Surreal Probabilities

J. DMITRI GALLOW

0 | INTRODUCTION

I'm going to flip a fair coin twice. How much more likely is it to land heads on the first flip, h_1, than it is to land heads on both flips, $h_1 h_2$? Answer: it is twice as likely.[1] After all, getting a heads on any flip is just as likely as getting a heads on any other, and the outcomes of different flips are independent of each other. In order for h_1 to be true, there must be one of these outcomes; and in order for $h_1 h_2$ to be true, there must be two. The probability of one of these outcomes is twice the probability of two. So h_1 is twice as likely as $h_1 h_2$.

Now I'm going to flip a coin infinitely many times. How much more likely is it to land heads on the first flip, h_1, than it is to land heads on every *odd* flip, $h_1 h_3 h_5 \ldots$? Answer: it is more than twice as likely, more than thrice as likely, and so on and so forth. For any integer n, it is more than n times as likely. In brief: it is *infinitely* more likely. After all, heads on any one flip is still just as likely as heads on any other, and these outcomes are still independent. In order for h_1 to be true, there must be one of these outcomes; and in order for $h_1 h_3 h_5 \ldots$ to be true, there must be infinitely many of them. The probability of one of these outcomes is infinitely greater than the probability of infinitely many of them.

How much more likely is the coin to land heads on every odd flip, $h_1 h_3 h_5 \ldots$, than it is to land heads on *every* flip, $h_1 h_2 h_3 \ldots$? Answer: again, it is more than twice as likely, more than thrice as likely, and so on and so forth. After all, nothing about the outcomes has changed, and in order for $h_1 h_2 h_3 \ldots$ to be true, there must be infinitely many *more* of these outcomes than there must be in order for $h_1 h_3 h_5 \ldots$ to be true. The probability of infinitely many more of these outcomes is infinitely lesser.

Draft of June 28, 2024; 🖂: dmitri.gallow@gmail.com, comments especially welcome

1. When I talk about how much more or less likely one proposition is than another, I mean *how many times* more or less likely it is. That is, I am asking: *by which factor* is the first's probability greater or lesser than the second's?

Or so it seems—but appearances may be misleading. My answers to the first two questions are uncontroversial conventional wisdom, but my answer to the third is heresy. Orthodox probability theory allows some propositions to be infinitely less likely than others, but only if their probability is zero. And any two propositions with zero probability are just as likely as each other. Orthodoxy permits us to say that propositions with positive probability are infinitely more likely than propositions with zero probability. So it allows us to divide the propositions up into two groups, with the propositions in one group being infinitely less likely than the propositions in the other. But it only allows us to do this once. We cannot draw similar distinctions within either of these groups. So once we've said that both $\mathrm{h}_1\mathrm{h}_2\mathrm{h}_3\ldots$ and $\mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\ldots$ are infinitely less likely than h_1, we cannot say that one of them is infinitely less likely than the other.

The reason is that orthodox probabilities are real numbers, and real numbers have the *Archimedean* property—for any two non-zero real numbers, x and y, there's always some integer n such that $nx > y$. Picturesquely, for any two non-zero real heights, you can always stack the smaller one on top of itself enough times that the stack ends up taller than the larger one. In the field of real numbers, zero is the only number which can't be stacked up enough times to overtake any other number. My answers to the second and third questions suggest that probabilities are not Archimedean. No matter how many times you stack up the probability of $\mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\ldots$, you'll never overtake the probability of h_1. And no matter how many times you stack up the probability of $\mathrm{h}_1\mathrm{h}_2\mathrm{h}_3\ldots$, you'll never overtake the probability of $\mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\ldots$.

Part of the difficulty in modeling non-Archimedean probabilities is finding an appropriate number system. One option is to use the hyperreal numbers. The hyperreals are a non-Archimedean field containing 'infinitesimals'—non-zero numbers infinitely smaller than any real number.[2] Hyperreal-valued probabilities have been touted and investigated by several philosophers, and criticized by several others.[3] Myself, I've not been moved by the critics. I think the best defenders of hyperreal probabilities have adequate responses. Hyperreal probabilities have their oddities, no doubt; but so too do real probabilities in infinite domains. No way out is pain-free; you have to pick your poison. Nonetheless, the hyperreal numbers can feel somewhat intangible. On the standard ultrapower construction, hyperreal numbers are identified with equivalence classes of ω-length sequences of real numbers, but these equivalence classes are

2. See Goldblatt, 1998.

3. See, for instance, Skyrms, 1980, Lewis, 1980, 1981, Williamson, 2007, Weintraub, 2008, Hájek, ms, Pruss, 2012, Hofweber, 2014, Easwaran, 2014, Wenmackers & Horsten, 2013, Wenmackers, 2016, Benci *et al.*, 2013, Benci *et al.*, 2016, and Parker, 2018, among others.

defined using a free ultrafilter.[4] While we can prove that such ultrafilters exist (using the axiom of choice), they can't be built up constructively. So in general, given two ω-sequences of real numbers, you won't be able to tell whether they pick out the same hyperreal number, or different ones. With the hyperreals, you can't be quite sure which numbers you're dealing with. Additionally, reasoning about the hyperreals can feel frustratingly circuitous. Because you can't get your hands on them directly, you have to prove that there's a way of translating claims about the reals, which you can get your hands on, to claims about the hyperreals. And you can prove that this translation (called the 'star map') is truth-preserving. So, to reason about the hyperreals, you first reason about the reals, and then translate your conclusion into one about the hyperreals. But you have to exercise care; the translation only says something about *some* of the hyperreals. What the original claim says about all of the reals, the new claim says only about the hyperreals in the image of the star map. For this reason, the translation of 'the powerset of X' needn't be the powerset of the translation of 'X'. And since 'the partial sums of the geometric series $1 + 1/2 + 1/4 + 1/8 + \cdots + 1/2^n + \ldots$ converge to 2' is true in the reals, its translation will come out true in the hyperreals, even though these partial sums do *not* converge in the hyperreals.

