November 8, 9, 2017; June 10, 21, November 8, 2018; January 19, 2019.

DRAFT

Chapter from a book, *The Material Theory of Induction,* now in preparation.

Uncountable Problems

John D. Norton[1]
Department of History and Philosophy of Science
University of Pittsburgh
http://www.pitt.edu/~jdnorton

1. Introduction

The previous chapter examined the inductive logic applicable to an infinite lottery machine. Such a machine generates a countably infinite set of outcomes, that is, there are as many outcomes as natural numbers, 1, 2, 3, … We found there that, if the lottery machine is to operate without favoring any particular outcome, the inductive logic native to the system is not probabilistic. A countably infinite set is the smallest in the hierarchy of infinities. The next routinely considered is a continuum-sized set, such as given by the set of all real numbers or even just by the set of all real numbers in some interval, from, say, 0 to 1.

It is easy to fall into thinking that the problems of inductive inference with countably infinite sets do not arise for outcome sets of continuum size. For a familiar structure in probability theory is the uniform distribution of probabilities over some interval of real numbers. One might think that this probability distribution provides a logic that treats each outcome in a continuum-sized set equally, thereby doing what no probability distribution could do for a countably infinite set. That would be a mistake. A continuum-sized set is literally infinitely more complicated than a countably infinite set. If we simply ask that each outcome in a continuum-sized set be treated equally in the inductive logic, then just about every problem that arose with the countably infinite case reappears; and then more.

[1] My thanks to Jeremy Butterfield for a close reading of this chapter that led to many corrections.

This chapter will explore what sorts of inductive logics can implement uniformity of chance over an outcome set of continuum size. The notion of uniformity used is label independence, as already developed in the previous chapter. To start, we will presume the outcome set is “bare,” that is, it has no further structure beyond its continuum size. Then, in Section 2 below, we shall see that label independence imposes on it an inductive logic something like the infinite lottery machine inductive logic, but with more sectors. This is an unfamiliar logic, remote from a probabilistic logic.

If we seek a sense of uniformity of chance compatible with a probabilistic logic, we must weaken the requirement of label independence. It will be weakened in successive sections in three stages. In Section 3, the unrestricted requirement of label independence is weakened by requiring that the independence holds only for permutations that preserve a σ-field of subsets of a continuum-sized outcome set. This is a natural first step, since probability measures in continuum sized outcome sets are standardly only defined over such subsets. We will find that this weakening is insufficient. A probability measure fails to conform with the weakened requirement of label independence. The failure is not remedied by a further weakening that only allows permutations that are involutions. The applicable logic turns out to be akin to that of the completely neutral support of Chapter 9.

In Section 4, label independence will be further weakened by assuming that the continuum outcome set has its own metrical structure, commonly the metrical geometry of a space. The permutations of label independence are restricted to those that preserve areas or volumes of this metrical geometry. This weakened version of label independence is, finally, compatible with a probabilistic logic: it is one that matches probabilities with the space’s areas or volumes.

The success, however, proves limited. For if the metrical space is infinite in area or volume, a probabilistic logic cannot provide uniformity of chances. It is easy to see that a metrically adapted label independence requires that this uniformity be expressed by the same inductive logic that applies to the infinite lottery machine. This inductive logic is the one that applies to the stochastic process of continuous creation of matter in Bondi, Gold and Hoyle’s steady state cosmology. Its application to this case is teased out in enough detail to return some curious results.

That this last inductive logic is applicable is demonstrated by decomposing the space into infinitely many parts. The parts are then reassembled in a way that respects the background metrical structure of the space, but precludes an additive measure. This construction is one of the simplest of a corner of mathematics that explores "paradoxical decompositions." This literature is introduced in Section 5. It has explored more thoroughly the difficulties faced when we seek to use additive measures to gauge the size of sets in a metrical space. The construction of Section 4 employed a decomposition into infinitely many parts. If our space had hyperbolic geometry, then a remarkable construction reported by Wagon (1994) shows that similar results can be achieved by decomposing the space into just three parts each of infinite measure.

This literature in paradoxical decompositions is the locus of nonmeasurable sets. These are sets in a metrical space to which no area or volume can be assigned consistently. While the difficulties for probability measures have so far arisen only in metrical spaces of infinite area or volume, these nonmeasurable sets become problematic for probability measures that match the areas and volumes of spaces with finite total area of volume. For such a probability measure will fail to assign a value to these nonmeasurable sets. Since these nonmeasurable sets impose a fundamental limitation on the use of probability measures in such spaces, they will be pursued in the remainder of the chapter.

Section 6 will review the simplest example, a Vitali set. Since a Vitali set is metrically nonmeasurable, it is beyond the reach of a probability measure adapted to the spatial metric. Instead, the chance that some outcome of a random process will be found in a Vitali set is shown to follow a familiar inductive logic, that of the infinite lottery machine. This section also discusses the awkwardness that nonmeasurable sets are not constructible by the means normally employed in set theory. Rather their existence is posited by the axiom of choice.

Finally in Section 7, I recount a nonmeasurable set described by Blackwell and Diaconis (1996) that comes closer to the sorts of systems commonly treated in accounts of inductive inference. It is a probabilistically nonmeasurable outcome set that arises with infinitely many coin tosses. In Section 8, I show that there is a weak inductive logic native to the example that I call an "ultrafilter logic."

Overall, this investigation shows that, in many cases for a continuum sized outcome set, a probabilistic logic fails to apply. Other, non-probabilistic logics do apply locally to the specific problem posed. To recount them, they appear as:

Section 3.6, Section 4.2, Section 6.2: variations on an infinite lottery machine logic.

Section 8: an ultrafilter logic.

2. The Inductive Logic of Uniform Chances in a Bare Continuum

How might an inductive logic provide equal support or equal chances to every outcome in a space of continuum size? To answer, we need to specify the applicable notion of equality or uniformity of chances. That condition was developed in the previous chapter. An infinite lottery machine selects among a countable infinity of numbers fairly, that is, without favoring any. Each of the infinity of outcomes was assigned a unique number label. The fairness of the lottery was expressed in the condition:[2]

Label independence

All true statements pertinent to the chances of different outcomes remain true when the labels are arbitrarily permuted.

That individual outcomes have equal chance is secured through propositions like:

Outcomes numbered "37" and "18" have the same chance.

The statement remains true no matter how we redistribute number labels across the outcomes. This indifference to the labels assigned to individual outcomes can only come about if all outcomes have the same chance. It is otherwise with statements like:

Outcome number "37" has greater chance than outcome number "18."

This statement cannot remain true under a relabeling that switches labels "37" and "18," assuming that the relation of "greater chance" is asymmetric.

The same applies to sets of outcomes:

The odd numbered set of outcomes has the same chance
as the even numbered set of outcomes.

This statement remains true no matter how we may permute the number labels over the outcomes. Once again, this indifference of the sets to the numbers that label their elements can only come

[2] Here and below, a permutation is a one-one map on the label set or, correspondingly, on the outcome set. In the previous chapter, these sets were countable. In conformity with modern usage, the term "permutation" will continue to be used when the label of outcome set is continuum sized. The term is synonymous with bijection.

about if the two sets have the same chance. From similar statements, it follows that two sets of outcomes have the same chance just in case there is a permutation of the number labels that reassigns the numbers labeling the first set to the second set.

We now apply label independence to an outcome set of continuum size. We saw in the previous chapter that the chance values assigned to sets of outcomes of an infinite lottery machine drawing were divided into two sectors, the finite sector and the infinite sector. Replicating the procedure of the previous chapter for the new case of a continuum-sized outcome set, we find a similar, but more complicated structure, with three sectors. In the continuum case, the chance of an outcome in various outcome sets has the indicated values and associated informal interpretation:

Finite set of outcomes of size n:

A countable infinity of values, $V(n)$, $n = 1, 2, 3, \ldots$; "very unlikely."

Countably infinite set of outcomes:

one value only, *V*(countably infinite); "unlikely."

Continuum-sized infinite set of outcomes:

For an outcome set of continuum size and whose complement is continuum sized,

one value only, *V*(continuum-co-continuum); "as likely as not."

For an outcome set of continuum size and whose complement is countably infinite,

one value only, *V*(continuum-co-countable); "likely."

For an outcome set of continuum size and whose complement is finite,

V(continuum-co-finite *n*), $n = 1, 2, 3, \ldots$; "very likely."

The strength of support grows as we move down this list. The distance between the sectors is very great since we step up the hierarchy of infinites. We could, presumably, find many results that match those of the infinite lottery machine logic and many more that are not in it, because of its extra structure. However I will pass over this exercise. What matters for our purposes is that the fullest implementation of uniformity in a continuum-sized outcome set leads to a logic that is quite different from a probabilistic logic.

3. Uniformity over a σ-Field of Outcomes

3.1 A Uniform Probability Distribution

The logic of the last section is very different from a probabilistic logic. We were driven to this logic by the requirement of label independence. If we are to find conditions more conducive to a probabilistic logic, we will need to weaken this requirement. To map a pathway for the weakening, we need to see our goal: a uniform probability distribution over a continuum-sized outcome set. Take the especially hospitable case[3] of outcomes labeled by real numbers in the interval [0,1], that is the set of real numbers x, such that $0 \leq x \leq 1$. The uniform probability distribution over this interval is derived from a probability density function

$$p(x) = 1 \tag{1}$$

and it is plotted in Figure 1.

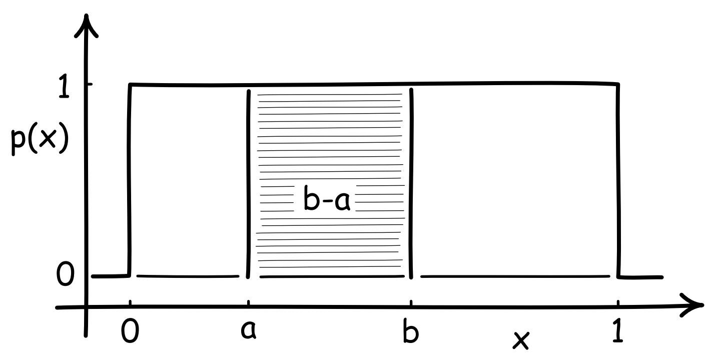

Figure 1. Uniform Probability distribution

We extract probabilities from this probability density for sets of outcomes by computing the corresponding areas under the curve. The probability of an outcome labeled by a real number in the interval [a, b], where $0 \leq a \leq b \leq 1$, is the area shown in the figure and, of course, is equal to b-a.

This distribution certainly *looks* like it is choosing without favor among the continuum sized outcome labeled by [0,1]. The curve in Figure 1 is flat. It is also free of a problem facing a

---

[3] It is hospitable since, otherwise, if either end of the interval extends to infinity, a uniform non-zero probability density over the interval integrates to an infinite probability over the whole interval.

uniform probability distribution over a countably infinite outcome space: there is no countably additive, uniform probability distribution over the set. For such a distribution, each outcome would have to be assigned the same probability. If that value is zero, then their countably infinite sum is also zero, in contradiction with the requirement that the probabilities of all mutually exclusive outcomes must sum to unity. In contrast, the probability density (1) can assign zero probability to each of its continuum many outcomes without a corresponding difficulty. The summation of an uncountable infinity of zeroes is not a well-defined operation in standard probability theory.

In spite of these encouraging signs, the uniform probability distribution fails to implement the requirement of label independence. Take statements such as

(Eq) The probability of events labeled by real numbers in [0, 0.5] is the same as the probability of events labeled by real numbers in [0.5, 1],

Since the permutations admissible under label independence are entirely unrestricted and can scatter the labels about in all imaginable ways, it is easy to see that this and many other statements like it fail to remain true when the number labels permuted. Some restriction on the permutations is needed if label independence is to apply.

3.2 The σ-Field

One of the founding results of modern measure theory is that an additive measure, such as a probability measure, cannot assign a measure to all subsets of points in a space if the space is sufficiently large. Then there are many nonmeasurable sets. In Section 6 below, we shall see the standard example that arises in the interval [0,1] of real numbers, a Vitali set. As a result, probabilities can be defined only for a preferred subset of all the subsets of real numbers in [0,1]. The resulting restriction on the scope of probability measures has been built into the modern mathematical formalism from the outset. Kolmogorov (1950), the *locus classicus* of the modern tradition, introduces the distinction in his definitions. A probability measure is defined in the context of a set of "elementary events." (p. 2, Ch. II) It is, for example, the set of outcomes labeled by real numbers in [0,1]. However a probability is not automatically defined for all subsets of this set. Rather, at the outset, probabilities are defined only for some of these subsets. These are the "random events" that form a field or algebra of sets. That is, the field or algebra is, by definition, closed under the finite union, finite intersection and complement of its members.

When the set of elementary events is infinite, the fields or algebras are required to be σ-fields or σ-algebras. That is, they are closed under countably infinite unions and intersections.

