6 + 1$ and subtends the arc containing these points as well as the point labelled by 1; the other is the reflection of the first in the diameter parallel to both). We can therefore compute:

$$P_2(e) = 2\frac{2}{① - 2} = \frac{4}{① - 2}.$$

In order to determine $P_2(s)$, note that each semicircle in $\mathcal{C}$ contains ①$/6 - 1$ chords perpendicular to the fixed diameter and shorter than the side of an inscribed equilateral triangle. The pairs $(2, ①), (3, ① - 1), (4, ① - 2),$ $\dots, (①/6, 5①/6 + 2)$ determine the relevant chords in one semicircle (the next pair of endpoints in this list determines the side of an equilateral triangle).

It can be deduced that the whole circle contains ①$/3 - 2$ chords, along the given direction of parallelism, that are shorter than the side of an inscribed equilateral triangle. Thus:

$$P_2(s) = \frac{① - 6}{3} \frac{2}{① - 2} = \frac{2}{3} - \frac{8}{3(① - 2)}.$$

Then, clearly:

$$P_2(l) = 1 - (P_2(e) + P_2(s)) = \frac{1}{3} - \frac{4}{3(① - 2)}.$$

Relative to the combinatorial argument from Section 3.1, the values obtained for this drawing procedure exhibit an infinitesimal discrepancy of order ①$^{-1}$ because the chords connecting consecutive discriminable points are systematically neglected. They were, on the contrary, included in the counts from Sections 3.1 and 3.2 (in the latter case, one of the consecutive points had to have the numeral label 1.). Nevertheless, the finite parts of $P(s), P_1(s)$ and $P_2(s)$, as well as those of $P(l), P_1(l)$ and $P_2(l)$, are the same.

Bertrand's original treatment of the drawing method just discussed leads to the probability value $1/2$. With the new instrumentality employed so far, the same value can be simulated, up to a discrepancy of order ①$^{-1}$, by setting up a probability model for the random selection of a point from a diameter, once it is declared that ① $+ 1$ points can be discriminated along a fixed diameter. It is worth emphasising that Bertrand could not offer a numerical model for the selection of single diameters (or directions), when describing his drawing methods. He thus resorted to the assumption that the draw ultimately reduces to picking a chord perpendicular to a diameter. In the presence of sharper numerical determinations, the random selection of a diameter can be explicitly described, and it gives rise to an ensemble of endpoints around the circle that suffices to set up the model from Section 3.1.

In the presence of this model, the selection of a chord from those perpendicular to an arbitrary diameter is entirely describable without superadding chords assigned to a uniform distribution of discriminable points along a particular diameter. That such superaddition introduces a distortion is revealed by the fact that, given a partition of the boundary of $\mathcal{C}$ into equal arcs, the discriminable chords orthogonal to a fixed diameter will not partition it into equal intervals. When the distortion is removed, probability values that are finitely accurate can still be obtained.

3.4. Selecting Midpoints of Chords

Let c designate the centre of $\mathcal{C}$ and let $\mathcal{C}'$ be $\mathcal{C}$ with c removed. Any point x in the interior of $\mathcal{C}'$ determines the unique chord perpendicular to the radius through x and c, of which x is the midpoint. Exploiting this fact, one might hope to reduce the selection of chords to the selection of points in the interior of $\mathcal{C}$. In view of the discussion from the previous subsection, one should be wary of identifying a probability model for the random selection of an interior point from $\mathcal{C}$ with a probability model for the random selection of a chord.

One may, however, require that discriminable interior points be midpoints of discriminable chords, in which case the model set up in Section 3.1 can be adopted. Its introduction does not immediately allow one to focus on discriminable interior points only, though, because, even if it is possible to identify each point in the interior of $\mathcal{C}'$ with the unique chord of which it is the midpoint, this identification breaks down for the centre c of $\mathcal{C}$, the midpoint of a continuum of diameters. The effect of identifying infinitely many diameters with their common midpoint cannot be evaluated under the canonical instrumentality of probability theory and, thus, there is no way of telling whether it fundamentally distorts the sought probability values.