So the hyperreals live behind a veil and speak to us only in code. This leads Hájek (2003) to complain that hyperreal probabilities are *ineffable*. He writes: "when philosophers gesture at infinitesimal probability assignments, I want them to give me a *specific example* of one. But this they cannot do; the best they can do is gesture."[5] Of course, for all their obscurity, we can nonetheless dimly perceive the hyperreal numbers. We can learn about them indirectly. We can decipher their messages. And the hyperreals can be fruitfully used to teach us things about the reals. But if you're primarily interested in non-Archimedean probabilities, it's natural to want a less Delphic system of numbers.

Here, I'll use John Conway's surreal numbers to give a more direct system for representing non-Archimedean probabilities. The surreal numbers allow us to meet Hájek's challenge and produce specific examples of infinitesimal probability assignments. With minimal assumptions, we'll be able to derive precise numerical values for the probabilities of propositions which are infinitely unlikely. For instance, we will

4. It doesn't matter what this is, but if you don't know and are curious, the relevant ultrafilter is just a collection of subsets of natural numbers which is closed under supersets and finite intersections and which is maximal in the sense that it either contains S or S's complement, for every set of natural numbers S; the ultrafilter is free if its intersection is empty. See Goldbring, 2022. In the ultrapower construction, you identify two ω-sequences of real numbers just in case the set of places in which they match is included in the ultrafilter.

5. Hájek, 2003, p. 292. See also the criticisms in Easwaran, 2014.

show that the probability of a fair coin landing heads on every odd flip, $h_1 h_3 h_5 \ldots$, is $2^{-1/4} \cdot \mathfrak{c}^{-1/2}$ (where $\mathfrak{c}$ is the cardinality of the continuum of real numbers). The candor of surreal probabilities is striking; it stands in stark contrast to the vague and hushed pronouncements of hyperreal probabilities.

The name 'surreal numbers' is due to Donald Knuth.[6] Conway just calls them 'numbers'—or 'all the numbers great and small'. 'Surreal numbers' is a much better name; but it can leave you with the impression that the surreals are of interest primarily for their peculiarity. In my view, this is a mistake. The surreal numbers are a serious number system, no more peculiar than Cantor's ordinals (which are contained within the surreals). In contrast to the hyperreals, the surreal numbers can be explicitly constructed. You can watch as they are born. There is no trouble picking out a particular infinitesimal surreal number, nor any trouble distinguishing two different infinitesimal surreals. There are simple and natural rules for adding and multiplying them together. And they give us all the probabilistic distinctions we could ask for—all the probabilistic distinctions the hyperreals give and more.

In section 1, I will give an introduction to the surreal numbers. This introduction will give us all we need to do some basic probability theory with surreal numbers. In section 2, I'll discuss which axioms we should impose on surreal-valued probabilities. And in section 3, I'll give a model of surreal probabilities for the case of flipping a fair coin infinitely often. We'll see that, from minimal assumptions, we will be able to define surreal probabilities for a large collection of infinitely unlikely propositions about how the coin lands. Along the way, I will respond to an argument from Williamson, 2007, which attempts to show that the probability of an infinite sequence of heads must be zero. But this paper won't offer a sustained defense of surreal probabilities over their real counterparts. Instead, I simply want to get surreal probabilities out on the table. To date, there has been almost no investigation of surreal probabilities in either mathematics or philosophy. In philosophy, Chen & Rubio (2020) assume surreal probabilities and use them to define surreal-valued *utilities* via an analogue of the von Neumann & Morgenstern representation theorem. They show how surreal-valued utilities can be used to address Pascal's wager and other puzzling decisions involving infinity.[7] I know of no work on surreal probabilities within mathematics. With such limited development, it is difficult to evaluate the prospects of using surreal numbers to model the kinds of non-Archimedean judgments I opened with, just as it is difficult to evaluate Chen & Rubio's proposal to use surreal numbers to put a value on Pascal's wager. My main goal here is to at least begin to remedy this situation. I will provide

6. Knuth, 1974.

7. See also §4.1 of Hájek, 2003.

a 'user's guide' to surreal probabilities, show how to calculate their values in the case of flipping a fair coin infinitely many times, and attempt to showcase some of their benefits.

1 | CONWAY'S PARADISE

In this section, I will provide an introduction to surreal numbers. I won't assume that you have any prior familiarity with them, though I will assume that you have some familiarity with ordinal numbers, as well as basic arithmetic on the real numbers. I won't be carefully proving properties of the surreal numbers—you can find the proofs elsewhere.[8] My goal is to just give you a basic understanding: how to build these numbers up, how to break them down, how to locate them on their number line, and how to add and multiply them together.

1.0 | Construction

Each surreal number has a birthday. Zero is born on the zeroth day. On the first day, one and negative one are born. The next day, -2, $-1/2$, $1/2$, and 2 are born. Each day, new numbers are born to fill in the gaps between the older numbers. When we awake on day n, we find the collection of numbers born before day n. We then cut these numbers in twain, dividing them into two sets, L and R (for *l*eft and *r*ight), where every number in L is less than every number in R. Then, a new number is born which is strictly greater than every number in L and strictly less than every number in R. A number like this is born for every possible way of cutting the already born numbers in twain (even the trivial cuts, which place all of the already existing numbers to one side or the other).

We can represent these cuts with '$\{L \mid R\}$', where L is the set of numbers to the left of the cut and R is the set of numbers to its right. At the start of day zero, there are no numbers, so the only cut is the trivial one, $\{\ \mid\ \}$.[9] This cut divides nothing in twain, leaving nothing on the left and nothing on the right. Since all of the none of the numbers on the left are less than all of the none of the numbers on the right, this counts as a cut. So the number corresponding to this cut is born. Call it '0'. On the morning of the first day, there is only one pre-existing number, so there are only two possible cuts: $\{\ \mid 0\}$ and $\{0 \mid\ \}$. Since all of the none of the numbers on the left are less than all

8. The interested reader can consult Conway, 1976 first, and then (in no particular order) Gonshor, 1986, Alling, 1986, Ehrlich, 1994, Ehrlich, 2001, Erlich, 2011, Rubinstein-Salzedo & Swaminathan, 2014, and Costin & Ehrlich, 2022.