Since a probability measure can assign probabilities only to some of the subsets of elementary event labeled by real numbers in [0,1], those sets have to be identified if the probability measure is to be adequately defined. The standard procedure is to work backwards from those probabilities that we cannot forego. In forming the probability distribution associated with (1), we expect that, whatever else, the probability assigned to all intervals of the form $[a,b]$ above is $b-a$. So we include in the σ-field all intervals of the closed form $[a,b]$ as well as half-open $[a,b)$, $(a,b]$ and and open (a,b).[4] We then require that the σ-field associated with the uniform distribution be one that contains all these intervals and is closed under all countable unions and intersections. It is not obvious that such a field should exist or, if so, that it is unique. Both are assured by the Extension Theorem (Kolmogorov, 1950, p. 17).[5]

3.3 σ-Field Adaptation

The uniform distribution does not assign probabilities to all subsets of the elementary events labeled by real numbers in [0,1]. It follows that the truth of statements concerning subsets of elementary events cannot be preserved under an arbitrary permutation of the numbering of the elementary events used in the statement. The permutation may take a set for which a probability is defined to one that is nonmeasurable. What is a true statement for the original set about its probability may fail to be true when those same number labels are applied to a nonmeasurable set, for the latter set has no probability. Thus the subsets in the σ-field are favored in the sense that a probability is defined for them only. Label independence fails.

If a probability density (1) is to conform with label independence, we need to weaken label independence. A first step in this weakening is to restrict the permutations so that they only map sets of events in the σ-field to sets of events in the σ-field.

σ-Field Adapted Label independence

---

[4] By the usual convention $[a,b)$ contains all x for which $a \le x < b$, *etc*.

[5] See Rosenthal (2006, Ch. 2) for a more expansive introduction to this result of great foundational importance.

All true statements pertinent to the chances of different outcomes remain true when the labels are permuted by all permutations that preserve the sets of the σ-field.

A consequence is that sets of elementary events labeled by some open, half-open or closed interval of real numbers, always remain labeled by such intervals under all permutations to be considered.

3.4 Failure

While σ-field adaptation is a necessary adaptation if the uniform probability density (1) is to be compatible with label independence, it turns out not to be sufficient. The uniform probability density (1) still does not conform with the weakened requirement. The permutations of the weakened requirement are continuous functions on x that invertibly map the interval [0,1] back to [0,1]. The condition of invertibility is essential. Otherwise the function would be redistributing the number labels in such a way that one elementary event is assigned more than one new number label. There are of course *very* many such invertible functions. Label independence requires that all of them leave the probability distribution unchanged. The trouble is that virtually all of them do not leave it unchanged.

One example illustrates the general behavior. We start with two events consisting of elementary events labeled by real numbers x in the intervals [0, 0.5] and in [0.5, 1.0]. The probability density (1) assigns equal probability of 0.5 to each event. As we saw above in (Eq), label independence requires that this statement remain true when we permute the numbers that label the elementary events. We use an invertible, continuous function to carry out the permutation. Let that function map each real number x in [0,1] to a new value y in [0,1] according to:

$$y = f(x) = \sqrt{1 - x^2} \qquad (2)$$

To use the function as a permutation of labels, we take the elementary event that was originally labeled y and assign it the new real number label x. The number x is "carried along" by the function. Under this permutation, as shown in Figure 2 on the left, the two events originally labeled with real numbers in the intervals [0, 0.5] and in [0.5, 1.0] are mapped to the events originally labeled with real numbers in the intervals [0.8666, 1] and in [0, 0.8666], respectively. These last events are now assigned the new, carried along number labels in the intervals [0, 0.5] and in [0.5, 1.0] respectively.

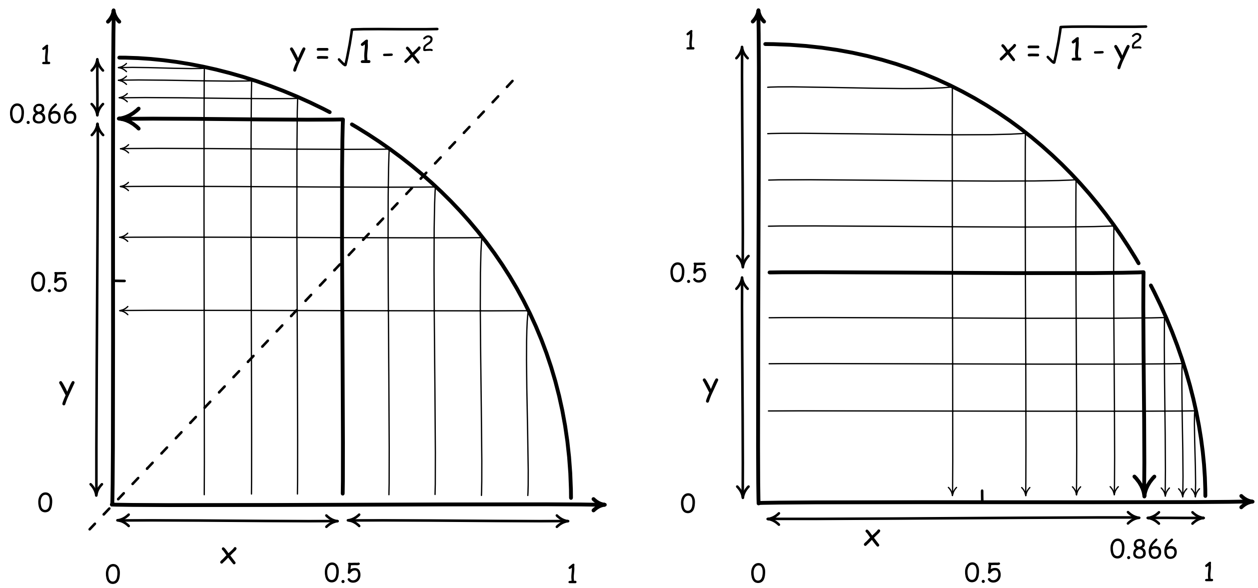

Figure 2. Uniformity of Probability Not Preserved under Permutation

These two intervals have unequal probabilities under the probability density (1): the probabilities are 1-0.8666 = 0.1333 and 0.8666 respectively. The permutation (2), however, assigns them new number labels in the intervals $[0, 0.5]$ and in $[0.5, 1.0]$ respectively. Statement (Eq) is false if we use the permuted number labels. Label independence is violated.

What would it take for label independence to be preserved? The condition needed is simple. A permutation like (2) can "carry along" the probabilities assigned to the origin set to the destination set. The key condition is that this carried along probability must match that originally assigned to the destination set. That is what failed for the permutation (2) above.

We can give this condition a general formulation as follows. The probability assigned to some small interval x to $x+dx$ is approximated by $p(x)dx$. Under the permutation, the number labels in the interval x to $x+dx$ are now reassigned to events originally labeled by numbers in the interval y to $y+dy$. These events were originally assigned a probability approximated by $\mathrm{p}(y)dy$. The condition that this original probability and the carried along probability agree is:

$\mathrm{p}(y)dy = p(x)dx$, in case dx and dy have the same sign; or

$\mathrm{p}(y)dy = -p(x)dx$, in case they differ in sign.

Taking the limit of dx and dy to zero, we have[6]

$$p(y) = p(x)\left|\frac{dx}{dy}\right| \tag{3}$$

Here $p(y)$ is the new probability density induced by the carrying along of the original probability density by the permutation, expressed in the original number labels.

A short calculation shows that the carried along probability density of (3) when computed for the permutation (2) and the source probability density (1) is

$$p(y) = \frac{y}{\sqrt{1-y^2}}$$

This induced probability density is no longer uniform over its argument, y. Thus, statement (Eq) will turn from true to false under permutation (3), violating label independence.

These last considerations lead directly to the general condition that must be satisfied by all permutations if label independence is to be preserved. It is simply

$$p(y) = p(x) \tag{4}$$

Comparing (3) and (4), we see that this equality of probability densities can only be secured if $\left|\frac{dx}{dy}\right| = 1$. This last condition is violated by almost every permutation of the number labels. For $y(x)$ a continuous, differentiable function of x, it is satisfied only by two cases $y=x$ and $y=1-x$.

The outcome is that the probability density (1) does not distribute the chances over a continuum set of elementary events indifferently, in the sense captured by the requirement of σ-field adapted label independence. For there are just two "right" ways to apply the numbering. That suggests that there is more structure hidden in the example than merely a continuum-sized set and its σ-field of subsets.

3.5 Involutions

Before proceeding, we should visit briefly with a tempting escape from the problems just developed. Might we propose that some x is the "right" labeling to use; that it has some property

---

[6] The absolute norm in $|dx/dy|$ keeps $p(y)$ positive in both cases above. Note that $|dx/dy|$ is either always positive or always negative, since the conditions of continuity and invertibility requires $x(y)$ to be everywhere increasing or everywhere decreasing.

intrinsic to the problem; and that a permutation y is somehow ill-suited, since it takes us to another labeling that lacks the property?

The particular function (2) above was chosen with just this possibility in mind. For it is an involution, which means it has the characteristic property that a double application of the function returns the original argument. That is $x = f(f(x))$. This means that there is a perfect symmetry in the relationship between x and y. Exactly the same functional form as (2) takes us back from y to x:

$$x = f(y) = \sqrt{1 - y^2}$$

Figure 2 on the right shows the inverse mapping of the interval y in [0, 0.5] to the interval x in [0.8666, 1]. The graph of an involution has the distinctive property of symmetry around the diagonal axis of the dashed line, $y=x$, shown in Figure 2. Clearly there are very many more involutions since this symmetry is all that is required.

The use of an involution responds directly to the idea that some labeling might be the "right" one. For it follows from the symmetry that, for any property that x bears with respect to y, there is a corresponding property that y bears with respect to x. Thus any decision that one of x or y is somehow favored cannot be derived from properties intrinsic to the parameters. For whatever case we make for favoring x based on the intrinsic properties of x, there is a corresponding case that can be made for y. What results is a further weakening of label independence:

> *σ-Field, Involution Adapted Label independence*
>
> All true statements pertinent to the chances of different outcomes remain true when the labels are permuted by all involutions that preserve the sets of the σ-field.

The existence of many involutions then shows that this proposal for escape fails. There is no intrinsic property of one labeling x that distinguishes it. A preference for x must be imposed by us externally by fiat. Such an external imposition breaks label independence. We may, however, find an external basis for the imposition, as we shall see in Section 4 below.

3.6 The Natural Inductive Logic on [0,1]

What if we forego the idea that inductive support must be represented probabilistically?[7] What inductive logic over the intervals of [0,1] conforms with these two weakened requirements of label independence? Even with these weakenings, it turns out that the only inductive logic admissible is akin to the infinite lottery machine logic.[8] The logic assigns the same neutral value I to any interval[9] (a, b), where $0 \leq a < b \leq 1$ in [0,1], excepting $(a, b)=(0,1)$:

$$\text{support}((a, b)) = I \qquad (5)$$

That this is the unique inductive logic conforming with the weakened label independence follows from two statements:

(i) In some real number labeling of the elementary events, all intervals (a, b) of equal size $|b\text{-}a|$ accrue the same support: support((0, 0.1)) = support((0.1, 0.2)) = support((0.2, 0.3)) =… etc.

(ii) For any $0<a<1$, $0<b<1$, there exists an involution on [0,1] that maps the interval $(0, a)$ to the interval $(b, 1)$. By label independence, they have the same support.[10]

Take any two intervals in the scope of (5): (a, b) and (c, d). By (i), they have the same support as $(0, b\text{-}a)$ and as $(1\text{-}(d\text{-}c), 1)$, respectively. Through (ii), label invariance entails that the intervals $(0, b\text{-}a)$ and $(1\text{-}(d\text{-}c), 1)$ have equal support. Hence all intervals in (5) have the same support, which we label as "I".

In this analysis, (i) is an assumption that amounts to requiring that there is at least some numbering that is naturally adapted to the equalities of support.[11] Statement (ii) is derived from

---

[7] For comparison, the transformational behavior of probability measures under involutions has been explored in greater detail in Norton (2008).

[8] As with the infinite lottery machine logic, different supports are assigned to sets of outcomes of finite size or countably infinite size.

[9] For simplicity of exposition, I consider only open intervals (a,b). The same results apply to half open and closed intervals.

[10] For the statement "Events labeled by $(0, a)$ have support X." must be true also of events labeled by $(b, 1)$ since this second set of elementary events can be relabeled through the involution by numbers in $(0, a)$.

the properties of involutions. Readers who are satisfied that the statement is correct might like to skip over the details that follow.

Statement (ii) can be demonstrated though two families of involutions that are jointly dense in the unit square, as displayed in Figure 3.

