Bertrand’s Paradox Revised

First formulated in the 19th century, Bertrand’s paradox rose to prominence as a refutation of the classical conception of probability and the principle of indifference (Keynes 1921; von Mises 1957; van Fraassen 1989; Gillies 2000). Despite its influence, the paradox remains poorly understood with continued and erroneous emphasis on the probability formalism, real-valued intervals, and uniform distributions. The first section of this paper presents two famous instances of Bertrand’s paradox: Circles and Chords from Bertrand (1889) and The Mystery Cube Factory from van Fraassen (1989). The second section outlines the standard analysis of the paradox and its shortcomings with respect to these instances. In the third and final section, I present a revised analysis of the paradox in terms of a conflict between the intuition that bijections preserve relative sizes and the actual behavior of intuitive measures. This revised analysis succeeds in both capturing all accepted instances of Bertrand’s paradox as well as generating several new ones. More importantly, however, it repositions Bertrand’s paradox as a paradox of infinity and pinpoints the mechanism underlying the paradox.

§1 Two Instances of Bertrand’s Paradox

1.1 Circles and Chords

Although Bertrand presented his paradox in a number of different settings, his most well-known formulation is geometric. Bertrand (1889, 4) begins with an apparently well-defined question:

Given a circle of radius \( r \), how likely is it that an arbitrary\(^1\) chord on this circle will be longer\(^2\) than the side of an inscribed equilateral triangle?

---

\(^1\)In Bertrand’s original writings, he writes of taking a chord “au hasard” which is typically translated as “at random”. I have substituted “arbitrary” instead to highlight the availability of several distinct interpretations for both this challenge and the paradox that follows. If, for example, “arbitrary” is interpreted as code for a particular distribution of physical propensities, then the resulting paradox pertains to physical propensities of this form. If “arbitrary” is interpreted instead as shorthand for a particular distribution in the mathematical theory of probability (e.g., the uniform distribution), then the paradox pertains to the mathematical theory of probability. If “arbitrary” is interpreted only as ruling out additional evidence about the chord selected, then the paradox pertains to absolute confirmation. Contemporary philosophy recognizes both a number of distinct notions of likelihood and a number of formal representations for these notions. Bertrand’s paradox belongs to no single notion or formalism, so I opt for neutral terminology whenever possible.

\(^2\) Bertrand (1889) actually asks after chords shorter than the side of an inscribed equilateral triangle. Contemporary discussion of the paradox, however, almost exclusively concerns chords that are longer, and so I too adopt this formulation of the paradox.
A chord on a circle is a line segment each of whose endpoints lie on the circle; an example is pictured in Figure 1 below. The side-length of the unique (up to rotation) equilateral triangle which can be inscribed in a circle with radius $r$ is $\sqrt{3}r$, pictured in Figure 2 below.

![Figure 1](image1.png)  ![Figure 2](image2.png)

Bertrand proposes three methods by which the desired likelihood could be found. Each proposal leverages a small amount of geometry in order to characterize both all available chords and chords with length greater than $\sqrt{3}r$ in a more intuitively tractable format.³

Circles and Chords - Distance from Origin to Chord Midpoint

A chord $AB$ is longer than $\sqrt{3}r$ if and only if the distance from the origin $O$ to the midpoint of $AB$, labeled $M_{AB}$ in figure 3, is between 0 and $\frac{r}{2}$ (compare with the inscribed equilateral triangle in Figure 2). Since the length between $O$ and $M_{AB}$ can be anywhere in the interval $[0, r]$, exactly half of all the possible lengths result in a chord longer than $\sqrt{3}r$. Since $AB$ was arbitrary, a chord longer than $\sqrt{3}r$ will then occur with likelihood $\frac{1}{2}$!