9. Strictly speaking, this should be written '$\{\{\} \mid \{\}\}$'. But it's standard to omit the inner set brackets in the interests of readability.

of the numbers on the right, $\{\ |\ 0\}$ is a cut, and corresponding to this cut, a number is born. Because of the cut that birthed it, this number must be less than zero. Call it '-1'. Similarly, since all of the numbers on the left are less than all of the none of the numbers on the right, $\{0\ |\ \}$ is a cut, and corresponding to this cut, a number is born. Because of the cut that birthed it, this number must be greater than 0. Call it '1'. On day two, we have four possible cuts: $\{\ \ |-1,0,1\}$, $\{-1\,|\,0,1\}$, $\{-1,0\,|\,1\}$, and $\{-1,0,1\ |\ \ \}$. For each of these cuts, a number is born. These numbers sit where their cuts were drawn, so let's call them '-2', '$-1/2$', '$1/2$', and '2', respectively. (For now, we can take these names to be stipulative; but once we have the rules for addition and multiplication, we'll find that $0 = \{\,|\,\}$ is the additive identity, and $1 = \{0\,|\,\}$ the multiplicative identity. We'll also have $1+1 = \{0\,|\,\}+\{0\,|\,\} = \{0,1\,|\,\} = 2$, $1/2+1/2 = \{0\,|\,1\}+\{0\,|\,1\} = \{0\,|\,\} = 1$, and so on. So the stipulative names are well-chosen.)

Any dyadic fraction—any rational number whose denominator is a power of two—is born in a finite number of days. These are all of the numbers which can be represented with a terminating binary decimal expansion. But no other numbers are born in a finite number of days. The square root of two is not yet born, nor is one third, nor π. But once infinitely many days have passed, we can awake on the ωth day (the next day after any finite number of days have passed) and once again start splitting the numbers in twain. On the ωth day, $1/3$ will be born. For $1/3$ is the number greater than $1/4$, $1/4 + 1/16$, $1/4 + 1/16 + 1/64$, $\ldots$, and less than $1/2$, $1/2 - 1/8$, $1/2 - 1/8 - 1/32$, $\ldots$. Similarly, every other real number will be born on day ω, since every real number has a decimal expansion in binary. For any real number, r, cut all of the dyadic fractions into those less than r and those greater than r. This will be a cut, and so on day ω, a new number will be born sitting in that cut's place. r is the only real number sitting at this cut's position, so call the number born of the cut 'r'.

But there are other numbers born on day ω. Consider the trivial cut with all of the pre-existing numbers on the left and nothing on the right. This is a number greater than every dyadic fraction. There is no such real number, but there is a surreal number like this. Call it 'ω', after Cantor's first infinite ordinal. Or consider the cut which places zero and every pre-existing negative number on the left and every pre-existing positive number on the right. This is a number strictly greater than zero but strictly less than every dyadic fraction. That is, it is greater than zero, but less than $1/2$, $1/4$, $1/8$, $\ldots$. There is no such real number, but there is a surreal number like this. Call it 'ϵ'. (We'll see later on that $\epsilon = \omega^{-1}$.) On day $\omega + 1$, we continue cutting, and we get even more numbers—$\omega + 1$, $\omega - 1$, $\epsilon/2$, and 2ϵ, for instance. In general, we keep cutting for as many days as there are ordinal numbers. The surreals are the proper class of numbers so-formed.

Another, equivalent, way of thinking about surreal numbers is like this: they are