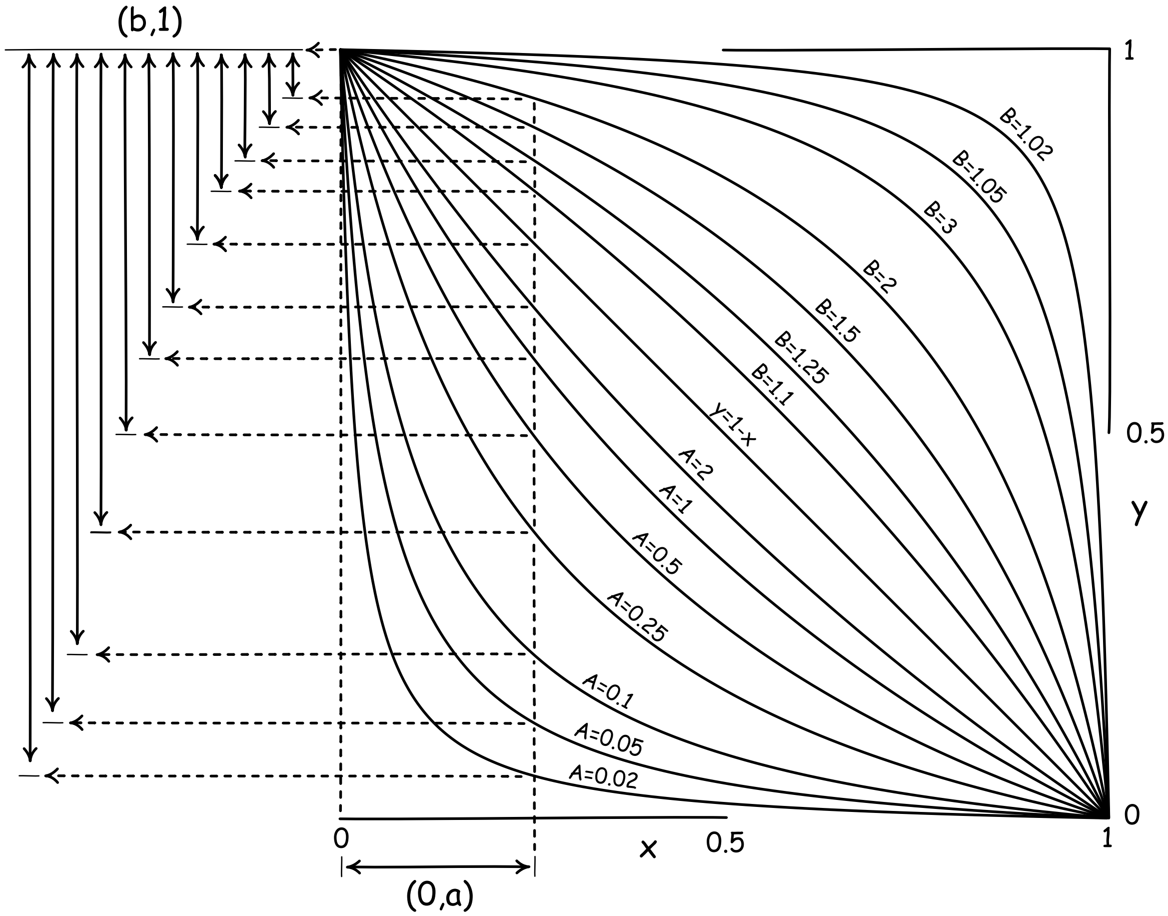

Figure 3. Two Families of Involutions on [0,1]

These involutions derive from the formulae:

$$y = \frac{A^2 + A}{x + A} - A \text{ , all } A>0, \text{ and } \quad y = \frac{B^2 - B}{x - B} + B \text{ , all } B>1.$$

That they are involutions can be seen by rearranging each to give

---

[11] Almost all of (5) can be derived with constructions like those of (ii). However no continuous involution can map all the equalities needed. None can map, say (0,0.5) to (0.1, 0.6). Something like assumption (i) is needed to complete derivation of (5).

$$(x+A)(y+A) = A^2+A \quad \text{and} \quad (x-B)(y-B) = B^2-B$$

Since x and y enter symmetrically into these rearranged formulae, it follows that, in each case, y has the same functional dependency on x as x does on y.

Consider the interval $(0, a)$ of (ii) for any $0<a<1$. It follows from the density of the involutions that there always exists one involution that maps $(0, a)$ to $(b,1)$ for any $0<b<1$. As Figure 3 shows, the A family of involutions, maps $(0, a)$ to $(b,1)$, where $0<b<1-a$. The B family of involutions maps $(0, a)$ to $(b,1)$, where $1-a<b<1$. The involution $y=1-x$, intermediate between the two families, covers the intermediate case of $b=1-a$, in which $(0, a)$ is mapped to $(1-a,1)$

4. Uniformity from Metrical Lengths, Areas and Volumes

4.1 Metrical Adaptation

If the uniform probability density (1) is to conform with label independence, we will need to weaken the requirement still further. In many important cases, a continuum-sized outcome set has further structure: a spatial metrical structure, to which the probability distribution is required to be adapted. Metrical structure assigns lengths in one-dimensional continua, areas in two-dimensional continua and volumes in three and higher dimensional continua.

When metrical structure is present, we often require adaptation of the chances to it. That means that sets of outcomes that are equal in length, area or volume have equal chances. These cases arise when, in accord with the material theory of induction, background facts warrant it. Here are some examples. A very long steel beam has defects randomly distributed through it. If it is stressed uniformly, this fact ensures that fracture is equally probable in portions of equal length. A dart is thrown at a dart board. Assuming disturbances from sufficiently many random factors, it is equally likely to strike regions of equal area. Under the physical principle of the maximization of thermodynamic entropy, a molecule of an ideal gas, free of external fields, is equally likely to be in parts of the containing vessel of equal volume.

This adaptation of chances to metrical structure can be implemented by restricting the set of the permutations in the requirement of label independence:

Metrically Adapted Label independence

All true statements pertinent to the chances of different outcomes remain true when the labels are permuted by all permutations that preserve the metrical measures of outcome sets.

A permutation preserves metrical measure just when labels identifying some metrically measurable set of outcomes are permuted to a new set of outcomes that has exactly the same metrical measure. In generic cases, such a permutation can switch any region with any other of the same metrical measure. In these cases, it follows from this weakened version of label independence that the chance of some outcome depends only on the length, area or volume associated with it. The statement "outcome A has chance such and such" must remain true when the labels identifying outcome A are relocated to any other part of the space under a metrical measure preserving permutation. The relocated outcome must have the same length, area or volume as the original, no matter how they may differ in their other properties.

These metrical measure-preserving permutations are allowed to preserve metrical measure patchwise. That is, they can divide up the space into patches and rearrange them, as long as the rearrangement preserves the measure of each patch. This last patchwise construction is a mainstay of traditional geometry. It is the standard method of proving equality of areas and volumes. Here is a rather pretty example that uses area-preserving permutations to prove Pythagoras's theorem. It is due to Rufus Isaacs (1975). The square on the left of Figure 4 shows four right angles triangles, each with sides of length a, b and hypotenuse c. They enclose a central square of area c^2, which is the "square on the hypotenuse" of Pythagoras' theorem. The area associated with this square is redistributed under a permutation shown in two steps in the central two squares. First two triangles are permuted so that their positions are moved down the figure. Then two of the triangles are moved together up the figure. The result, shown in the square on the right, is that the region forming the square of area c^2 has been relocated to a new region consisting of two squares, one of area a^2 and another of area b^2. These are the "squares on the other two sides." They are shown by this construction to be equal to the square on the hypotenuse.

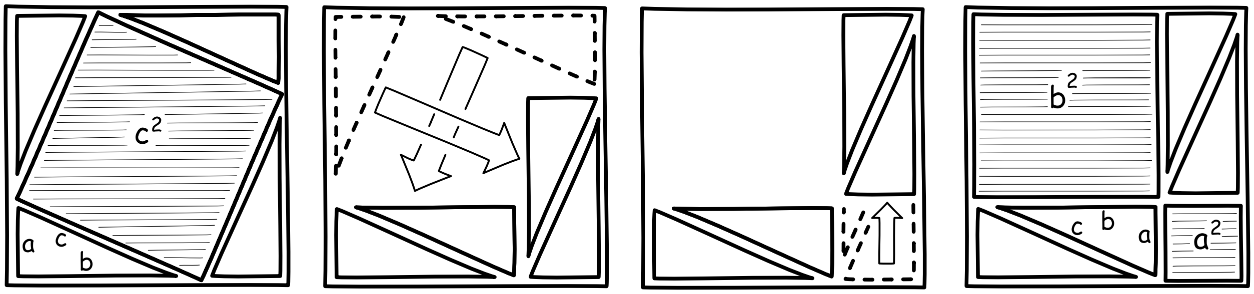

Figure 4. A Metric Preserving Permutation Proves Pythagoras' Theorem

If the chances are expressed by probabilities, metrically adapted label independence requires equal lengths, areas and volumes to be equally probable. Familiar cases work just as we would expect. These successful applications of the probability calculus arrive easily. It is because an additive metrical structure is already present in the physical assumption that the spatial continua have lengths, areas or volumes native to them. Chances acquire that additive structure upon adaptation to the metrical structure. Disjoint volumes add to give the combined volume, so the chances of outcomes in them add also to give the disjoined chances. Since the total system length, area or volume may have an arbitrary magnitude, all that remains is to normalize the adapted chances to unity to recover probability measures. If the total area of dart board is 144 square inches, then the probability of the dart striking any nominated square inch area 1/144.

4.2 The Infinite Lottery Machine Logic, Again

We can now see which will be the troublesome cases: those in which the lengths, areas or volumes of the total system are infinite. For then normalization over a uniform measure is no longer possible. If the dartboard is infinite in area, then the probability of the dart striking any nominated square inch is $0 = 1/\infty$. Since the infinite area is a countable infinity of unit areas, the chance relations among them turn out to be the same as in the infinite lottery. That is, the requirement of metrically adapted label independence leads us to the same inductive logic as applies to an infinite lottery machine.

An easy way to see this is to continue with the infinite dart board, that is, the example of areas on an infinite Euclidean plane. A process identifies a point in the plane in such a way that its chances conform with metrically adapted label invariance. We can divide this plane into infinitely many tiles of equal, finite area. For convenience, let us pick square tiles. We consider the outcome that the point selected is in one or more of these tiles. Each will have an equal chance. Infinitely many real number pairs label each square uniquely. Since there are a countable infinity of tiles, we can relabel them with a single natural number, 1, 2, 3, …. The resulting relabeling will now conform with the original, unrestricted requirement of label independence. Since the labels are natural numbers, the arguments of the previous chapter apply. The chances of outcomes in various sets of the tiles conform with the infinite lottery logic.

It now follows that all areas consisting of finitely many, n, tiles have the same chance and, as with the infinite lottery, they are assigned the chance value V_n. Since the areas of the tiles are additive, we have the further property of the additivity of these chance values. For all finite m and n,

$$V_{m+n} = V_m + V_n$$

These finite cases can be developed further in obvious ways. The more interesting cases, however, are outcomes in parts of the plane of infinite area. Crudely, under metrical adaptation, we expect trouble, since all infinite areas are equal. Using arguments carried over from the analysis of the infinite lottery machine, we will find that the chances of outcomes in all infinite-co-infinite regions have the same value, called V_∞ in the infinite lottery case.

To see this, divide the infinite plane into four quadrants, I, II, III and IV. We can then reproduce the argument concerning the sets *one*, *two*, *three* and *four* of the infinite lottery machine. We first number the tiles in the quadrant I with the numbers in the set

$$one = \{1, 5, 9, 13, ...\}$$

and then continue for quadrants II, III and IV with the numbers in the sets

$$two = \{2, 6, 10, 14, ...\}$$

$$three = \{3, 7, 11, 15, ...\}$$

$$four = \{4, 8, 12, 16, …\}$$

respectively, as shown on the left in Figure 5.