![Figure 3](image3.png)

³Rowbottom (2013) and Shackel (2007) rightly observe that there is cause for worry with all three of Bertrand’s proposals. Each of the proposed methods treats a subset of the available chords and then lifts this treatment to the total space. No guarantee of the adequacy of such a method is ever explicitly provided, and in some cases the operation is highly unintuitive, e.g., in the third method, the centerpoint of the circle corresponds with uncountably many chords while other points correspond to only one (Rowbottom 2013).
Circles and Chords - Angle from Tangent

For any point \( A \), we can inscribe an equilateral triangle with \( A \) itself as one of the triangle’s points. The endpoint \( B \) of the chord is entirely determined by the angle \( \theta \) between \( A \) and the line tangent to the circle at \( A \) as in Figure 4. Since the sides of the inscribed equilateral triangle occur at 60\(^\circ\) and then again at 120\(^\circ\) from the tangent, the chord \( AB \) is longer than \( \sqrt{3}r \) if and only if \( \theta \) is strictly between 60\(^\circ\) and 120\(^\circ\). Since \( \theta \) can be drawn from anywhere in the interval \([0^\circ, 180^\circ]\), exactly one third of all possible angles result in a chord longer than \( \sqrt{3}r \). A chord longer than \( \sqrt{3}r \) will then occur with likelihood \( \frac{1}{3} \).

Circles and Chords - Area of Midpoint Regions

Just as above, a chord \( AB \) is longer than \( \sqrt{3}r \) if and only if the distance from the origin \( O \) to the midpoint of \( AB \) is less than \( \frac{r}{2} \) (compare with Figure 2). Since any point within the circle \( C \) is the midpoint of some chord and only those points within the smaller circle \( C' \) of radius \( \frac{r}{2} \) are midpoints of chords longer than \( \sqrt{3}r \), we need only take the area of \( C' \) divided by the area of \( C \):

\[
\frac{\text{Area}(C')}{\text{Area}(C)} = \frac{\pi(\frac{r}{2})^2}{\pi r^2} = \frac{\frac{1}{4}r^2}{r^2} = \frac{1}{4}
\]

A chord longer than \( \sqrt{3}r \) will then occur with likelihood \( \frac{1}{4} \).

Each of Bertrand’s proposed methods thus generates a distinct value. Since the requested likelihood cannot be simultaneously \( \frac{1}{2} \), \( \frac{1}{3} \), and \( \frac{1}{4} \), at most one of these methods is legitimate.

1.2 The Mystery Cube Factory

The best-known instance of Bertrand’s paradox is The Mystery Cube Factory first elaborated by van Fraassen (1989):

The Mystery Cube Factory - Length.

A factory produces perfect cubes with side-length \( \leq 2 \) cm. Given this, how likely is it that a cube produced by this factory has side-length \( \leq 1 \) cm?
The expected answer is that the requested value is $\frac{1}{2}$. If cubes may have any side-length less than 2 cm, then those side-lengths less than 1 cm intuitively make up half the total interval.

Of course, the description in terms of side-length above isn’t necessary; the mystery cubes at issue also have areas, volumes, and so forth. Consider a similar question then with area instead of side-length:

*The Mystery Cube Factory - Area.*

A factory produces perfect cubes with faces whose areas are $\leq$ 4 cm$^2$. Given this, how likely is it that a cube produced by this factory has a face with area $\leq$ 1 cm$^2$?

The intuitive answer is now that the requested likelihood is $\frac{1}{4}$. If cubes may have any area less than 4 cm$^2$, then those with area less than 1 cm$^2$ intuitively make up a quarter of all possibilities.

The paradox is sealed by noticing that a cube with side-length less than 2 cm is a cube whose faces have area less than 4 cm$^2$. Similarly, a cube with side-length less than 1 cm is a cube whose faces have area less than 1 cm$^2$. The intuitive answers of $\frac{1}{2}$ and $\frac{1}{4}$ are thus inconsistent.