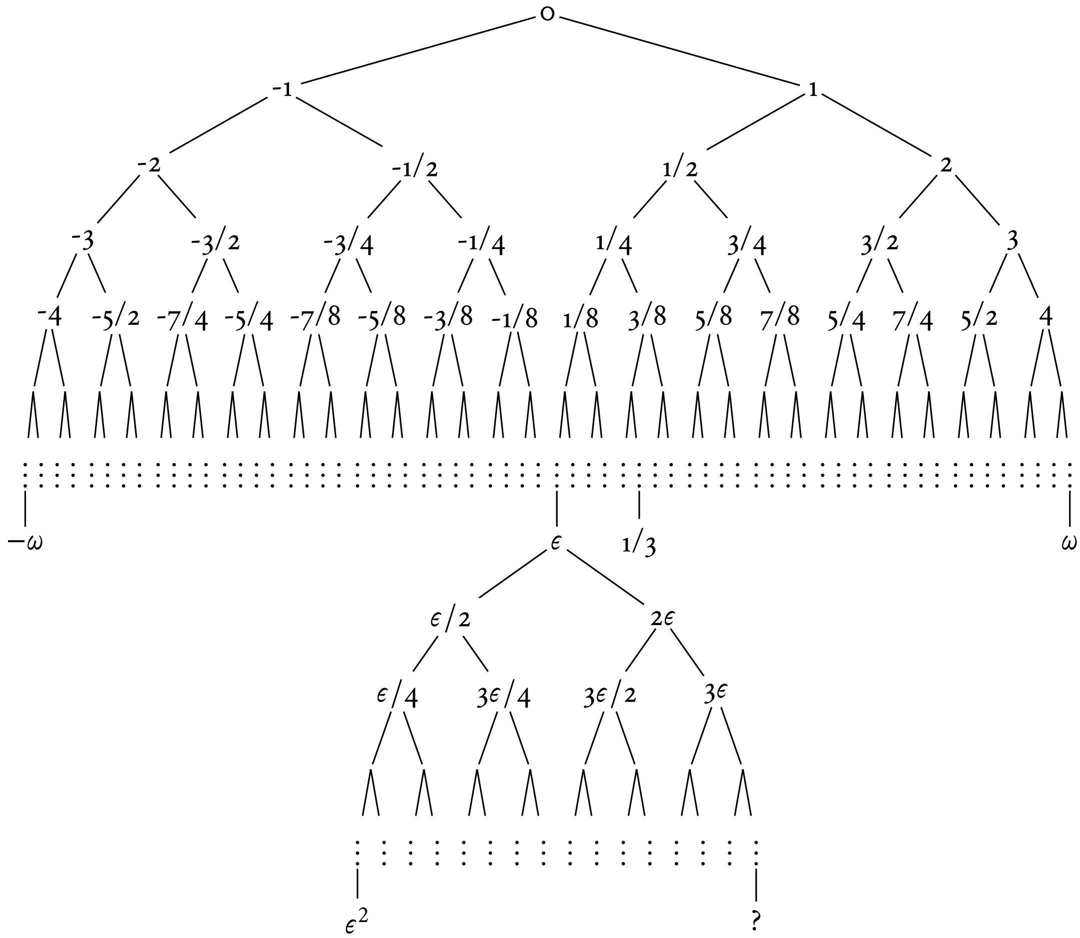

FIGURE 1: The surreal numbers are the nodes in the complete binary tree.

the nodes of the complete binary tree. This tree starts with a single root node (the number zero) which branches into two children (one and negative one); each of its children branch into two children, and so on and so forth. After ω many branches, the tree continues growing. There is a new node which sprouts from the bottom of each infinite path down the tree. The tree thereupon continues branching, with new branches growing at each ordinal. There are exactly two children of every node in this tree, and exactly one child of every downward path through the tree which is the length of a limit ordinal. (See figure 1.)

Associated with any of these numbers is a sequence of decisions of whether to travel left (in the negative direction) or right (in the positive direction) at each node. The length of this sequence is the birthday of the associated number. For instance, 3/2 is associated with two steps in the positive direction followed by one step in the negative direction, $(+, +, -)$, and 2ϵ is associated with the $\omega + 1$-length sequence of one step in the positive direction, infinitely many steps in the negative direction, and one step in the positive direction, $(+, -, -, -, \ldots, +)$.[10] The order of the numbers can be read off of these associated sequences. Take any two numbers, x and y, and go to the first position in their associated sequences which don't match; if x's value in that position is greater than y's value in that position, then x is greater than y. (By stipulation, $-$ is less than undefined, which is less than $+$.) This will give us a strict total order.[11]

You can find Cantor's ordinal numbers sitting at the right-most branch at every level of the binary tree. Whereas one of Cantor's ordinals is built from the set of all its predecessors, $\beta = \{\alpha\}_{\alpha<\beta}$, one of Conway's ordinals is built from the *left* set of all its predecessors, $\beta = \{\alpha \mid \}_{\alpha<\beta}$. We use the ordinals to count birthdays. A number's birthday is the ordinal it was born with—it is the ordinal at the same level as that number in the binary tree. If one number's birthday precedes another's, then say that the first number is *simpler* than the second. So zero is the simplest number of all, 3/4 is simpler than 5/8, and every dyadic fraction is simpler than 1/3. Then, we can introduce a convention for naming surreal numbers: if L and R are any sets of numbers such that every number in L is less than every number in R, then let $\{L \mid R\}$ name the simplest number greater than all the numbers in L and less than all the numbers in R. For instance, $\{-2 \mid 3\}$ is another name for zero, and $\{-4 \mid -7/8\}$ is another name for -1. If x is any surreal number, we can always find some collection of 'representative' numbers less than x, x^L, and some collection of 'representative' numbers greater than x, x^R.[12] In the binary tree, if there's a most recent predecessor of x from which you have to step right (in the positive direction) in order to reach x, x^L, and some most recent predecessor from which you have to step left (in the negative direction) in order to reach x, x^R, then $x = \{x^L \mid x^R\}$. If there's no most recent predecessor like this, then we'll have to consider some representative sequence which approaches x from the left or the right. For instance, on your path to 2ϵ, you most recently stepped right at ϵ and there's no most recent left-hand step, but rather an infinite sequence of steps with no last left-hand step. So we can write $2\epsilon = \{\epsilon \mid \ldots, 1/8, 1/4, 1/2\}$. That is: 2ϵ is the simplest number greater than ϵ but less than 1/2, 1/4, 1/8, and so on.

Consider the surreal number

$$\{\epsilon, 2\epsilon, 3\epsilon, \ldots \mid \ldots, 1/8, 1/4, 1/2\}$$

10. Gonshor, 1986 defines surreal numbers as these sequences of pluses and minuses.

11. That is: (a) no surreal is greater than itself; (b) if x is greater than y, y is not greater than x; (c) if x is greater than y and y is greater than z, then x is greater than z; and (d) for any two distinct surreal numbers, x and y, either x is greater than y or y is greater than x.

12. Conway (1976) calls these numbers 'typical', rather than 'representative'.

This is the simplest number greater than any integer multiple of ϵ and less than any fraction of 1. You reach it in the binary tree by taking one step right, in the positive direction, to one, then taking infinitely many steps left, in the negative direction, arriving at ϵ, and then taking infinitely many steps right, in the positive direction. It is associated with the 2ω-length sequence of one plus followed by infinitely many minuses, followed by infinitely many pluses, $(+, -, -, -, \ldots, +, +, +, \ldots)$. A riddle for the reader (to be answered below): what name should we give this number?