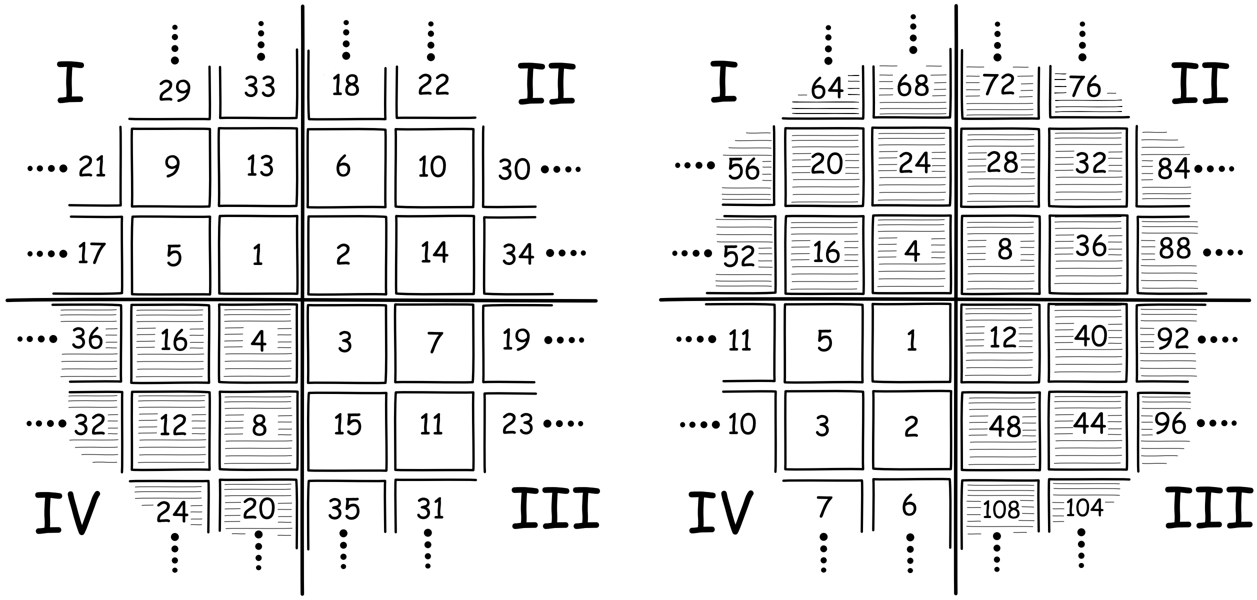

Figure 5. Rearranging Tiles over the Quadrants of an Infinite Plane

Since each quadrant contains a countable infinity of tiles, we can proceed just as we did with the infinite lottery machine. We can rearrange the tiles so that all those in quadrant I fill both quadrants I and III, while those formerly in quadrants II, III and IV fill just quadrants II and IV. Or we can rearrange the tiles so that those in quadrant IV fills quadrants I, II, and III, while those formerly in quadrants I, II and III just fill quadrant IV. This rearrangement is shown on the right in Figure 5. Since the rearrangement of tiles is merely a permutation of the labeling, it preserves chances. With further similar permutations, we can conclude:

$$\begin{aligned} \text{Ch (I)} &= \text{Ch (I or II)} = \text{Ch(I or III)} = \ldots = \text{Ch (III or IV)} \\ &= \text{Ch(I or II or III)} = \ldots = \text{Ch (II or III or IV)} = V_\infty \end{aligned} \qquad (6)$$

where "Ch(I)" designates the chance of an outcome in quadrant I.

Since this inductive logic has been elaborated more fully in the previous chapter, there is no need to duplicate the analysis here. Similar manipulations can show that this same inductive logic applies to one-dimensional continua with length and three- and higher dimensional continua with volume, if the chance processes in them conform with metrically adapted label independence. The next section provides an illustration in a science of this logic in a three dimensional space.

4.3 Continuous Creation of Matter in Steady State Cosmology

The steady state cosmology of Bondi, Gold and Hoyle enjoyed considerable attention with its initial formulation of 1948, until it eventually succumbed to several empirical problems. The most notable was an enduring difficulty in explaining naturally the cosmic background radiation observed by Penzias and Wilson in 1964. The cosmology is based on the "perfect cosmological principle." It goes beyond the more familiar cosmological principle in asserting that the universe presents the same average aspect to us not just at all positions in space, but at all times as well.

We know from measurements of the velocities of distant galaxies that the matter of the universe is everywhere expanding. That would normally entail that the average density of matter is everywhere decreasing, so it is lesser at later times. This decrease would violate the perfect cosmological principle. So steady state cosmology posits the continual creation of matter at just the right rate to maintain a constant, average matter density through time. Since ordinary matter is particulate in nature, this continual creation must be a discrete process with particles popping into existence stochastically. In Bondi and Gold's (1948) original proposal, the rate of creation was (p. 256):[12]

> The required rate of creation … can be estimated as at most one particle of proton mass per litre per 10^9 years.

By the time of writing of Bondi (1960), the requisite creation rate was updated with new astronomical measurements of the rate of expansion of the universe. Bondi now estimated it as (1960, p. 143)[13]

> …on an average the mass of a hydrogen atom is created in each litre of volume every 5 x 10^{11}years.

The difference between creation of a particle of proton mass and of hydrogen atom mass is inconsequential. A hydrogen atom consists of a proton and an electron and the proton comprises roughly 99.9% of the atom's mass.

For our purposes, the delicate question is just what stochastic rules govern the creation of these particles. The theorists ruled out the initially plausible possibility of matter creation within

---

[12] This corresponds to a mass creation rate of approximately 10^{-43} g/sec cm.^{3} (p. 265).

[13] This corresponds to a mass creation rate of approximately 10^{-46} g/sec cm^3 (p. 143).

stars. Insufficient newly created matter could escape from stars to form new galaxies. (Bondi and Gold, 1948, p. 266; Bondi, 1960, p. 149). On grounds of simplicity, the theorists proposed creation processes uniformly distributed through space. From Bondi and Gold (1948, p. 268):

> According to this view the probability of creation taking place in any particular four-dimensional element of volume (spatial volume element x element of time) is simply proportional to its (four-dimensional) volume, the factor of proportionality being a function of position. By our argument in 4.1 this factor cannot vary very much from point to point.

From Bondi (1960, p. 151):

> It seems simplest to suppose that the probability of creation in any small four-dimensional element of space-time is simply proportional to its four-dimensional volume.

On the strength of these remarks, we shall proceed in assuming the following stochastic model. In some fixed interval of cosmic time, there is an equal chance of creation of a hydrogen atom in each region of space of the same volume. Creation events are independent of each other.

Bondi and Gold assume that chance in this model can be probabilistic. They are mistaken. Since the space of steady state cosmology is Euclidean and, thus, infinite, this stochastic model conforms with metrically adapted label independence and will be governed by the infinite lottery machine inductive logic. As a result, the process of continual creation that they describe will not proceed quite according to normal expectation.

To explore the application of this logic to continual creation, imagine the Euclidean space of the cosmology divided into two infinite parts, "left" and "right" by some infinite plane. We will ask after the distribution of new particle creation events on the two sides of the plane in the course of a year. Since the average creation rate per unit volume of space is assumed non-zero, infinitely many particles will be created in each side over the year. Is this creation rate the same in both sides? That is, in the long run, are one in two creation events on the left side?

It is tempting to give the quick answer that the rate is infinitely many particles per year in both. Therefore, they are equal. This equality is something less than it seems. It does not support the further conclusion that one in two creation events are, in the long run, on the left side. Take the case in which the rate of particle creation per unit volume per year on the left side is 1,000 times greater than on the right side. Since both volumes are infinite, this case too yields a creation rate of infinitely many particles per year on both sides. Yet we do not expect one in two of them to be in this left side in the long run. It seems that a more refined means of comparing the rates of creation is needed.

In the course of a year, infinitely many particles will be created, but it will be a countable infinity. (There are a countable infinite of equal volumes of space. In each at most a finite number of particles will be created, usually zero or one.) If we track these creation events one by one, we can form the ratio of left side particle creation events to the total number. Among N particle creation events, there will be N_L creation events in the left side.

Since left and right are equally favored, our expectation is that the ratio of N_L/N will stabilize towards one half as we let N go to infinity. This expectation is not supported by the infinite lottery inductive logic. This case is isomorphic to the frequency of even numbers in repeated drawings from an infinite lottery machine. We saw in the previous chapter (§10.7) that the relative frequency of even numbers among all drawn does not stabilize to any definite value.

This result may seem to contradict the symmetry of right and left. Surely half of all creation events must happen on the left in the long run; and half must happen on the right? That expectation depends on the tacit assumption that there is an average in the long run to the fraction of creation events. We now see that there is not. The symmetry of left and right is preserved in the sense that no stable fraction arises in the long run for *both* left and right.

This result arises from tracking creation events in infinite volumes of space. If we restrict our consideration to finite volumes of space, then the normal probabilistic analysis succeeds. Over time, constant mass is preserved on average in each finite volume of space, as required by steady state cosmology.

Finally, as a minor point, this analysis involves a technical complication. It requires an enumeration of the particle creation events in the year by 1, 2, 3, 4, …, N, … so that the limit of the ratio N_L/N can be formed. Such an enumeration is possible since there are only a countable infinity of creation events. However the enumeration must be dictated by a rule that is independent of whether the event is in left region or right region. The simplest such rule is to number the creation events by the their time order. We would number the temporally first event 1, the second 2, and so on. The difficulty is that there may be no first event if the creation times have an accumulation point towards the past. That arises if, for example, the creation events happen at times (in years) 1/100, 1/101, 1/102, 1/103, … There can be multiple such accumulation points. If there are accumulation points towards the future, then the enumeration can never pass them.

I believe the following rule will solve the problem. Divide the year into 1/10ths and assign 1, 2, 3, … to the first event in each 1/10th, if there is one in each 1/10th. Next divide the year in 1/100ths and assign the next numbers to the first unnumbered events in each 1/100th, if there is one in each 1/100th. Continue for 1/1000ths, 1/10000th... If several events have *exactly* the same creation time, assign them the same number and increment both N and N_L in one step.[14]

5 Paradoxical Decompositions

5.1 What They Are

The construction of Section 4.2 above is just the first of many that yield results troublesome to additive measures. It is one of the simplest instantiations of what is known as a paradoxical decomposition. Their specification is rather general. Following Wagon (1994, Ch. 1), such decompositions arise in the context of a set E that can be partitioned into a countable collection of pairwise disjoint subsets, $A_1, A_2, A_3, \ldots, B_1, B_2, B_3, \ldots$ [15]

$$E = A_1 \cup A_2 \cup A_3 \cup \ldots \cup B_1 \cup B_2 \cup B_3 \cup \ldots$$

There must also be a group G that acts on the set E. Its elements map these subsets to other subsets of E. The original set E admits a paradoxical decomposition if elements of the group can map the A-sets of the partition to sets whose union exhaust E; and correspondingly for the B-sets. That is, there are elements of G, $g_1, g_2, g_3, \ldots$ and $h_1, h_2, h_3, \ldots$, such that we have

$$E = g_1(A_1) \cup g_2(A_2) \cup g_3(A_3) \cup \ldots$$

$$E = h_1(B_1) \cup h_2(B_2) \cup h_3(B_3) \cup \ldots$$

The standard definitions (Wagon, 1994, Def. 1.1, p. 4; p.7) do not explicitly allow for a common and important case: the mapping of the disjoint A-sets and B-sets onto E can be inverted. That is,

---

[14] This method will fail if infinitely many events have exactly the same time of creation. I presume this is not expected to happen.

[15] In the case that the A-subsets and the B-subsets each are finite in number, they do *not* need to be the same number.

a partition of the entire set E can be mapped back to either the A-sets or B-sets by elements of G.[16] When this inversion is possible, then elements of the group G can map the A-subsets onto the B-subsets; and conversely.

The construction of Section 4.2 above conforms to the conditions of paradoxical decomposition. Quadrant IV might correspond to the A-sets and the union of quadrants I, II and III might correspond to the B-sets. The group is the group of isometries of a Euclidean space. These are the maps on the space that preserve metrical distance and thus also areas. They comprise translations, rotations and reflections. Moving a tile from one part of the space to another, while preserving its area, corresponds to allowing one of the isometries to act on it. In this case, it is a translation.

The conditions for a paradoxical decomposition are realized since a rearrangement of the tiles in quadrant IV can cover the whole space; and the same is true of the tiles in the union of quadrants I, II and III. The case that concerned us, however, was the further case in which inversions are possible. Then the tiles in quadrant IV can be swapped with those in quadrants I, II and III. The import of several swaps of this type was the non-additive chances (6).

5.2 How They Extend the Analysis

There are two aspects of the argument of Section 4 for these non-additive chances that could be strengthened. First, the argument requires a decomposition into infinitely many subsets that are then rearranged to give the final result. One might worry that there is some trickery peculiar to the infinitude of the decomposition.

(i) Can the construction still proceed if the decomposition is into finitely many parts only?

Second, the total area of the Euclidean plane involved in the paradoxical decomposition is infinite.

(ii) Are paradoxical decompositions possible if we require the total area or, more generally, the total volume of the space to be finite?

The literature in paradoxical decompositions has provided affirmative answers to both questions.

---

[16] This inversion can fail if, for example, the image sets $g_1(A_1)$, $g_2(A_2)$, $g_3(A_3)$, … are not disjoint.

A paradoxical decomposition with finitely many subsets and using the group of isometries is not possible in the Euclidean plane. It is possible, however, if we move to non-Euclidean geometries. After the geometry of Euclid, the next simplest geometries are the spaces of constant positive and negative curvature. The second case of constant negative curvature is a hyperbolic geometry. It is a space of infinite area. In it Euclid's axiom of the parallels fails in this way: There is more than one straight line through a point, parallel to a given straight line elsewhere in the space. It can be visualized, piecewise, as the geometry induced on a saddle shaped surface in a higher dimensional Euclidean space.

Wagon (1994, pp. 61-68) shows that it is possible to divide up a two-dimensional hyperbolic space into three disjoint parts whose union exhausts the space and provides a paradoxical decomposition, using the isometry group. (See Wapner (2005, pp. 45-48) for a simplified and engaging development.) Call the disjoint parts A, B and C. If we choose a suitable axis of rotation, Wagon shows that it is possible to rotate A by 120^{o} so that it now coincides with B. A further rotation by 120^{o} then leaves A coincident with C. These rotations are isometries, so they preserve the areas of the parts rotated.