In presence of Sergeyev's computational methodology, however, a numerical estimate of the distortion can be obtained by looking again at the model from Section 3.1 and taking the discriminable interior points of $\mathcal{C}$ to be the midpoints of discriminable chords determined by pairs of labelled endpoints. In this case the centre of $\mathcal{C}$ is the common midpoint of ①/2 discriminable diameters. The discriminable midpoints in the interior of $\mathcal{C}$ are therefore:

$$\frac{①(①-1)}{2} - \left(\frac{①}{2} - 1\right) = \frac{①(①-2)+2}{2}.$$

Since a diameter is longer than the side of an inscribed equilateral triangle, the number of discriminable midpoints of chords shorter than such a side is the same as in Section 3.1, namely $①(①-3)/3$. Calling $P_3(s)$ the probability of selecting the midpoint of a chord shorter than the side of an inscribed equilateral triangle, it is now easy to compute:

$$P_3(s) = \frac{①(①-3)}{3} \frac{2}{①(①-2)+2} = \frac{2}{3} - \frac{2}{3}\left(\frac{1 - \dfrac{2}{①}}{① - 2 + \dfrac{2}{①}}\right),$$

whose discrepancy from $P(s)$ is of order $①^{-1}$, *i.e.*, finite agreement holds.

Thus, given a preliminary specification of the number of discriminable diameters, their identification with c does not affect a probability estimate if only accuracy of order $①^0$ is required. However, in order to reach this conclusion, a numerical specification of the infinite collection of discriminable diameters is to be given, and this cannot be done by declaring a distribution of points in the interior of $\mathcal{C}$ alone, but only by specifying the totality of discriminable chords as

was done in Section 3.1. The viability of the drawing method based on interior points is subject to the introduction of the parametrisation adopted in Section 3.1. When this drawing method is viable, it must be in finite agreement with the results obtained for the original model.

Bertrand's application of the same drawing method, however, leads to the probability value $3/4$ for $P_3(s)$. The latter value can be simulated under Sergeyev's instrumentality, up to an infinitesimal error, by a model that specifies the discriminable points in the interior of $\mathcal{C}$ by taking ① discriminable points on a fixed radius and then assuming that the circle through the nth discriminable point ($1 \leq n \leq$ ①) contains n discriminable points. This model imposes a uniform distribution upon a collection of points that are not homogeneously spread over the interior of $\mathcal{C}$: fewer and fewer points are discriminable as one approaches its centre. The model also describes a random choice unrelated to the selection of a chord from an ensemble homogeneously distributed around the circle, which was dealt with in Section 3.1.

In view of the previous subsections, it is possible to conclude that, when such homogeneous distribution is fixed as the geometrical configuration of reference, the drawing methods proposed by Bertrand are in finite agreement and only generate infinitely small discrepancies. If, on the other hand, one replaces the geometrical configuration attached to the parametrisation of chords as pairs of labelled endpoints with other geometrical ensembles (points on a diameter, interior points), which in turn lead to distinct random selection processes, then probability values proliferate.

4. TWO CANONICAL RESOLUTIONS

In Section 2, I argued that a satisfactory approach to Bertrand's paradox requires an expansion of the canonical instrumentality of probability theory. In Section 3, I have shown that, under the expansion afforded by Sergeyev's computational methodology, a numerical treatment of Bertrand's paradox can be given, under which the three drawing methods are in finite agreement when regarded as approximations to a model that describes the random selection of a chord from the totality of all chords in a circle $\mathcal{C}$. Two recent papers present results that seem to be at variance with these conclusions.

On the one hand, [Aerts and Sassoli de Bianchi, 2014] purports to provide a resolution of Bertrand's paradox by canonical means and obtains the value $1/2$ for $P(s)$. On the other hand, [Gyenis and Rédei, 2015] defuses Bertrand's paradox by offering a mathematical account of the proliferation of probability values as an unproblematic phenomenon. In the next two subsections I shall explain why the resolution proposed by Aerts and Sassoli de Bianchi is unsatisfactory and in what way the analysis provided by Gyenis and Rédei is not only consistent with the study of Bertrand's paradox articulated in Section 3 but indirectly confirms it.

4.1. Averaging Over Drawing Methods

Aerts and Sassoli de Bianchi draw a distinction between an easy problem and a hard problem raised by Bertrand's paradox. The easy problem is to

figure out why the fact that distinct probability values arise from distinct chord selection procedures does not contradict the principle of indifference (very roughly, the principle that no particular selection outcomes are more likely to occur than others). The hard problem is to obtain a uniquely determined value for $P(s)$. In order to tackle the easy problem, Aerts and Sassoli de Bianchi note that the question posed by Bertrand (stated at the beginning of Section 3) admits of distinct empirical reifications, up to a certain degree of idealisation.