§2 The Standard Analysis of Bertrand’s Paradox

Discussion of Bertrand’s paradox has largely focused on resolving particular instances rather than characterizing the underlying phenomenon. One of the few commentators to diverge from this trend is Keynes (1921, 52):

In general, if $x$ and $f(x)$ are both continuous variables, varying always in the same or in the opposite sense, and $x$ must lie between $a$ and $b$, then the probability that $x$ lies between $c$ and $d$, where $a < c < d < b$, seems to be $\frac{d-c}{b-a}$, and the probability that $f(x)$ lies between $f(c)$ and $f(d)$ seems to be $\frac{f(d)-f(c)}{f(b)-f(a)}$. These expressions, which represent the probabilities of necessarily concordant conclusions, are not, as they ought to be, equal.
Nearly a century later, Bangu (2009, 31) provides a nearly identical analysis:

One begins with a variable $x$ uniformly distributed in an interval $[a, b]$ and then one considers a scaling transformation $\theta$ such that $x' = \theta(x)$. For a fixed value $c$ such that $a \leq c \leq b$, two questions (or ‘problems’, as Bertrand calls them) are formulated:

Q1: What is $p_1 = P(x \in [c, b], \text{if } x \text{ is random in } [a, b])$?

Q2: What is $p_2 = P(x' \in [c', b'], \text{if } x' \text{ is random in } [a', b'])$?

These two questions are said to be ‘identical’, so they should receive the same answer. However, if we calculate the probabilities using [the principle of indifference] (and the standard Lebesgue measure of the intervals), we notice that $p_1 = \frac{|b - c|}{|b - a|}$ is different from $p_2 = \frac{|b' - c'|}{|b' - a'|}$, for many transformations $\theta$, and values of $a, b$ and $c$.


Let $I_{a,b}$ denote an interval from $\mathbb{R}$ of the form $[a, b], (a, b), (a, b)$, or $(a, b)$ with $a < b$. The common core for these accounts can be summarized as:

**Standard Analysis of Bertrand’s Paradox:**

Every instance of Bertrand’s paradox invokes a space of possibilities $\Omega$ and an event $E \subseteq \Omega$ together with

(i) A bijection $g : \Omega \to I_{a,b}$ such that $g[E] = I_{c,d}$

(ii) A bijection $f : I_{a,b} \to I_{f(a),f(b)}$ (or $I_{f(b),f(a)}$) such that $f[I_{c,d}] = I_{f(c),f(d)}$ (or $I_{f(d),f(c)}$)

where

(iii) $\frac{|d-c|}{|b-a|} \neq \frac{|f(d)−f(c)|}{|f(b)−f(a)|}$.

Further, in so far as this structure clearly obtains, we likewise have an intuitive instance of Bertrand’s paradox.
On this analysis, the paradox is induced by representing the space of possibilities $\Omega$ in terms of two intervals $I_{a,b}$ and $I_{f(a),f(b)}$ wherein some single event $E$—represented by the two subintervals $I_{c,d}$ and $I_{f(c),f(d)}$—receives two different “sizes”, namely $|d-c|$ and $|f(d)-f(c)|$.

Applying the analysis to the *The Mystery Cube Factory*, the space of possibilities $\Omega$ is the set of cubes which can be produced by the factory, viz. cubes with side-length $\leq 2$ cm and face-area $\leq 4$ cm$^2$. The event $E$ in question is then the desired subset of these cubes, viz. cubes with side-length $\leq 1$ cm and face-area $\leq 1$ cm$^2$. Note that both sets are well-defined. Problems arise, however, when this space and event are viewed under two intuitive bijections, $g$ and $f \circ g$. The first of these bijections, $g$, maps the set of possible cubes to their side-lengths in $[0, 2]$. The second bijection, $f \circ g$, maps the set of possible cubes to their face-areas in $[0, 4]$. To complete the diagram, note that side-lengths and face-areas are themselves related by an intuitive bijection $f(x) = x^2$. Finally, the size of $E$ relative to all possible cubes varies depending on which of $g$ and $f \circ g$ is used, satisfying condition (iii) of the standard analysis.

As Rowbottom and Shackel (2010) have also observed, *Circles and Chords* does not conform to the standard analysis. Of his three methods, only Bertrand’s last—*Area of Midpoint Regions*—appears to supply a bijection between the set of all possible chords and another space. This other space, however, is $\mathbb{R}^2$ not $\mathbb{R}$, and Bertrand uses relative area rather than relative interval length to generate the associated likelihood. To make matters still more complicated, there is only the appearance of a bijection here because, while any point in a circle besides the center is the midpoint of a unique chord, the center of a circle is the midpoint for uncountably many chords (Shackel 2007). Even if we emphasize appearance and overlook this failure to construct a bijection, Bertrand’s third method still breaks with the standard analysis by using $\mathbb{R}^2$ rather than $\mathbb{R}$.