We might pause at this moment and imagine that a point is chosen randomly in the space such that metrically adapted label independence is respected. These rotations by 120^{o} are metrically adapted permutations that can swap the labeling among the three sets A, B and C. Thus they have equal chances. If we assign probabilities to the chosen point being in A or in B or in C, we must then have

$$P(A) = P(B) = P(C) = 1/3$$

so that $P(A) + P(B) + P(C) = 1$.

The trouble is that rotations around a different point in the space lead to different results. With a different, suitably chosen axis of rotation, a rotation of A by 180^{o} leaves it coincident with the union of B and C. Applying the same reasoning, we now arrive at probability assignments

$$P(A) = P(B \cup C) = P(B) + P(C) = 1/2$$

They are incompatible with the first set of probability assignments. Once again we find that these chances cannot be represented by probabilities.

A curious sidelight is that this case of a hyperbolic space could almost be applied directly to the example of steady state cosmology of Section 4.3. The spacetime of steady state cosmology is a de Sitter spacetime. Bondi, Gold and Hoyle introduced a cosmic time that slices the spacetime into spaces at different instants of cosmic time. They chose a slicing that yields Euclidean spaces. A de Sitter spacetime is rich in symmetries. It turns out that there are other ways of slicing it that admit different cosmic times. In another choice of cosmic time, the spaces at each cosmic instant are hyperbolic in their geometry. If we ask for matter to be created continuously by some stochastic process that is uniform in the hyperbolic space, the construction just sketched, promoted to a three dimensional space, shows that this uniformity cannot be represented probabilistically. The demonstration does not require decomposition into infinitely many parts, but just the three indicated. However the cogency of this more elegant construction is lessened by the fact that a slicing of a de Sitter spacetime into hyperbolic spaces is uncongenial to steady state cosmology. For in this slicing, the radius of curvature of the space would vary with cosmic time. (See Bondi, 1960, p. 145.) While this variant slicing is simply another way of displaying the spacetime structure of the steady state cosmology, its associated cosmic time is not one in which the perfect cosmological principle can be expressed.

The areas A, B and C of this construction are not as simple geometrically as the quadrants of Euclidean space used in Section 4.2. Each consists of infinitely may parts, with the parts touching only at points, as shown in the rather pretty diagrams in the references above. However decomposition of the hyperbolic space into these three parts is notable in one aspect: *it does not require the axiom of choice*. The significance of this statement will be clarified below.

The hyperbolic space is infinite in area and the three parts A, B and C are also infinite in area. That infinity allows them to be rotated into one another in ways that preclude a finite, additive measure for the areas. For when areas are infinite, we can write all of the following:

$$\text{Area}(A) = \text{Area}(B) = \text{Area}(C) = \infty$$

$$\text{Area}(A) = \text{Area}(B \cup C) = \text{Area}(B) + \text{Area}(C) = \infty$$

Since these equalities cannot all be satisfied if the areas of the parts are finite, one might expect that a paradoxical decomposition of a space of finite area or volume is not possible.

That expectation proves incorrect. There are paradoxical decompositions of spaces of finite volume. The celebrated example is the Banach-Tarski paradox. It has been discussed in such detail elsewhere, that it needs only the barest statement here. See Wapner (2005, Ch.5) for a very clear development; and Wagon (1994) for a mathematically more thorough treatment. The basic result is that a sphere in three-dimensional Euclidean space can be decomposed into five parts. The parts are then rearranged in space, where the rearrangement employs only volume preserving isometries. The result is two spheres, each with the same volume as the original sphere.

The air of paradox reflected in the name derives from the apparent impossibility of the process. We decompose a sphere into parts that can be recombined into two spheres whose total volume is double that of the original sphere, where all the rearrangments are isometries. The air of paradox is dispelled, however, once we find that four of the five parts in the standard decomposition are nonmeasurable in the background metric of Euclidean space. They are not simple volumes of the type normally encountered in geometry. They are scatterings of infinitely many points that defy simple geometric description. No volume can be consistently assigned to them.[17] Thus the constructions are revealed as very fancy versions of a more familiar decomposition. We can take a countable infinity of entities labeled 1, 2, 3, … and divide them into the set of odd labeled entities and the set of even labeled entities. If we now relabel the entities in each set with 1, 2, 3, … and 1, 2, 3, …, we have doubled the set of entities, or at least that is what the labeling indicates.

While Banach-Tarski like constructions have proven enormously stimulating to mathematical inquiry,[18] the most important contribution to our concerns here arises at the outset.

---

[17] A point to which we will shortly return: the axiom of choice is needed to arrive at their existence.

[18] When one first encounters these constructions, one might be quite amazed that a mortal mathematician could discover them. Or at least that was my reaction. What I found very helpful was the recognition that the more complicated constructions derive from a simple piece of group theory. The elements of the free group with two generators *a* and *b* consist of finite strings of symbols like $abba^{-1}b^{-1}$a of arbitrary, but always finite length. It is easy to see that a paradoxical decomposition is possible in this set of group elements. Any good treatment shows it. All that remains is to realize the generators in some geometrical setting, for example as rotations in space, and in a way that preserves the free group properties. Banach-Tarski like paradoxes then appear and they require three dimensions of space, since in two dimensions the two generators *a* and *b* cannot be realized. The complications of the geometry of the rotations mask the constructions' simple origins.

It is that there are nonmeasurable sets. Their existence represents some sort of obstacle to the universal applicability of additive probability measures in inductive inference. The next section turns to look at how these nonmeasurable sets arise.

6. A Nonmeasurable Set

6.1 A Vitali Set

The simplest example of a nonmeasurable set, used almost universally as in introduction to the general idea, is a Vitali set. (See Kharazishvili, 2004, ch. 1; Wagon, 1994, pp. 7-8; Wapner, 2005, pp. 132-35) The version to be developed here will be a subset of the interval of real numbers [0,1), that is all real numbers x such that $0 \le x < 1$. These real numbers will be the angular coordinates that cover a circle as shown in Figure 6.

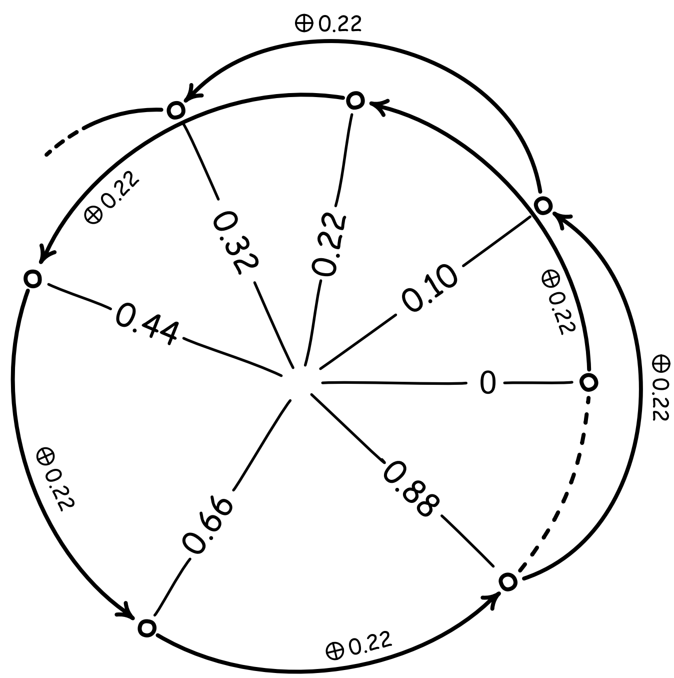

Figure 6. Equivalent numbers used in the construction of a Vitali Set

Two real numbers are defined to be equivalent under the relation "~" if they differ only by a rational number. That is $x \sim y$ just in case there is a rational number r such that $y = x \oplus r$. Addition "⊕" is modulus 1 addition. To compute it, the numbers x and r are added by ordinary arithmetic. If the resultant exceeds one, one is subtracted. If it is negative, one is added. This modular rule ensures that the sum shown always remains in the interval [0,1). Figuratively, addition by r just steps us repeatedly round the circle of Figure 6. This figure shows points equivalent under successive addition of the rational number 0.22 = 11/50, that is 0, 0.22, 0.44, 0.66, 0.88, 0.10, 0.32, …

Since the relation is an equivalence relation, it divides all the real numbers in [0,1) into disjoint equivalence classes. They are distinguished by a number that, as I shall say, "seeds" them. The rational number 0 seeds an equivalence class that contains all the rational numbers in [0,1). This shows immediately that each equivalence class has infinitely many seeds: every rational number in [0,1) seeds the same class. Irrational numbers seed other classes. The irrational $1/\sqrt{2} = 0.7071\ldots$ seeds a class that contains $(\sqrt{2}-1)/2 = 0.2071\ldots$ since

$$\frac{1}{\sqrt{2}} - \frac{1}{2} = \frac{\sqrt{2}-1}{2}$$

The simple graphic of Figure 7 displays the partition of [0,1) into the equivalence classes.

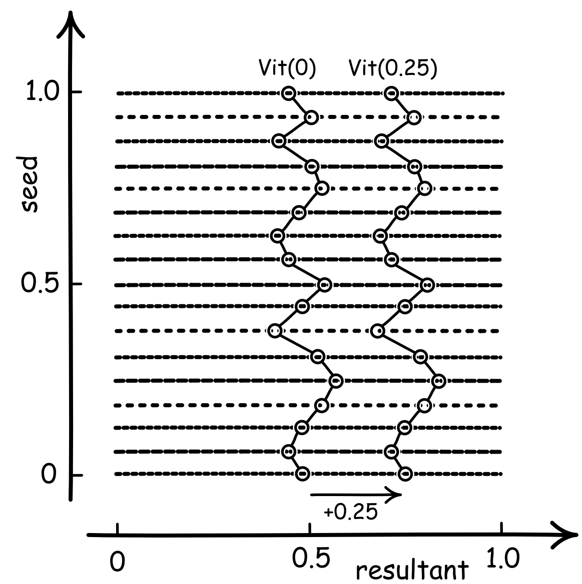

Figure 7. Choices that Form a Vitali Set

The points in the square are all the real numbers in [0,1). Each is uniquely picked out by the seed of the equivalence class to which is belongs and the rational increment added to the seed to arrive at it. The vertical axis shows the seeds used to create each equivalence class. The axis has many gaps in it since all duplicated seeds are eliminated. Its seeds include only one rational number and only one of $1/\sqrt{2}$ and $(\sqrt{2}-1)/2$. The horizontal axis shows the values in [0,1) that the various members of each equivalence class can take after all the rationals are added to the seed of the equivalence class. Each equivalence class is represented by a single horizontal line.

A Vitali set is formed by taking just one number from each equivalence class. This means that the difference between two elements in the set cannot be a rational number. Forming the set amounts to taking a vertical section in the square shown in Figure 7. It seems obvious at this point that such a section can be taken. (This is a point to which we will return shortly.) Moreover there are very many ways that this section can be taken, so very many sets can be Vitali sets. We just need to settle on one to proceed. Call it *Vit*(0).

To demonstrate that this is a nonmeasurable set, we need a measure, for a set can be nonmeasurable only with respect to some specified measure. We take the uniform distribution (1) over [0,1) as that measure. Its uniformity gives it the property of translation invariance. That is, if the probability density assigns some probability $P(A)$ to a subset A of [0,1), then it assigns the same probability to the set A_x produced by translating all numbers in A by the same amount x:[19] $P(A_x) = P(A)$. Applying a uniform translation by r to all the numbers in the Vitali set *Vit*(0), we form the translated set *Vit*(r). Figure 7 shows *Vit*(0.25).

It is easy to see that the set of translated Vitali sets, for all rational numbers r, partition the interval [0,1), as shown in Figure 8.

---

[19] That is, A_x is $\{y \oplus x: y \in A\}$.

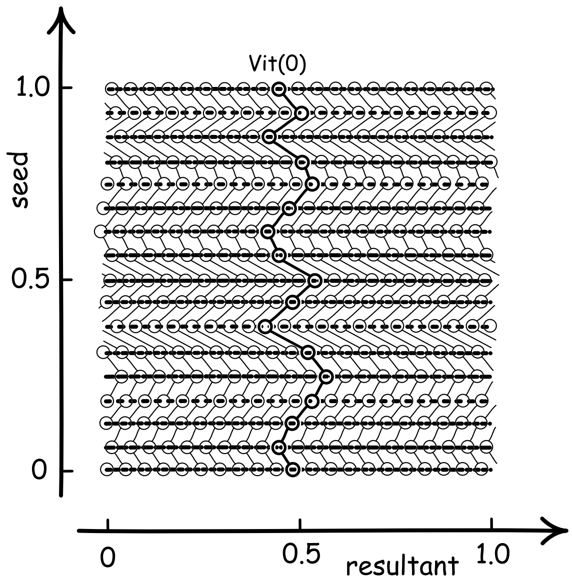

Figure 8. Vitali Sets Partition [0,1)

That is their union is [0,1) and the translated sets are pairwise disjoint. The first follows by construction, since every number in [0,1) is either in *Vit*(0) or arrived at from an element of *Vit*(0) by adding a rational r to it, which means that it is a member of *Vit*(r). Two translated Vitali sets $V(r)$ and $V(s)$ are disjoint for unequal rational numbers r and s. For otherwise they share a common element of the form $x \oplus r = y \oplus s$, where both x and y are elements of $V(0)$. However this last equation entails that x and y differ by a rational number. This cannot be true of any two distinct elements of $V(0)$, since each is drawn from a distinct equivalence class.