1.1 | Arithmetic

Arithmetic operations on the surreal numbers are defined recursively in terms of their representative bounds. In general, if $x = \{x^L \mid x^R\}$ and $y = \{y^L \mid y^R\}$ (where these representative bounds could be empty, they could be single numbers, or they could be sequences of numbers),

$$\begin{aligned} x + y &= \{x + y^L\,,\ y + x^L \mid x + y^R\,,\ y + x^R\} \\ -x &= \{-x^R \mid -x^L\} \\ xy &= \{x^L y + x y^L - x^L y^L\ ,\ x^R y + x y^R - x^R y^R \\ &\qquad \mid x^L y + x y^R - x^L y^R\ ,\ x^R y + x y^L - x^R y^L\} \end{aligned}$$

In the case where the left- or right-hand representative bounds are sequences, you apply these definitions to *all* of the numbers in the sequences. If there are no representative bounds on the left or right, then you apply them to all of the none of the numbers in those bounds. The fact that we can start with empty bounds allows the recursive definitions to get going without a base case.[13]

To add x to y, you put on the left all of the sums of x and the left bounds of y and all of the sums of y and the left bounds of x; and you put on the right all of the sums of x and the right bounds of y and all of the sums of y and the right bounds of x. (By the symmetry of this definition, we'll automatically have that $x + y = y + x$.) For instance, to find the sum $0 + 0 = \{\,|\,\} + \{\,|\,\}$, we put on the left all of the none of the sums of zero and the left bounds of zero, and we put on the right all of the none of the sums of zero and the right bounds of zero, yielding $\{\,|\,\}$. So $0 + 0 = 0$. To add $0 = \{\,|\,\}$ and $1 = \{0\,|\,\}$, we put on the left the sum of zero and the left bound of one (along with all of the none of the sums of one and the left bounds of zero) and we put on the right all of the none of the sums of one and the right bounds of zero, together with all of the

13. These definitions will still be correct, even when you don't use the representative bounds; but if you use bounds less simple than the numbers being added, there's no guarantee that applying the definitions recursively will eventually terminate.

none of the sums of zero and the right bounds of one, yielding $\{0 + 0 \mid\} = \{0 \mid\} = 1$. So $0 + 1$ is 1. From here, we can show

$$\begin{aligned} 1 + 1 &= \{0 \mid\} + \{0 \mid\} = \{1 + 0, 1 + 0 \mid\} = \{1 \mid\} = 2 \\ 0 + 1/2 &= \{\mid\} + \{0 \mid 1\} = \{0 + 0 \mid 0 + 1\} = \{0 \mid 1\} = 1/2 \\ 1 + 1/2 &= \{0 \mid\} + \{0 \mid 1\} = \{1 + 0, 0 + 1/2 \mid 1 + 1\} = \{1/2, 1 \mid 2\} = 3/2 \\ 1/2 + 1/2 &= \{0 \mid 1\} + \{0 \mid 1\} = \{0 + 1/2 \mid 1/2 + 1\} = \{1/2 \mid 3/2\} = 1 \end{aligned}$$

If we let n range over all the positive integers, then we can write ω as '$\{n \mid\}$'. Then,

$$\begin{aligned} \omega - 1 = \omega + (-1) &= \{n \mid\} + \{\mid 0\} \\ &= \{n - 1 \mid \omega + 0\} \\ &= \{n \mid \omega\} \end{aligned}$$

(Since n ranges over all positive integers, the bound $n - 1$, or $0, 1, 2, \ldots$, is no different from the bound n, or $1, 2, 3, \ldots$.) So $\omega - 1$ is the simplest number greater than every positive integer but smaller than ω. In the binary tree, you get to $\omega - 1$ by traveling right in the positive direction for infinitely many steps, arriving at ω, and then taking a single step left in the negative direction. It is associated with the $\omega + 1$-length sequence of infinitely many pluses followed by a single minus: $(+, +, +, \ldots, -)$. Likewise, for any fixed positive integer m, $\omega - m = \{n \mid \omega - m + 1\}$ is the simplest number greater than every positive integer but less than $\omega - (m - 1)$.

The definition of multiplication is more complicated. It is justified by the observation that, since x is greater than x^L and y is greater than y^L, we should have $(x - x^L)(y - y^L) > 0$. So if we want the numbers to behave, we should have $xy > x^L y + x y^L - x^L y^L$. So we should have $x^L y + x y^L - x^L y^L$ as a left bound of xy. Similarly, we should have $(x - x^L)(y^R - y) > 0$, giving us the right bound $x^L y + x y^R - x^L y^R$. The other bounds follow from the requirements that $(x^R - x)(y - y^L) > 0$ and $(x^R - x)(y^R - y) > 0$. Since 0 has no representative bounds, multiplying by zero will just return zero, so $x \cdot 0 = 0$ for any x. And, for any x, we'll have

$$x \cdot 1 = \{x^L \mid x^R\} \cdot \{0 \mid\} = \{x^L \cdot 1 + x \cdot 0 - x^L \cdot 0 \mid x^R \cdot 1 + x \cdot 0 - x^R \cdot 0\} = \{x^L \mid x^R\}$$

From here, you can verify that $1/2 \cdot 2 = 1$ and $-1 \cdot 1 = -1$, for instance. For a more interesting example, take $\omega/2$:

$$\begin{aligned} \omega \cdot 1/2 &= \{n \mid\} \cdot \{0 \mid 1\} \\ &= \{n/2 + \omega \cdot 0 - n \cdot 0 \mid n/2 + \omega \cdot 1 - n\} \end{aligned}$$

$$= \{n/2 \mid \omega - n/2\}$$
$$= \{n \mid \omega - n\}$$

(Since n ranges over all positive integers, $n/2$ is the same bound as n, and $\omega - n/2$ is the same bound as $\omega - n$.) So $\omega/2$ is the simplest number greater than every positive integer n and less than $\omega - n$, for any positive integer n. In the binary tree, you get to $\omega/2$ by traveling infinitely many steps to the right, arriving at ω, and then traveling infinitely many steps to the left. It is associated with the 2ω-length sequence of infinitely many pluses followed by infinitely many minuses: $(+, +, +, \dots, -, -, -, \dots)$.