*Circles and Chords* thus greatly complicates the status of the standard analysis. Taking all three methods seriously suggests that the standard analysis ought to be generalized along two distinct dimensions. First, Bertrand’s use of relative area in his third method suggests that the paradox ought to extend beyond the relative length of intervals. Second, Bertrand’s apparent use of partitions in his first and second methods suggests that the paradox does not require explicit bijections for the entire space of possibilities. In both cases, the structure outlined by the standard analysis is overly restrictive.

---

4The existence of a well-defined set of possibilities (or sample space) is sometimes missed by commentators, e.g., Burock (2005) and Pettigrew (2016). Nevertheless, this is a key part of paradox. A situation wherein different likelihoods are ascribed over different sets of possibilities is unremarkable; the contradictory solutions are troubling precisely because the underlying space and event are fixed.

5While it is possible to construct a superficially similar situation which does fit, Bertrand clearly makes use of premises which deviate from those of the standard analysis. Since paradoxes (or instances of paradoxes) are individuated in part by their premises, constructing clever bijections cannot save the standard analysis here. It is important not that (i)-(iii) could be satisfied for $\Omega$ and $E$ but rather that (i)-(iii) are the premises which lead to paradox.
§3 Revised Analysis of Bertrand’s Paradox

Bertrand’s paradox has long been intimately associated with continuous intervals, relative interval lengths, and the mathematical theory of probability. Despite this long association, none of these are an essential feature of the paradox. Rather, the core of the paradox is a conflict between the intuition that bijections ought to preserve relative sizes and a host of particular measures which do not. The two generalizations suggested by Circles and Chords are well taken.

Infinite collections have long been associated with paradox. Zeno’s paradoxes are a famous case from antiquity but difficulties continued throughout the middle ages and well into the 20th century. Galilei (1638) provides a classic example with the intuitive attractiveness of both (1) and (2):

(1) Natural numbers are more numerous than squares of natural numbers.
(2) There are just as many squares of natural numbers as there are natural numbers.

The initial plausibility of (1) stems from the part-whole intuition, the conviction that proper parts are smaller than the whole. Since \( \{0, 1, 4, 9, 16, \ldots \} \subset \{0, 1, 2, 3, 4, \ldots \} \)

it ought to be that the natural numbers are strictly larger than set of all squares of natural numbers. (2) meanwhile is supported by a similarly appealing bijection invariance intuition, the conviction that bijections between sets preserve size. Since squaring natural numbers

\[
\begin{array}{cccccccc}
0 & 1 & 2 & 3 & 4 & \ldots \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
0 & 1 & 4 & 9 & 16 & \ldots 
\end{array}
\]

clearly produces a bijection between the natural numbers and their squares, these sets ought to be the same size. As Mancosu (2009) argues at length, the proper response to this situation is to recognize that both part-whole and bijection invariance encode consistent intuitions about size but that the move to infinite sets requires a choice between them.

Bertrand’s paradox is a similar situation only for the relative size of sets. Suppose we wish to know how the relative size of a set \( E \) compares to the relative size of a set \( E' \) where \( E, E' \subseteq \Omega \). If \( \Omega \) is not itself equipped with a salient notion of relative size, it is natural to represent \( \Omega \) in terms of a set which is so equipped and thereby obtain our desired comparison. Bertrand’s paradox appears when two or more of these representations contradict one another.
Revised Analysis of Bertrand’s Paradox:

Every instance of Bertrand’s paradox invokes a space of possibilities $\Omega$ and events $E, E' \subseteq \Omega$ together with