Assume for purposes of a reductio argument, that the Vitali set is measurable under the uniform density (1) and that is has a probability P. Since the probability density is invariant under translation, it follows that all uniformly translated Vitali sets *Vit*(r) have the same probability. The set of rational numbers is countable.[20] Therefore there are countably many translated Vitali sets. The countable sum of their probabilities must be unity. That is, the summation of a countable infinity of probabilities P must be unity. No real number P can satisfy this condition. If P is zero, the countably infinite sum is zero. If P is greater than zero, no matter

[20] For each rational can be represented by the ratio p/q of natural numbers p and q. The pair can then be mapped one-one to an infinite subset of the natural numbers by the formula 2^p3^q.

how small, the countably infinite sum is infinite. We have arrived at a contradiction. The Vitali set *Vit*(0) is not measurable under the uniform density (1).

6.2 The Infinite Lottery Machine Logic, Again

How does the existence of nonmeasurable sets like a Vitali set affect inductive inference? We can set up an inductive inference problem that uses this Vitali set by assuming that a real number has been chosen in the interval [0,1). We will assume that the choice is uniform in the sense that the chance of selection in any set, if defined, is unchanged by translations of the set. It follows that the distribution of chances in the space conforms with metrically adapted label independence, where the permutations are translations that preserve the metric associated with the probability density (1). It now follows that each of the translated Vitali sets *Vit*(*r*) must have equal chances. For any pair of Vitali sets, *Vit*(*r*) and *Vit*(*s*), a translation by *s-r* shifts the labels on the first set to the second.

The inductive problem is to determine the chances that the point selected lies in one of the Vitali sets, or in some union of them. The probability measures derived from the uniform density (1) cannot supply chances for these outcomes, for it is not defined on them. Rather, the applicable logic is the infinite lottery machine logic of the previous chapter. To see this note that the countably many Vitali sets *Vit*(*r*) can be relabeled by the natural numbers 1, 2, 3, … Each Vitali set *V*(1), *V*(2), … has the same chance and, under the new labeling, conforms with the original, unrestricted requirement of label independence. These are just the conditions to which an infinite lottery machine conforms. By repeating the arguments concerning it, we can infer that:

Chance that the point chosen is in some finite set of Vitali sets of size *N* is V_N.

Chance that the point chosen is in some infinite-co-infinite set of Vitali sets is V_∞.

Chance that the point chosen is in some infinite-co-finite set of Vitali sets, where the complement is of size *N*, is V_{-N}.

The familiar results now follow. There is the same chance that the point chosen is in the infinite set of Vitali sets that have even numbered labels, in those with odd numbered labels, in those with labels that are powers of ten: 1, 10, 100, 1000, … etc. On many repetitions there is no stabilization of frequencies such as would conform with a probability measure. We do not stabilize with roughly half the points selected in the odd numbered set and half in the even numbered set.

6.3 The Axiom of Choice

This last analysis assumes that a logic of induction should accommodate outcomes in nonmeasurable sets like the Vitali sets. However these nonmeasurable sets have a disputed status in mathematics. The difficulty derives from a key step in the analysis. The Vitali set $V(0)$ was formed by selecting just one element from each of the equivalence classes above. It was simply assumed that such a selection is possible. To see that matters are not quite so simple, one should reflect on just how we are to make the selection. Might we choose the smallest or largest element in each equivalence class? That fails since there might be no smallest or largest element. Might we choose that element that is the median value, that is, the one that comes half way through? Since the equivalence classes are infinite, "half way through" is ill-defined. Might we choose the element whose value coincides with the mean of all members in the equivalence class? That fails since there may be so such element.

We might suspect that all these failures derive from a poor imagination. There is some recipe, we might hope, even if very complicated, that lets us specify which set is our Vitali set $V(0)$. However it turns out that no one has been able to find a constructive formula that can specify the uncountable infinity of choices needed. There are formal results that suggest but do not prove that no such constructive formula is possible. Rather, the best we can now do is simply to assume that there does exist a set comprised of just one element from each equivalence class. At first glance, the existence of such sets seems so straightforward that it can hardly be doubted. But then we find reasons for doubt. Since a Vitali set results from an uncountable infinity of selections of numbers from an uncountable infinity of equivalence classes, if there are any Vitali sets, then there are very many of them. Yet when we try positively to specify just one, we can find no way to do it. If they exist, all we can say is that they are somewhere in very great numbers in the mathematical universe. We just cannot specify precisely where.

These last considerations have been codified into more precise mathematics. The standard treatment of sets is the Zermelo-Fraenkel set theory.[21] Its axioms were developed to

---

[21] For an easier introduction, see Stoll (1979, Ch. 7).

rescue set theory after Russell's paradox showed its naïve foundations fatally flawed. In the naïve set theory, we assume that a set can be formed as those things that satisfy any condition we can specify. Famously Russell used this rule to create the set of all sets that do not contain themselves as elements. The set is contradictory in that it can be member of itself if and only if it is not a member of itself.

To avoid this problem, Zermelo-Fraenkel set theory is restrained in just what sets it allows to exist. Its axioms do provide cautiously for the existence and behaviors of certain sets and include what amount to principles of set construction. The axiom schema of subsets tells us that we can always create a new set as a subset from another by placing some restrictive condition on elements in the original set. This replaces the problematic naïve rule with a benign rule, since its set delineating condition can only carve off a set from an already existing set. It does not permit formation of Russell's set. Other axioms assert the existence of the null set; of the union of two sets that are already elements of another set; of the power set of all subsets of a set; and of an infinite set constructed by specific conditions.

Constructive axioms of this type proved able to recover much of set theory. However they are not rich enough to provide for the sets like the Vitali sets of the last section. It turned out that their existence could only be secured by introducing a new, non-constructive axiom that merely asserted the existence of certain sets, but gave no recipe for their construction. That axiom is the "axiom of choice," or something equivalent to it. That axiom amounts to the assertion that, if we have a set of member sets that are pairwise disjoint, then there exists another set comprised of just one element from each of the member sets. The Vitali set $Vit(0)$ formed above is just such a set. The presumption that it exists amounts to applying the axiom of choice.

The axiom of choice has been surrounded by an air of uncertainty. A major motivation for the uncertainty was the discovery of the Banach-Tarski paradox, for the formation of the sets in the paradox require the axiom. As a result, treatments of the paradox routinely include labored discussions of the cogency of the axiom. See, for example the ominously numbered Chapter 13 of Wagon (1994). As far as I can see the question of the admissibility of the axiom and thus of nonmeasurable sets remains open simply in virtue of the lack of any well-principled means to decide for or against it.

The original basis for arguments against it was the intuitive inadmissibility of results like the Banach-Tarski paradox. To block the paradox, one had to overturn something in the foundations of set theory. The axiom of choice stood out as the easiest target because of its non-constructive character. But if one reconciles to the Banach-Tarski paradox so that it becomes the more benignly labeled Banach-Tarski theorem, then this basis for rejecting the axiom of choice is lost. Other reasons for rejecting it are hard to find. Its truth is not empirically decidable. There is no physical test we can perform to detect the existence of nonmeasurable sets of points specifically in some physical space. The axiom has been shown to be consistent with the other axioms of the Zermelo-Fraenkel set theory, so there no problem in logic in adding it to the axioms of the theory.

Correspondingly, however, there seems to be no decisive grounds for adding the axiom of choice to the other axioms of Zermelo-Fraenkel set theory. Just as there is no empirical way to falsify the axiom, there is no empirical way to demonstrate it. Rather the principal motivation for employing it seems to be pragmatic: much useful mathematics depends upon it. For example, Zorn's lemma, which is equivalent to the axiom of choice, is needed to demonstrate that every vector space has a basis.[22]

This pragmatic attitude is perhaps not so different from a simpler one. No measurement can distinguish whether a physical magnitude is an irrational real number or some nearby rational number. Any measurement has some inexactness. We can never affirm by direct measurement that the hypotenuse of a right-angled triangle with unit sides is exactly the irrational number $\sqrt{2}$ = 1.41421… as opposed to the nearby rational numbers 14/10 or 141/100 or 1414/1000 and so on. However if we forego the possibility of irrational lengths in space, we forego the right-angled triangles of Pythagoras' theorem. Instead the best we would have would be many triangles, all with sides of rational length, that come arbitrarily close to the side lengths of Pythagoras' theorem. We may congratulate ourselves on the purity of our prudence in restricting ourselves to the observationally more secure. Our reward would be mathematical complexities that would propagate pain and misery through the entirety of our physical theories.

Our question here is not simply that of the admissibility of the axiom choice. It is a slightly different one. Should an account of inductive inference be responsible for relations among propositions that pertain to nonmeasurable sets? To forego exploring these relations

---

[22] See Brunner et al. (1996) for an extended analysis of the role of the axiom of choice in the mathematics of quantum theory.

would require positive reasons for precluding nonmeasurable sets. I do not see them unless we are prepared to entertain anthropocentric perspectives on the world. That might happen if we are so committed a subjectivist that we reduce the scope of inductive inference to relations among things that we can construct. That attitude seems quite presumptuous to me. That nonmeasurable sets outstrip our constructive prescriptions seems to me quite reasonably explained by the weakness of those prescriptions. They are weak, as we have repeatedly learned. We would like a finite axiom system whose theorems would include all the truths of arithmetic. Goedel's famous theorem shows that no finite axiom system can do this. It tells us that our arithmetic axiomatic methods are weak in their reach. If finite prescriptions are essential to us, we run into trouble at the very start of mathematics. There is an uncountable infinity of real numbers in [0,1]. Yet our language admits of only countably many sentences for describing them. Most real numbers outstrip our descriptive reach. Return now to nonmeasurable sets. Are there, we might ask, nonmeasurable sets of points in our physical space? Whether there are or not is a physical fact about space and true whether our finite constructive devices allow us to give a precise description of them.

In my view, as long as the status of these sets remains open, we should consider what an inductive logic must do to accommodate them. For a general understanding of the nature of inductive inference must be expansive enough include these accommodations. To do otherwise is to prejudge the status of nonmeasurable sets and artificially restrict the scope of inductive inference. It is in that permissive spirit that the explorations of this chapter are undertaken.

7. Blackwell and Diaconis' Nonmeasurable Coin Toss Event

Most instances of nonmeasurable sets arise in the esoteric realm of abstract mathematics. When we use the sets to specify chancy events, that makes the events seem distant from the concerns of an inductive logic that may apply to real science. It would help to reduce that distance if we could find nonmeasurable events that arise in the archetypal probabilistic problem of sequential coin tosses. Blackwell and Diaconis (1996) have described such events. An account of them will be given in this section. An interesting bonus is that the apparatus needed to describe the events enables specification of another inductive logic that, while very weak, applies to events that are otherwise probabilistically nonmeasurable.

7.1 Tail Events

Blackwell and Diaconis' event arises in the case of infinitely many coin tosses. Each toss has a probability of 1/2 of a head H or a tail T; and the tosses are all probabilistically independent. Our elementary events will be infinite sequences of heads and tails. If we let variables a_1 = H or T, a_2 = H or T, a_3 = H or T, …, then such an infinite sequence is represented by the infinite tuple $\mathbf{a} = < a_1, a_2, a_3, \ldots >$. The nonmeasurable event will be one of what is called a "tail set," or, as I shall call them here, "tail event." These are events whose properties (such as the probability, if defined) depend only on the long term behavior of the infinite sequence, that is, on its tail.

Such events are familiar and important. For example, elementary events like

<H, T, H, T, H, T, H, T, …> and <H, H, T, T, H, H, T, T, …>

are distinctive in that the limiting relative frequency of heads H is 1/2. This distinctive property is shared by many other elementary events that differ in finitely many of the individual coin tosses. For example

< *H, H, H, H, H, H, H, H, H, H,* H, T, H, T, H, T, H, T, …>

differs from <H, T, H, T, H, T, H, T, …> only in its first few tosses. It will still return a limiting relative frequency of heads H of 1/2. The heavy weighting towards H in the early tosses is eventually and inexorably swamped by the later tosses.

Each of these elementary events has a probability given by the infinite product 1/2 x 1/2 x 1/2 x …. That is, each has probability zero. There are infinitely many elementary events that return this limiting relative frequency. We combine[23] them disjunctively to form the event "*half*": that the infinitely many coin tosses return a limiting relative frequency of heads H of 1/2. Since the individual tosses are probabilistically independent and each of probability 1/2, we can apply the strong law of large numbers to conclude that the event *half* will occur with probability one, *P*(*half*) = 1.

---

[23] If we think of the events as propositions, then we are "or"ing them together. If we think of them as elements of a set, we are collecting them into a set.