I won't bother proving it here, but it can be proven that surreal addition and multiplication are associative and commutative, that multiplication distributes over addition, and that $-x$ is the additive inverse of x. There is a way of defining the multiplicative inverse of x, $1/x$, and it can be shown that $x \cdot (1/x) = 1$. So the surreal numbers satisfy the field axioms—and they would be a field, but for the fact that they are too large, forming a proper class rather than a set. Conway calls them a Field, on the grounds that proper classes, like proper names, should be capitalized.

1.2 | Archimedean parts

The real numbers have the *Archimedean* property that, for any two positive real numbers, r and s, there's some integer n such that $nr > s$. For any two positive numbers, one can be stacked up enough times that it overtakes the other. In one sense, the surreals are not Archimedean. For any integer n, $n\epsilon < 1$. So we could say that 1 is infinitely larger than ϵ. Similarly, there's no integer n such that $n\epsilon^2 > \epsilon$. So we could say that the square of ϵ is infinitely smaller than ϵ itself.

As an aside, I think that there's a deeper sense in which the surreals *are* Archimedean. For the surreals have their own distinct kind of integer. Say that n is a *surreal* integer iff n is the simplest surreal number between $n - 1$ and $n + 1$, $n = \{n - 1 \mid n + 1\}$.[14] It can be shown that the surreal numbers have the following Archimedean-*like* property: for any two positive surreal numbers, x and y, there's some *surreal* integer n such that $nx > y$.[15] The surreals appear non-Archimedean if you're looking at them from the perspective of the reals, and only allowing yourself to stack a number up a *real* integer number of times. But within the Field of surreal numbers, you should allow yourself to stack a number up any *surreal* number of times. Then, for any two positive surreal numbers, you'll be able to stack the one up enough times that it overtakes the other.

14. Conway calls them 'omnific' integers.

15. See Conway, 1976, ch. 5.

But put that to the side. Let's confine our attention to the *real* integers, and continue using 'integer' to mean *real* integer. Say that two positive surreal numbers, x and y, are on the same Archimedean level iff there's some integer n so that $nx > y$ and some integer m so that $my > x$. Being on the same Archimedean level is an equivalence relation, so it partitions the positive surreal numbers into equivalence classes, which we can call *Archimedean levels*—or just *levels*. These levels inherit the ordering of their members. One is higher than another iff the numbers in one are greater than the numbers in the other. (This will give us a strict total order.) If X and Y are two Archimedean levels of positive numbers, with X higher than Y, then the numbers in X are infinitely larger than the numbers in Y, and the numbers in Y are infinitely smaller than the numbers in X.

Each level has a unique simplest member. It is the firstborn of its level. Conway calls it the 'leader'. For instance, all of the positive reals are on the same level. (Though they are not the *only* numbers on this level; they are joined by numbers like $5 - 6\epsilon$ and $12 + \epsilon^5/\sqrt{\omega^\omega}$. In general, these levels are convex: if x and y are both on the same level, and $x < z < y$, then z is on the level, too.) Amongst all these numbers, 1 is the firstborn. So 1 is the leader of its level. Similarly, ω and ϵ are the leaders of their levels.

Let ω^0 be the simplest leader of all. That is, $\omega^0 = 1$. Let ω^1 be the simplest positive number larger than any multiple of ω^0, and let ω^{-1} be the simplest positive number smaller than any fraction of ω^0. That is: $\omega^1 = \omega$ and $\omega^{-1} = \epsilon$. We can likewise let $\omega^{1/2}$ be the simplest positive number greater than any multiple of ω^0 and less than any fraction of ω^1. That is,

$$\omega^{1/2} = \left\{0, n\omega^0 \,\middle|\, \frac{\omega^1}{2^n}\right\} = \left\{0, n \,\middle|\, \frac{\omega}{2^n}\right\}$$

where n ranges over all integers. In the binary tree, you reach $\omega^{1/2}$ by taking infinitely many steps in the positive direction, arriving at ω, then infinitely many steps in the negative direction, arriving at $\omega/2$. You then take infinitely many more steps in the negative direction, arriving at $\omega/4$, followed by infinitely many more in the negative direction, arriving at $\omega/8$, and so on and so forth, until you've taken an infinite number of left steps an infinite number of times. $\omega^{1/2}$ is associated with an ω^2-length sequence of ω-many pluses followed by ω-many ω sequences of minuses.

In general, if x is any surreal number, there will be a leader ω^x, defined to be the simplest positive number greater than any multiple of ω^{x^L} and less than any fraction of ω^{x^R}.

$$\omega^x \overset{\text{def}}{=} \left\{0, n\omega^{x^L} \,\middle|\, \frac{\omega^{x^R}}{2^n}\right\}$$

These leaders can be thought of as powers of ω, since they obey the usual laws of