(i’) A set $U_1$ and notion of relative size $\lesssim_1$ such that either

(a) there exists a bijection $g_1 : \Omega \to U_1$ or

(b) there exists a partition $P$ of $\Omega$ such that, for any choice of $P_i \in P$, there is a corresponding bijection $g_1^i : P_i \to U_1$, and for any $i, j$, both

$$g_1^i[E \cap P_i] = g_1^i[E \cap P_j]$$

$$g_1^i[E' \cap P_i] = g_1^i[E' \cap P_j]$$

(ii’) A set $U_2$ and notion of relative size $\lesssim_2$ such that either

(a) there exists a bijection $g_2 : \Omega \to U_2$ or

(b) there exists a partition $P$ of $\Omega$ such that, for any choice of $P_i \in P$, there is a corresponding bijection $g_2^i : P_i \to U_2$, and for any $i, j$, both

$$g_2^i[E \cap P_i] = g_2^i[E \cap P_j]$$

$$g_2^i[E' \cap P_i] = g_2^i[E' \cap P_j]$$

where

(iii’) $\lesssim_1$ and $\lesssim_2$ do not agree on the relative sizes of $E$ and $E'$ (under $g_1$ and $g_2$ respectively).

Further, in so far as this structure clearly obtains, we likewise have an intuitive instance of Bertrand’s paradox.
More precisely, Bertrand’s paradox is the result of two representations \( \langle U_1, \leq_1 \rangle \) and \( \langle U_2, \leq_2 \rangle \) which engender conflicting relative size intuitions over two sets \( E \) and \( E' \) drawn from a common superset \( \Omega \). A set \( U \) and notion of relative size \( \leq \) may qualify as a representation either by means of a bijection \( g : \Omega \to U \) or by means of a collection of uniform bijections from the parts in some partition of \( \Omega \) to \( U \). A conflict between representations can be constructed whenever relative size is not preserved under bijection, viz. when it is not the case that \( E \leq E' \) if and only if \( g[E] \leq g[E'] \) for any bijection \( g : \Omega \to \Omega' \).

All of the standard instances of Bertrand’s paradox can be explained on the revised analysis. The Mystery Cube Factory, for example, presents a space of possibilities \( \Omega \) containing all cubes of side-length \( \leq 2cm \) (and thus also face-areas \( \leq 4cm^2 \)) and then asks after the relative size of a particular subset \( E \), the cubes with side-length \( \leq 1 \) cm and face-areas \( \leq 2 \) cm\(^2 \). The paradox arises when these sets of cubes are distilled into just lengths \( (g_1[E] \text{ and } U_1) \) and areas \( (g_2[E] \text{ and } U_2) \) respectively, representations which possess intuitive measures of relative size which vary under bijection. The underlying comparative nature of the disagreement can be seen by comparing \( E \) with its complement \( E' \), the set of cubes with a side-length of \( 1 - 2 \) cm and face-area of \( 1 - 4 \) cm\(^2 \). In the side-length representation, \( g_1[E'] \leq_1 g_1[E] \) but in the face-area representation \( g_2[E'] \not\leq_2 g_2[E] \). By construction, the result is a pair of inconsistent relative size judgments despite clear bijections.

The adequacy of the revised analysis requires not only a satisfactory treatment for accepted instances of Bertrand’s paradox but also that further instances of the revised analysis give rise to further instances of the paradox. To motivate this latter claim, I present a new instance of Bertrand’s paradox but also that further instances of the revised analysis give rise to construction, the result is a pair of inconsistent relative size judgments despite clear bijections.

For an infinite collection \( Y \) of natural numbers and \( X \subseteq Y \), one intuitive notion of relative size is given by a minor generalization of natural density:

\[
\mu_{\text{rel}}(X, Y) = \lim_{n \to \infty} \frac{|X_{\leq n}|}{|Y_{\leq n}|}
\]

for \( n \in \mathbb{N} \) where

\[
X_{\leq n} = \{ x \in X \mid x \leq n \}
\]

\[
Y_{\leq n} = \{ y \in Y \mid y \leq n \}.
\]

It is this notion of size which leads us to say, for instance, that even numbers make up half the natural numbers—\( \mu_{\text{rel}}(\text{Even}, \mathbb{N}) = \frac{1}{2} \)—while multiples of four make up one quarter of the naturals and one half of all evens—\( \mu_{\text{rel}}(\text{MultipleOf4}, \mathbb{N}) = \frac{1}{4} \text{ and } \mu_{\text{rel}}(\text{MultipleOf4}, \text{Even}) = \frac{1}{2} \).