This last paragraph describes the distinctive property of a tail event: its probability is unaffected by whatever may happen in finitely many of the tosses that comprise it. More precisely:

> *Tail event characterization 1*: a tail event is probabilistically independent of the outcome of any finite set of tosses.

Recall that two events *A* and *B* are probabilistically independent just if $P(A \,\&\, B) = P(A).P(B)$. This defining property means that *half* is independent of the conjunction $(a_1 = \text{H})$ & $(a_2 = \text{H})$ & $(a_7 = \text{H})$ & $(a_{63} = \text{H})$, so that:

$$P(half \,\&\, (a_1 = \text{H}) \,\&\, (a_2 = \text{H}) \,\&\, (a_7 = \text{H}) \,\&\, (a_{63} = \text{H}))$$

$$= P(half) \,.\, P((a_1 = \text{H}) \,\&\, (a_2 = \text{H}) \,\&\, (a_7 = \text{H}) \,\&\, (a_{63} = \text{H}))$$

and similarly for any other finite set of tosses.

There are many other tail events. For example:

quarter: the limiting relative frequency of heads H is 1/4. *P*(*quarter*) = 0.

three-quarters: the limiting relative frequency of heads H is 3/4. *P*(*three-quarters*) = 0.

interval-no: the limiting relative frequency of heads H lies in some interval of reals that does *not* contain 1/2: *P*(*interval-no*) = 0.

interval-yes: the limiting relative frequency of heads H lies in some interval of reals that *does* contain 1/2: *P*(*interval-yes*) = 1.

even-H: an infinite number of even numbered tosses are head H. *P*(*even-H*) = 1

Tolstoy: the infinite sequence contains, infinitely often, the entirety of Tolstoy's *War and Peace*, encoded in binary using H and T, as well as every variant of the same length created by all possible typographical errors. *P*(*Tolstoy*) = 1.

It may at first seem that this list of examples is uncreative in the sense that every probability is a zero or a one. Those zeroes and ones are unavoidable however. The Kolmogorov (1950, pp. 69-70) Zero-One Law asserts that all tail events to which probability can be assigned are of probability zero or one only.

The proof of the law involves some mathematical complications. Rosenthal (2006, §3.5) gives a serviceable formulation as well as a helpful account of tail events. The basic idea behind the proof, however, is so simple and striking as to bear mention. As we saw above, the defining characteristic of a tail event we shall call "*tail*" in infinitely many coin tosses is that it is probabilistically independent of any event formed from only finitely many coin tosses, such as one we will here call "*finite*." That means

$$P(tail \,\&\, finite) = P(tail)\ .\ P(finite)$$

for all possible *finite*. The unusual circumstance is that the event *tail* is a member of the infinite set of events formed from all possible instantiations of *finite*, when closed under finite and countable unions and intersections.[24] This leads eventually to the curious result that *tail* is independent of itself! Substituting *tail* for *finite* in this last equation and noting that *tail* & *tail* = *tail*, we have

$$P(tail) = P(tail \,\&\, tail) = P(tail)\ .\ P(tail)$$

This equation admits only two solutions

$$P(tail) = 0 \quad \text{and} \quad P(tail) = 1$$

Since we will shortly be dealing with nonmeasurable events, we will need another characterization of tail events that does not explicitly invoke probability measures. That condition is simply that

> *Tail event characterization 2*: if $\mathbf{a} = < a_1, a_2, a_3, \ldots >$ is an elementary event within some tail event and $\mathbf{b} = < b_1, b_2, b_3, \ldots >$ is any elementary event that differs from it in only finitely many tosses, then **b** is also in the tail event.

This new characterization entails the original one above in case the events concerned have well defined probabilities. To see this, pick any finite set, such as a_1 and a_3. Let us say that

$$\mathbf{a}_{\text{H.H}\ldots} = < a_1 = \text{H}, a_2, a_3 = \text{H}, a_4, a_5, a_6, \ldots >$$

is an elementary event in some tail event where $a_2, a_4, a_5, a_6, \ldots$ have some values that are kept fixed in what follows here. The new condition requires that all combinations of alternative values of a_1 and a_3 appear in other elementary events in the tail event. These additional events are

$$\mathbf{a}_{\text{H.T}\ldots} = < a_1 = \text{H}, a_2, a_3 = \text{T}, a_4, a_5, a_6, \ldots >$$

$$\mathbf{a}_{\text{T.H}\ldots} = < a_1 = \text{T}, a_2, a_3 = \text{H}, a_4, a_5, a_6, \ldots >$$

$$\mathbf{a}_{\text{T.T}\ldots} = < a_1 = \text{T}, a_2, a_3 = \text{H}, a_4, a_5, a_6, \ldots >$$

The probabilistic contribution to the tail event by these four elementary events is

---

[24] That is, the σ-algebra formed from all instantiations of *finite*.

$$P(\mathbf{a}_{H.H\ldots} \vee \mathbf{a}_{H.T\ldots} \vee \mathbf{a}_{T.H\ldots} \vee \mathbf{a}_{T.T\ldots}) = P(\mathbf{a}_{H.H\ldots}) + P(\mathbf{a}_{H.T\ldots}) + P(\mathbf{a}_{T.H\ldots}) + P(\mathbf{a}_{T.T\ldots})$$

$$= P(a_1 = \mathrm{H}) \,.\, P(a_3 = \mathrm{H}).\, P(<a_2, a_4, a_5, a_6, \ldots>) +$$

$$P(a_1 = \mathrm{H}) \,.\, P(a_3 = \mathrm{T}).\, P(<a_2, a_4, a_5, a_6, \ldots>) +$$

$$P(a_1 = \mathrm{T}) \,.\, P(a_3 = \mathrm{H}).\, P(<a_2, a_4, a_5, a_6, \ldots>) +$$

$$P(a_1 = \mathrm{T}) \,.\, P(a_3 = \mathrm{T}).\, P(<a_2, a_4, a_5, a_6, \ldots>)$$

$$= P(<a_2, a_4, a_5, a_6, \ldots>).$$

This is just the probabilistic contribution to the tail arising when tosses a_1 and a_3 are excluded, which shows the probability is independent of the tosses a_1 and a_3. Repeating for all other finite combinations of tosses, we see that the probability of the tail event is independent of any of these finite combinations, which is the first characterization of tail event above.

7.2 An Intermediate Tail Event E[25]

We can start with a tail event of probability zero. By adding new elementary events to it, we can expand it to a tail event of probability one. For example, we might start with the tail event *interval-no* that is defined by the limiting relative frequency of heads lying in the interval 0.9 to 1.0. Since 0.5=1/2 is not in that interval, this tail event has zero probability. We continuously expand the interval by adding more elementary events until the interval becomes 0.4 to 1.0. At the moment when the interval expands to include the limiting relative frequency of

---

[25] Alex Pruss has pointed out another way that a nonmeasurable tail event may be formed in this coin tossing example. Each elementary event has a reversed event in which every H is replaced by T and every T by H. We form maximal equivalence classes of elementary events, such that two events in the same class differ only in finitely many of the individual coin toss outcomes. For each such equivalence class U there is reversed class U^r consisting of the reversals of the elementary events in U. The entire outcome set is partitioned by an infinity of (unordered) pairs of such classes: $\{U, U^r\}, \{V, V^r\}, \ldots$ Using the axiom of choice for collections of two-membered sets, we choose one equivalence class from each pair. Their union is the tail event N. The entire outcome set is partitioned by N and its reversal N^r. The event N satisfies conditions (a) and (b) of Section 7.3 and thus is nonmeasurable. See http://alexanderpruss.blogspot.com/2017/11/heres-simple-construction-of-non.html

heads of 0.5, its probability will flip from zero to one. Writing “rf” for the limiting relative frequency of heads and assuming that the intervals include their end points, we have

P(rf in 0.6 to 1.0) = 0
P(rf in 0.55 to 1.0) = 0
P(rf in 0.51 to 1.0) = 0
P(rf in 0.50001 to 1.0) = 0
P(rf in 0.5 to 1.0) = 1
P(rf in 0.49999 to 1.0) = 1
P(rf in 0.45 to 1.0) = 1

This last example suggests that, as we assemble sets of elementary events into events, we find no tail events intermediate between events with probability zero and those with probability one. Certainly there are none in the sequence just considered. However that last sequence included by construction only tail events with well-defined probabilities. What Blackwell and Diaconis demonstrate is that there are very many tail events, intermediate between events with zero and one probability, and that these tail events are probabilistically nonmeasureable. No probability can be assigned to each of them.

We begin assembling Blackwell and Diaconis’ event “E” as a set of elementary events, making our focus the presence of H toss outcomes. The first elementary event in E is just one that consists of all H:

$$\mathbf{a}_{\text{all-H}} = \langle \mathrm{H}, \mathrm{H}, \mathrm{H}, \mathrm{H}, \mathrm{H}, \mathrm{H}, \mathrm{H}, \mathrm{H}, \ldots \rangle$$

We now add to E all elementary events that differ from $\mathbf{a}_{\text{all-H}}$ in only finitely many tosses. They include:

<T, H, H, H, H, H, H, H,…>
<H, T, H, H, H, H, H, H,…>
<H, H, T, H, H, H, H, H,…>
…
< T, T, H, H, H, H, H, H,…>
< T, H, T, H, H, H, H, H,…>
…

Call them “infinite H, finite T” elementary events. There are as many of these elementary outcomes as there are subsets of the natural numbers. That is, there is a higher order of infinity of them. Nonetheless, the probability of the event just formed is zero. It is a tail event characterized by a limiting relative frequency of heads of one. Our starting point is essentially the same as the growing intervals of tail events above.

We will add many, many more elementary events to E but in a way that avoids the flipping of probability from zero to one. We achieve this by adding elementary events in a way that conforms with a specific set of rules. To express them, we need to define the intersection operation ∩ on elementary events. The intersection of elementary events **a** and **b** is the elementary event **a** ∩ **b** that has H in every position that has H in both **a** and **b** and T otherwise. For example:

$$\mathbf{a} = \langle \text{H, T, H, T, H, T, H, T, } \ldots \rangle$$
$$\mathbf{b} = \langle \text{H, H, T, T, H, H, T, T, } \ldots \rangle$$
$$\mathbf{a} \cap \mathbf{b} = \langle \text{H, T, T, T, H, T, T, T, } \ldots \rangle$$

The complement $\mathbf{a}^c$ of an elementary event is just that same event **a** with each occurrence of H switched to T and each occurrence of T switched to H. For example:

$$\mathbf{a} = \langle \text{H, T, H, T, H, T, H, T, } \ldots \rangle$$
$$\mathbf{a}^c = \langle \text{T, H, T, H, T, H, T, H,} \ldots \rangle$$

The event E is a set of elementary events, where we write elementary event **a** is a member of E as $\mathbf{a} \in E$.

The rules for forming E are that the following conditions are respected as the elementary events are added:

I. The "no-H" elementary event $\mathbf{a}_{\text{no-H}} = \langle \text{T, T, T, T, } \ldots \rangle$ is not in E.

$\mathbf{a}_{\text{no-H}} \notin E$.

II. ("containment") If $\mathbf{a} \in E$ and **b** arises by replacing some T in **a** by H, then **b** is also in E.

If $\mathbf{a} \in E$ and $\mathbf{a} \cap \mathbf{b} = \mathbf{a}$, then $\mathbf{b} \in E$.

III. ("intersection") The intersections of elementary events in E are also in E.

If $\mathbf{a} \in E$ and $\mathbf{b} \in E$, then $\mathbf{a} \cap \mathbf{b} \in E$.

IV. ("exhaustion") For every element **a**, either **a** or its complement $\mathbf{a}^c$ is in E.

For all **a**, either $\mathbf{a} \in E$ or $\mathbf{a}^c \in E$

V. ("free") The infinite intersection of all elementary events in *E* is the "no-H" event.

$$\bigcap_{\mathbf{a} \in E} \mathbf{a} = \mathbf{a}_{no-H}$$

Those with mathematical interests will recognize these five conditions as defining a free ultrafilter. The first three specify a filter. The fourth makes the filter an ultrafilter; and the fifth makes it a free ultrafilter.[26]

These conditions impose a definite structure on the elementary events that comprise *E*. From III. and I., we have that every intersection of elementary events in *E* must have some H toss outcomes. Thus, for all elementary events **a**, just one of **a** or its complement $\mathbf{a}^c$ can be included in *E*. Condition V. ensures that every elementary event in *E* must contain infinitely many H toss outcomes.[27]

The set of "infinite H, finite T" elementary events along with $\mathbf{a}_{\text{all-H}}$ satisfies all these conditions, excepting IV.[28] While we have not fully specified the content of *E*, we can already see at this stage that any possible set *E* must include this set. This follows from II and the fact that I requires that some H must be present in all the events of any possible set *E*.

---

[26] Blackwell and Diaconis do not implement the ultrafilter structure directly on the tuples that form the elementary events. Rather they form sets of indices of the locations of H in the tuples. For example, <H, T, H, T, H, T, …> yields the set of odd numbers {1, 3, 5, 7, …}. The ultrafilter is implemented in the set of all these subsets of the natural numbers.