In particular, one may concretely mark off the extent of a chord by throwing a stick on a circular surface (provided the stick always falls on the surface). The observation, found in [Rowbottom, 2013], concerning the restrictiveness of Bertrand's drawing methods is translated by Aerts and Sassoli de Bianchi into the remark that these methods are not satisfactory models of the random throw of a stick. This conclusion is consistent with the analysis of Section 3, which pointed out in what ways some of Bertrand's probability values can be simulated by describing random selection processes distinct from the selection of a chord.

As for the hard problem, Aerts and Sassoli de Bianchi start from the observation that the probability values generated by Bertrand's methods are to be seen as particular, biassed values, constrained by a certain restriction on the drawing procedure.

In view of this, they go on to produce a clever construction that allows them to obtain a universal mean of selection procedures, each represented by a density function that is the limit of suitable step functions. They finally identify the value of their universal mean, namely $1/2$, with $P(s)$. The correctness of their mathematical treatment does not lead to a value for $P(s)$ because the objects averaged over do not encode faithful information about the geometrical character of Bertrand's problem.

This is true in the light of Aerts's and Sassoli de Bianchi's own remarks, but it is even more apparent when the results of Section 3 are taken into account. These results showed not only that some drawing procedures distort the original random selection problem, when modelled as geometrical selections other than a direct selection of chords (*e.g.*, as the random selection of a point from a segment), but also that Bertrand's drawing procedures lead to the same finite estimate of $P(s)$, if they are regarded as restricted versions of a random selection process on an ensemble of homogeneously distributed chords. Aerts and Sassoli de Bianchi cannot avail themselves of the last conclusion, because it is inaccessible from the point of view of the canonical instrumentality of probability theory. Thus, they seem willing to accept that there is nothing better to do than averaging over arbitrary distortions of the original selection problem. This strategy may be forced upon one in possession of exclusively the canonical instrumentality, but it must be dismissed once the insights provided by an application of Sergeyev's computational methodology are available.

4.2. Defusing the Paradox

Gyenis and Rédei frame their discussion of Bertrand's paradox in the context of what they call the elementary classical interpretation of probability theory. The motivation for this approach is the common interpretation of the paradox as a violation of the principle of indifference. Gyenis and Rédei show that what is involved in Bertrand's paradox is not a violation of the principle of indifference but a violation of a distinct property called labelling invariance.

Their argument begins with the observation that it is easy to satisfy a condition of neutrality, which corresponds to the principle of indifference, on finite sample spaces, by imposing uniform, discrete distributions upon them. Under the canonical instrumentality of probability theory, infinite sample spaces do not allow the introduction of uniform, discrete distributions but it is possible to reinstate neutrality by a suitable addition of topological structure.[8]

If one then works within a suitable category of topologically enriched measure spaces, a version of the principle of indifference can still be satisfied, and Bertrand's paradox continues to hold. Its occurrence does not therefore violate a neutrality condition, but turns out to violate another property, namely labelling invariance.

In order to understand what this is, note that, for finite $X = \{x_1, \ldots, x_n\}$, it amounts to the fact that the probability of an event $A \subseteq X$ is not altered by a reassignment of numerical indices, *i.e.*, a relabelling. When X is infinitely large, relabellings must be defined as measurable bijections with measurable inverses. Invariance then amounts to the fact that any relabelling is a measure-theoretic isomorphism between two spaces *describing the same phenomenon*.[9] Its violation is then nothing but the fact that the probability value $P(s)$ is not preserved across Bertrand's probability models. In view of Section 3, this kind of violation may be regarded as a pointer to differences between the probability models that cannot be fully detected by the canonical instrumentality. In other words, it is an indicator of the level of canonical discriminability between these models.

Under the enriched instrumentality supplied by Sergeyev's computational methodology, discrimination power does not only increase, but becomes mathematically informative. This is because, if one tries to simulate the probability values obtained by Bertrand using Sergeyev's computational methodology, one ends up with probability models whose sample spaces do not contain the same number of elements (*e.g.*, the first drawing method takes into account ①-1 discriminable chords whereas the second drawing method takes into account ①$+1$ discriminable points).[10]

---

[8] Details are not important here but may be found in [Gyenis and Rédei, 2015, pp. 355–356].