Since \( \mu_{\text{rel}} \) ascribes different sizes to the set of even numbers and the set of multiples of four relative to \( \mathbb{N} \), \( \mu_{\text{rel}} \) does not preserve relative size under bijection. This can be immediately leveraged

---

6In the case where \( U_1 \) and \( U_2 \) are equipped not just with comparative size relations \( \leq_1 \) and \( \leq_2 \) but quantitative measures \( \mu_1 \) and \( \mu_2 \) over a shared scale, explicit reference to a second set \( E' \) can be dropped and (iii') collapsed into a disagreement over the relative size of \( E \), namely \( \mu_1(g_1[E]) \neq \mu_2(g_2[E]) \). While the standard instances of Bertrand’s paradox all use quantitative measures of relative size, purely comparative variations of, for example, The Mystery Cube Factory are easy to construct.
into an intuitive instance of Bertrand’s paradox:

**Tupperware Factory.** A factory produces tupperware with a natural number $n$ printed on the lid and another $m$ printed on the bottom. Further, every natural number appears exactly once in each position. How many of the tupperware have an even number on their lid? How many of the tupperware have a multiple of four on their bottom?

The intuitive responses (à la $\mu_3$) are that evens make up exactly $\frac{1}{2}$ of all possible lid values, and so the first likelihood is $\frac{1}{2}$. It is similarly plain that multiples of four make up exactly $\frac{1}{4}$ of all possible bottom values, and so the second likelihood is $\frac{1}{4}$.

Bertrand’s paradox now appears when it is revealed that the factory pairs lids and bottoms in the somewhat odd fashion described by $f$:

$$f(x) = \begin{cases} 2x & x = 2n \text{ for } n \in \mathbb{N} \\ n + 1 + (n + 1) \text{ Div } 4 & x = 2n + 1 \text{ for } n \in \mathbb{N}. \end{cases}$$

The function $f$ pairs all even numbered lids with bottoms bearing a multiple of four and shuffles all odd numbered lids to the remaining bottoms:

$$\begin{array}{cccccccc}
0 & 1 & 2 & 3 & 4 & 5 & \ldots \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
0 & 1 & 4 & 2 & 8 & 3 & \ldots
\end{array}$$

As a result, the set of tupperware with an even numbered lid is the set of tupperware with a bottom bearing a multiple of four, and so the intuitive values of $\frac{1}{2}$ and $\frac{1}{4}$ contradict one another. $\mu_3$ has been brought into conflict with the clear bijection between tupperware lids and tupperware bottoms.

While the revised analysis greatly expands the scope of Bertrand’s paradox, not all intuitive measures of relative size are susceptible. The most obvious example of a relative measure which is immune to the paradox is counting elements in finite sets. For a finite set $\Omega$ and $E \subseteq \Omega$, define

$$\mu_{\text{fin}}(E, \Omega) = \frac{|E|}{|\Omega|}.$$ 

Unlike previous examples, this measure is fixed under any bijection $g : \Omega \to \Omega'$ and thus satisfies the bijection invariance intuition. Supposing that the counting measure $\mu_{\text{fin}}$ is the only acceptable metric for relative size over finite sets, Bertrand’s paradox cannot then arise in a finite setting. An infinite set of possibilities is an essential component of the paradox, and so Bertrand’s paradox can be properly accounted a paradox of infinity.

Like Galilei (1638)’s classic paradox of infinity, Bertrand’s paradox dramatizes the inconsistency of two intuitive alternatives. On the one hand, it is natural to think that relative sizes—just like sizes—ought to be preserved under bijection. On the other, many intuitive measures of relative size do not, in fact, preserve relative size under bijection. Among these latter are not only relative interval length in $\mathbb{R}$ and relative area in $\mathbb{R}^2$ but also relative interval length in $\mathbb{Q}$ and natural density.
As a result, Bertrand’s paradox is best viewed as a paradox of infinity for relative sizes with only a tangential connection to attributions of likelihood. While the resolution of the paradox has ramifications for discussions of rational credence, confirmation, and probability, the paradox itself generalizes well beyond all three.

References