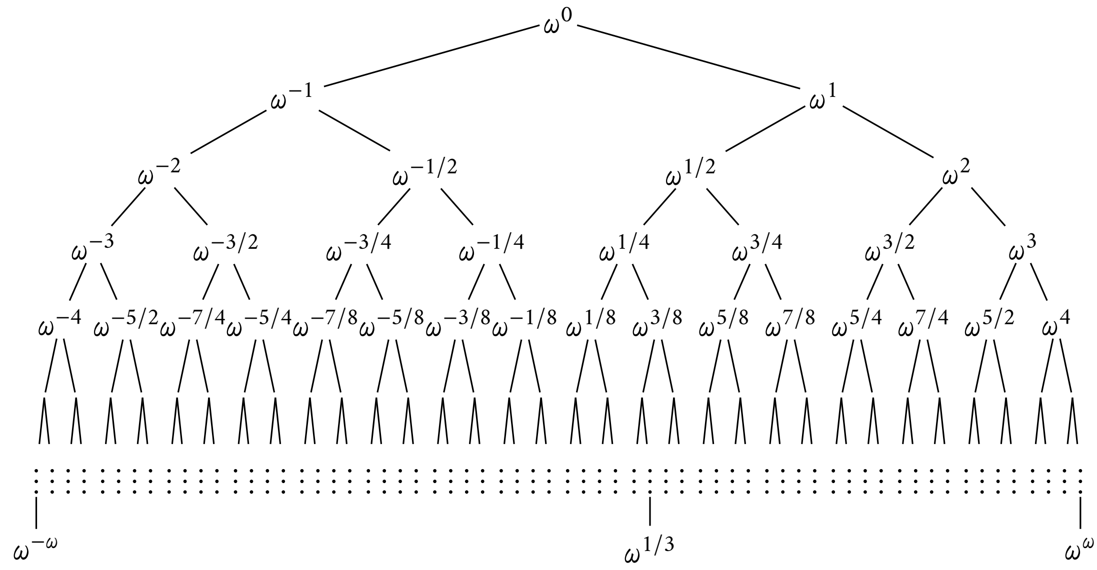

Figure 2: The leaders of the Archimedean levels of positive surreal numbers have the same order and simplicity structure as the surreal numbers themselves.

exponents. In particular, $\omega^0 = 1$, $\omega^x \cdot \omega^y = \omega^{x+y}$, and $\omega^{-x} = 1/\omega^x$.

The leaders are ordered in the same way as the surreal numbers themselves—for they are built up in the very same way. (See figure 2.) Since the order of the Archimedean levels is the same as the order of their leaders, this means that the levels are also ordered in the same way as the surreal numbers. The levels of the leaders to the left of $\omega^0 = 1$ contain all of the positive numbers infinitely smaller than 1, and the levels of the leaders to the right of ω^0 contain all of the positive numbers infinitely larger than 1. (Since our topic is surreal *probabilities*, we will be concerned exclusively with the level of $\omega^0 = 1$ and the levels to its left.)

Any surreal number can be decomposed into a (perhaps infinite) sum of real multiples of these leaders. Take any surreal number x. x is on some Archimedean level or other. Let ω^{y_0} be the leader of this level. Then, there will be some real number r_0 such that x is r_0 times ω^{y_0} plus something infinitely smaller than ω^{y_0}—call it 'x_1'. That is, $x = r_0 \cdot \omega^{y_0} + x_1$. It could be that $x_1 = 0$, in which case we are done. But if not, x_1 will be some number infinitely smaller than ω^{y_0}. So x_1 will be on its own level. Let 'ω^{y_1}' be the leader of that level. (Note that, since x_1 is infinitely smaller than x, y_1 must be smaller than y_0.) Then, there will be some real number r_1 such that x_1 is r_1 times ω^{y_1} plus something infinitely smaller than ω^{y_1}—call it 'x_2'. That is, $x_1 = r_1 \cdot \omega^{y_1} + x_2$, and therefore $x = r_0 \cdot \omega^{y_0} + r_1 \cdot \omega^{y_1} + x_2$. We can carry on in this way for as long as we have to until we have no infinitesimal bits left over. In this way, we'll have written out the surreal number x as a (perhaps infinite) sum of real multiples of leaders.

In general, then, for any surreal number x, we can find a decreasing sequence of surreal numbers $(y_\alpha)_{\alpha<\beta}$ (where α and β are ordinals), and a corresponding sequence of real numbers $(r_\alpha)_{\alpha<\beta}$, such that

$$x = \sum_{\alpha<\beta} r_\alpha \cdot \omega^{y_\alpha}$$

This is called the *Conway normal form* of the surreal number x. It gives us a decomposition of a surreal number into real multiples of its component Archimedean parts. All surreal numbers have a Conway normal form, and any two distinct surreal numbers have distinct Conway normal forms.

Doing surreal arithmetic from the foundational definitions can be confusing and tedious. (But try proving that $1/2 + 1/2 = 1$ in the field of real numbers starting from the foundational definitions!) Breaking surreal numbers down into real multiples of their Archimedean parts affords us a much simpler and more natural way of adding and multiplying them. We can just add and multiply the real multiples of the Archimedean parts using the comfortable and familiar rules for the reals, together with the natural rules for the powers of ω.[16] For instance, if we add $5+6\omega^{-1}+40\omega^{-2}$ to $4\omega^{-1}+5\omega^{-2}+\omega^{-3}$, we get $5 + 10\omega^{-1} + 45\omega^{-2} + \omega^{-3}$. So addition works just as you would expect. Likewise for multiplication. To get the product of $3\omega^{-1/2}+4\omega^{-3/4}$ and $\omega^{-1/2}$, just do what comes naturally:

$$(3\omega^{-1/2} + 4\omega^{-3/4}) \cdot \omega^{-1/2} = 3\omega^{-1} + 4\omega^{-5/4}$$

More carefully, if $r \cdot \omega^y$ appears in x's normal form ($r \neq 0$), say that x has an ω^yth part. For instance, $5 + 6\epsilon$ has an ω^0th part and an ω^{-1}th part. And $12\omega + 7$ has an ω^1th part and an ω^0th part. Now, suppose you want to take the sum or product of x and y. Let $(z_\alpha)_{\alpha<\beta}$ be a decreasing sequence of surreal numbers such that, for every $\alpha < \beta$, either x or y (or both) has an ω^{z_α}th part, and every Archimedean part of either x or y appears somewhere in $(\omega^{z_\alpha})_{\alpha<\beta}$. Then, we can write x as $\sum_{\alpha<\beta} r_\alpha \cdot \omega^{z_\alpha}$ and y as $\sum_{\alpha<\beta} s_\alpha \cdot \omega^{z_\alpha}$, where $(r_\alpha)_{\alpha<\beta}$ and $(s_\alpha)_{\alpha<\beta}$ are real numbers—though some of these real numbers could be zero, if one of x or y has an ω^{z_α}th part and the other does not. Then, we will have that

$$x + y = \sum_{\alpha<\beta} (r_\alpha + s_\alpha) \cdot \omega^{z_\alpha}$$

$$\text{and} \quad xy = \sum_{\alpha<\beta} \sum_{\gamma<\beta} \omega^{z_\alpha + z_\gamma} \cdot r_\alpha \cdot s_\gamma$$