[9] The qualification in italics is explicitly assumed by Gyenis and Rédei. For a rigorous definition of labelling invariance, see [Gyenis and Rédei, 2015, pp. 357–358].

[10] Different numbers of chords are obtained even when Bertrand's drawing methods are described as approximations to the random selection from Section 3.1.

Even if these models are based on sample spaces that are classically indistinguishable relative to size, if one relies on the finer numerical distinctions of size afforded by the numeral system based on ①, it becomes clear that what looked like indistinguishable collections cannot in fact be related by bijections. In effect, as was argued in Section 3, the probability models that simulate Bertrand's distinct probability values may not even be seen as models of the same phenomenon. This difference was not visible under the canonical instrumentality other than as a failure of labelling invariance, whereas it becomes transparent after the shift to a numerically more expressive instrumentality is carried out.

It follows that Gyenis and Rédei do not only provide a subtle analysis of Bertrand's paradox within a canonical context, but also pinpoint the canonical property whose failure corresponds to an actual differentiation between models once Bertrand's problem is endowed with the canonically missing numerical determinations. It is noteworthy that, under Sergeyev's methodology, labelling invariance holds, either because there are no bijections joining the relevant spaces or because a straightforward generalisation of this notion for a finite sample space is available.

5.   SUMMARY

This paper explored one aspect of mathematical thinking, which is equally significant in a pure and an applied context, and may be referred to as the dynamics of determination and instrumentality. Certain mathematical problems, as well as mathematised empirical problems, occur within an enquiry as objects of investigation calling for symbolic instruments adequate to their character and, thus, capable of tackling them. It may well be the case that a canonical array of instruments should prove insufficient to carry out a successful intervention upon a problem, in which case the forging of new instruments is required if progress in enquiry is to be made.

Bertrand's paradox nicely illustrates a situation in which canonical instruments are not effective because they cannot render into computationally serviceable terms certain features of the problem at hand, namely numerical specifications of infinite collections of chords. Once new computational instruments, such as those coming from Sergeyev's methodology, are introduced, greater insight into the paradox can be gained and certain difficulties produced by resort exclusively to the canonical instrumentality of probability theory are overcome.

REFERENCES

Aerts, D., and M. Sassoli de Bianchi [2014]: 'Solving the hard problem of Bertrand's paradox', *Journal of Mathematical Physics* **55**, 083503, http://dx.doi.org/10.1063/1.4890291.

Benci, V., and M. Di Nasso [2003]: 'Numerosities of labelled sets: A new way of counting', *Advances in Mathematics* **173**, 50–67.

Bertrand, J. [1889]: *Calcul des probabilités*. Paris: Gauthier-Villars.

Borel, E. [1901]: *Éléments de la théorie des probabilités*. Paris: Hermann et Fils.

Gyenis, Z., and M. Rédei [2015]: 'Defusing Bertrand's paradox', *British Journal for the Philosophy of Science* **66**, 349–373.

Jaynes, E.T. [1973]: 'The well-posed problem', *Foundations of Physics* **3**, 477–493.

Klyve, D. [2013]: 'In defense of Bertrand: The non-restrictiveness of reasoning by example', *Philosophia Mathematica* (3) **21**, 365–370.

Lolli, G. [2015]: 'Metamathematical investigations on the theory of Grossone', *Applied Mathematics and Computation* **255**, 3–14.

Mosteller, F. [1965]: *Fifty Challenging Problems in Probability.* Reading, Mass.: Addison-Wesley.

Rowbottom, D. [2013]: 'Bertrand's paradox revisited: Why Bertrand's "solutions" are all inapplicable', *Philosophia Mathematica* (3) **21**, 110–114.

Sergeyev, Ya. D. [2003]: *The Arithmetic of Infinity.* Rende: Edizioni Orizzonti Meridionali.

———— [2009a]: 'Numerical computations and mathematical modelling with infinite and infinitesimal numbers', *Journal of Applied Mathematics and Computation* **29**, 177–195.

———— [2009b]: 'Numerical point of view on calculus for functions assuming finite, infinite, and infinitesimal values over finite, infinite, and infinitesimal domains', *Nonlinear Analysis Series A: Theory, Methods and Applications* **71**, e1688–e1707.