[27] To see this, assume otherwise that there is an elementary event **fin**(*n*) in *E* that has finitely many H—say, *n* of them. If *n*>1, then there is an elementary event **a** such that $\mathbf{a} \cap \mathbf{fin}(n)$ and $\mathbf{a}^c \cap \mathbf{fin}(n)$ each have one or more H, but each is strictly fewer than *n*. Since just one of **a** and $\mathbf{a}^c$ is in *E*, it follows from the intersection condition III that there is another elementary event in *E* with fewer H than *n*. Iterating, it follows that, if there is an elementary event in *E* with finitely many H, then there is an elementary event in *E* with just one H. This elementary event **fin**(1) must be contained in every elementary event in *E*. Otherwise the intersection of **fin**(1) with some elementary event in *E* would be $\mathbf{a}_{\text{no-H}}$ so that $\mathbf{a}_{\text{no-H}}$ must also be in *E* by III, which then violates I. But if $\mathbf{a} \cap \mathbf{fin}(1) = \mathbf{fin}(1)$ for all **a** in *E*, then the free condition V is violated.

[28] They are equivalent to a Fréchet filter.

To satisfy exhaustion IV, we need to add further events. We have many choices over which to add. For example, we must add one of the elementary events in *half*

a = <H, T, H, T, H, T, H, T, …>

or its complement

$\mathbf{a}^c$ = < T, H, T, H, T, H, T, H,…>

But we cannot add both. Next, we must choose among

b = <H, H, T, T, H, H, T, T, …>

$\mathbf{b}^c$ = < T, T, H, H, T, T, H, H, …>

Adding the tail event *half* flipped the probability of the continuously growing set of tails events above from zero to one. We now see that this tail event cannot be a subset of *E*. For all four of **a**, $\mathbf{a}^c$, **b** and $\mathbf{b}^c$ are included in *half*. It also suggests that no tail event with a relative frequency in the vicinity of 0.5 can be in *E*. That these tail events are precluded from *E* gives the first indication that our path leads away from events with well defined probabilities. We may avoid the flipping of probability from zero to one by including only parts of these tail events in *E*.

We need to make many, many, many decisions of this type. We get a rough estimate of the number by noting that there are as many elementary events as there are members of the power set of the natural numbers, that is the set of all subsets of the natural numbers.[29] We then make about that many choices of inclusion between each elementary event and its complement. This suggests that the number of ways of forming distinct *E*s is two orders of infinity higher than the natural numbers:[30] it has the cardinality of the power set(power set (natural numbers)). There are *very* many possible events *E*!

The supposition, here, is that, if we persist in adding elementary events to *E* prudently, we will arrive at a set conforming with all the conditions. In particular, exhaustion IV will be satisfied. This is an apparently innocent supposition and essential to the formation of *E*. It is, however, a non-constructive assumption of existence. We have not specified just which elementary events can be added to satisfy exhaustion IV and, were we to try, our efforts to do so

---

[29] Each subset of the natural numbers corresponds to an elementary event. The odd numbers {1, 3, 5, …} corresponds to <H, T, H, T, H, …>.

[30] A more precise analysis shows that this is the cardinality of the set of ultrafilters on the natural numbers. See Comfort and Negrepontis (1974, p. 147).

would fail. The existence assumption turns out to be of a similar character to the axiom of choice described above. More precisely, the existence of *E* is proved by the ultrafilter theorem. Its proof commonly employs Zorn's lemma, which is equivalent to the axiom of choice. However, the ultrafilter theorem is logically weaker than the axiom of choice, as displayed in Herrlich (2006, p. 18).

Nonetheless all the vacillations that surround the earlier construction of the Vitali sets arise again here. As reported above, my view is that we should persist in exploring these systems. To do otherwise is to prejudge the admissibility of axioms like the axioms of choice and thus to restrict artificially the scope of our inductive logics.

7.3 Event E is Probabilistically Nonmeasurable

We can now prove that any event *E* conforming with the conditions I.-V. is nonmeasurable. For purposes of a reductio argument, assume that event *E* is measurable; and thus so also is its complement event E^c, the set of all elementary events not included in *E*. We will find that

(a) from a symmetry, $P(E) = P(E^c) = 0.5$; and

(b) since E is a tail event, by the Kolmogorov Zero-One Law, $P(E) = 0$ or 1.

Since (a) and (b) contradict, the reductio is completed. The set *E* is not measurable.

To see (a), note that there is a one-one correspondence between elementary events in *E* and those in E^c: each $\mathbf{a} \in E$ corresponds to $\mathbf{a}^c \in E^c$. To implement the correspondence, we just flip H to T and T to H in each elementary event **a**. It follows that each set of elementary events **a** in *E* is mapped to a corresponding set in E^c with a mirror image structure, under the flipping of H and T. Thus, if a probability is defined for the first set, then the corresponding set has the same probability. An easy way to see this is to note that we turn some set of elementary events in *E* into the corresponding set in E^c, without making any changes to the physical tosses, merely by imagining that the labels on each of the coins is switched from H to T or T to H. If follows that, if *E* is probabilistically measurable, then so is E^c; and they have the same probability. Since $P(E) + P(E^c) = 1$, we infer that $P(E) = P(E^c) = 0.5$.

To see (b), consider some elementary event $\mathbf{a} \in E$. Let **b** be any elementary event that differs from **a** in finitely many of its toss outcomes. From exhaustion IV, we have that one of **b** or $\mathbf{b}^c$ is in E. If $\mathbf{b}^c$ is in E, then so must $\mathbf{a} \cap \mathbf{b}^c$. But since $\mathbf{a}$ and $\mathbf{b}^c$ agree only on finitely many toss outcomes, it follows that $\mathbf{a} \cap \mathbf{b}^c$ has only finitely many H. We saw above that all elementary events in E have infinitely many H. Therefore $\mathbf{b}$ is in E. That is, for every elementary event in E, the event E also contains every other elementary event that differs from it in only finitely many toss outcomes. Recalling *Tail event characterization 2* above, it now follows that E is a tail event. By the Kolmogorov Zero-One Law, it has probability zero or one.

8. The Ultrafilter Logic

The analysis above shows that probabilistic reasoning over the outcomes of infinitely many coin tosses cannot proceed if our considerations include the very many nonmeasurable events of type E. The probability calculus falls silent over them.[31]

There are so very many elementary events arising with infinitely many coin tosses that we run into problems with standard methods even prior to attempting probabilistic analysis. For example, we might try to characterize the event consisting of all elementary events in which there are (in some sense) more heads than tails. One natural approach employs limits. We consider a finite sequence of coin tosses and compute the ratio of the number of heads to the number of tails. The event of interest consists of all elementary events in which that ratio is greater than one. We then take the limit as the number of coin tosses goes to infinity. The event that results will be something less than what we sought. For it is easy to contrive elementary events for which the ratio in question has no limit. All of these will be omitted from the event.

Should we despair of inductive inferences that encompass all the elementary events of the infinite coin toss? It turns out that, if we are willing to consider rather weak systems of inductive logic, we can find one that applies. It is embodied in the conditions I.-V. of the last section that characterizes an ultrafilter. A popular way of explaining the import of an ultrafilter is that it is a specification of which sets are large. In this case, a set of elementary events satisfying the conditions I.-V. contain a large number of H; all the rest do not. What makes this a natural

[31] We can get no help from upper and lower probabilities. Blackwell and Diaconis (1996) also show that the lower probabilities of both E and E^c are zero. Thus the correspondingly intervals are maximally large.

understanding is that these conditions admit only elementary events with infinitely many H; and condition II explicitly continued to populate *E* with all those elementary events with more H in them. The notion of “large” at issue here is, in intuitive terms, vague. Let us simply turn this around and assert that what we mean by “large” is membership in some set *E* that conforms with I.-V.

What results is a two-valued inductive logic that responds to the evidence that the actual outcome of infinitely tosses contains many H. The elementary events in *E* are “supported” (one value) by the evidence as having many H. The remainder are “not supported” (the other value). The axioms of the logic are the conditions I.-V. above. They play the same role as the Kolmogorov axioms of probability theory.

There are infinitely many possible sets *E* of elementary events. This infinity enables the logic to have a dynamics loosely akin to that of conditionalization in probabilistic analysis. We start out with the choice of applicable *E* left entirely open. This is as evidentially neutral a starting point as the logic admits. We can then carry out the analog of conditionalization by restricting the admissible sets *E* to those with some particular elementary event or some set of elementary events. Loosely speaking this restriction introduces the new information that something in these elementary events is close to the actual outcome. More precisely, to conditionalize on some elementary event **a** in this way is say that some infinite subsequence of **a** must be common to all elementary events in *E*. For axiom III., in conjunction with the other axioms, requires that every elementary event in *E* have an intersection with **a** that has infinitely many H in it.

As with the probability calculus, there are restrictions on the events on which we can conditionalize. In the ordinary probability calculus, we cannot conditionalize on events with zero probability. Correspondingly, if we have conditionalized on a set of elementary events containing

$$\mathbf{a} = <\mathrm{H, H, H, H, T, H, H, H, H, T, H, H, H, H, T, \ldots}>$$

we cannot then conditionalize on a set containing its complement

$$\mathbf{a}^{c} = <\mathrm{T, T, T, T, \ H, T, T, T, T, \ H, T, T, T, T, \ H, \ldots}>$$

For the axioms preclude membership of both in *E*.

The logic is weak. It is merely two-valued and, as a practical matter, no finitely specifiable set of evidence will lead to complete determination of the membership of *E*. For, as we have seen, the existence of ultrafilters must be assumed without a finite recipe for the construction of any one of them. If we exclude highly contrived fantasies, I cannot now think of a factual scenario whose background facts would require axioms I.-V. to govern our inductive inferences.

The value of the logic lies in reminding us that many logics of inductive inference are possible. If we infer probabilistically over outcomes of infinitely many coin tosses, we do arrive at many strong results. However their cost is all these inferences fall silent over the nonmeasurable events. If we are prepared to accept a weaker inductive logic, then we see that there is a logic native to the mathematical structure that does embrace all events.

9. Conclusion

The considerations of this chapter have been wide-ranging. They are, however, unified by a single question. How might an inductive logic represent the uniformity of chances over an outcome set of continuum size? It might have seemed that this is an easy case for a probabilistic logic. Is it not realized by a uniform probability density over some continuum-sized set such as the interval [0,1]? That proves not to be the case. If we define the uniformity of chances through the requirement of label independence, the inductive logic that arises is very far from a probabilistic logic.

The bulk of the chapter has tried to find how we may alter the requirement of uniformity until it matches what the probability calculus can provide. These alterations were introduced by weakening the requirement of label independence until we arrived at a version adapted to a background spatial metric. Even this weakening and the addition of background metrical structure met with limited success. For the inductive logic adapted to spaces of infinite area or volume is not probabilistic. Further, nonmeasurable sets arise in spaces of finite area and volume. They escape the reach of a probability measure if its probabilities are to match the spatial areas and volumes. The only escape from this last problem seems to be to find reasons to ignore these sets. That they are non-constructible is a tempting way to banish them from our consideration. However this escape comes at the cost of supposing that all that exists in mathematics and in the physical world described by mathematics is what we can construct by our meager, finite methods.

References

Blackwell, David and Diaconis, Persi (1996) "A Non-measurable Tail Set," Statistics, Probability and Game Theory: IMS Lecture Notes—Monograph Series, **30**, pp. 1- 5.

Brunner, Norbert; Svozil, Karl; Baaz, Matthias (1996) "The Axiom of Choice in Quantum Theory," *Mathematical Logic Quarterly*, **42**, pp. 319-40

Bondi, Hermann (1960) *Cosmology*. 2nd ed. Cambridge: Cambridge University Press.

Bondi, Hermann and Gold, Thomas (1948) "The Steady-State Theory of the Expanding Universe," *Monthly Notices of the Royal Astronomical Society*, **108**, pp. 252- 70.

Comfort, W. Wistar and Negrepontis, Stylianos (1974) *The Theory of Ultrafilters*. Berlin: Springer-Verlag.

Herrlich, Horst (2006) *Axiom of Choice*. Berlin: Springer-Verlag.

Kharazishvili, A. B. (2004) *Nonmeasurable Sets and Functions*. Amsterdam: Elsevier.

Kolmogorov, Andrei N. (1950) *Foundations of the Theory of Probability*. Trans N. Morrison. New York: Chelsea Publishing Company.

Norton, John D. (2008) "Ignorance and Indifference." *Philosophy of Science,* **75**, pp. 45-68.

Stoll, Robert R. (1979) *Set Theory and Logic*. New York: Dover.

Rosenthal, Jeffrey S. (2006) *A First Look at Rigorous Probability Theory*. 2nd ed. Singapore: World Scientific.

Isaacs, Rufus (1975) "Two Mathematical Papers without Words," *Mathematics Magazine*, **48**, p. 198.

Wagon, Stan (1994) *The Banach-Tarski Paradox*. Cambridge: Cambridge University Press.

Wapner, Leonard M. (2005) *The Pea and the Sun*. Wellesley, MA: A. K. Peters, Ltd.