16. See theorem 16 of Ehrlich, 2001.

This is why we gave the name 'ϵ^2' to the surreal number $\{0 \mid \epsilon/2^n\}$. This is the number you reach in the binary tree by taking one step to the right from zero, then infinitely many steps to the left from one, arriving at ϵ, and then taking infinitely many more steps to the left. (See figure 1.) It associated with the 2ω-length sequence $(+, -, -, -, \dots, -, -, -, \dots)$. And it is ω^{-2}, since it is the simplest positive number smaller than any fraction of $\omega^{-1} = \epsilon$. So it is the leader of the Archimedean level indexed by $\{\ \mid -1\} = -2$. Using the rule for multiplication, $\omega^{-2} = \omega^{-1} \cdot \omega^{-1} = \epsilon \cdot \epsilon$.

Go back to the riddle from §1.0: what name should we give the surreal number

$$\{\epsilon, 2\epsilon, 3\epsilon, \dots \mid \dots, 1/8, 1/4, 1/2\}?$$

You may have been tempted to answer '$\omega\epsilon$'. After all, ω is what we get when we take infinitely many steps to the right from one, so shouldn't $\omega\epsilon$ be what we get when we take infinitely many steps to the right from ϵ? The rules for multiplication teach us that this tempting thought is incorrect. For $\omega\epsilon = \omega^1 \cdot \omega^{-1} = \omega^0 = 1$. Instead, since this is the simplest number greater than any multiple of ω^{-1} but less than less than any fraction of ω^0, it is the leader indexed by $\{-1 \mid 0\} = -1/2$. So the answer to the riddle is '$\omega^{-1/2}$', or '$1/\sqrt{\omega}$'—or, since $\epsilon = \omega^{-1}$, we could call it '$\sqrt{\epsilon}$'.

Conway normal form allows us to evaluate some (but not all) infinite series. Take, for instance,

$$\epsilon + {}^{\epsilon}/_{2} + {}^{\epsilon}/_{4} + \cdots + {}^{\epsilon}/_{2^n} + \dots$$

The nth partial sum of this series is $\sum_{i=0}^{n} {}^{\epsilon}/_{2^n}$. Now, we cannot talk sensibly about these partial sums converging in the surreals, if what we mean by 'convergence' is the epsilon-delta definition from real analysis. Take any increasing ω sequence of surreals, $x_1, x_2, \dots$, and take any supposed limit of this sequence, x. If it's going to be a limit, then x must be greater than each x_i. But then, there will be a surreal number $y = \{x_1, x_2, \dots \mid x\}$ which is strictly greater than every number in the sequence and strictly less than x. So let $\varepsilon = x - y$, and we cannot have $x - x_n < \varepsilon$ for any x_n.[17] (Here, ε is just some small number; it is not the simplest infinitesimal ϵ.) But Conway provides a different way of evaluating this infinite series. Each of the partial sums $\sum_{i=0}^{n} {}^{\epsilon}/_{2^n} = \left(\sum_{i=0}^{n} {}^{1}/_{2^n}\right) \cdot \omega^{-1}$ is a real multiple of ω^{-1}. And, in the field of real numbers, these multiples converge to 2. So we can evaluate the infinite series as $2 \cdot \omega^{-1} = 2\epsilon$.

17. There is a corresponding notion of limit for the surreals, but it only applies to ordinal-length sequences, not ω length sequences. See Rubinstein-Salzedo & Swaminathan, 2014.

More generally,, suppose we have an infinite sum

$$\sum_{i=1}^{\infty}\left(\sum_{\alpha<\beta} r_\alpha^i \cdot \omega^{y_\alpha}\right)$$

where the numbers being summed all have the same Archimedean parts ω^{y_α} and variable real multiples $r_\alpha^i \neq 0$. And, suppose that, for each α, the set of finite sums $\sum_{i=1}^n r_\alpha^i$ converge to r_α in the reals. Then, we can assign this infinite series the value of $\sum_{\alpha<\beta} r_\alpha \cdot \omega^{y_\alpha}$. However, if for any α, the finite sums of the real multiples $\sum_{i=1}^n r_\alpha^i$ fail to converge, then the infinite sum of surreal numbers will be undefined.

That's all we need to know about the surreal numbers to get going. In the next section, I'll discuss which axioms we should use for surreal-valued probabilities. And in §3, I'll define a particular surreal-valued probability for the case of flipping a fair coin infinitely many times.

2 | ASSUMPTIONS ABOUT SURREAL PROBABILITIES

I'll assume that surreal-valued probabilities have the same basic properties as real-valued probabilities. In particular, I'll assume that they are...

... *non-negative*: no proposition's probability is less than zero;

... *normalized*: a proposition's probability is one if it is guaranteed to be true and zero if it is guaranteed to be false;

... *monotonic*: if A entails B, then the probability of B is no less than the probability of A; and

... *additive*: the sum of A's and B's probability is equal to the sum of the probabilities of their union and intersection: $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cup B) + \mathbb{P}(AB)$.

I'll follow the usual conventions for real-valued probabilities by assuming that two propositions are independent iff the probability of their conjunction is the product of their probabilities, $\mathbb{P}(AB) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. And I'll understand a *conditional* probability to tell us how much less likely a conjunction is than one of its conjuncts, $\mathbb{P}(A \mid B) = \mathbb{P}(AB) \div \mathbb{P}(B)$. So two propositions are independent iff their conditional probabilities are equal to their unconditional probabilities, $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

In addition, I'll assume that surreal-valued probabilities are *regular*, meaning that a proposition's probability is zero only if it is guaranteed to be false. This assumption is pre-theoretically plausible, but it has to be rejected if probabilities are real-valued, since any way of assigning positive real numbers to each of an uncountable number of possibilities will either violate additivity or normalization. But surreal probabilities won't encounter this difficulty. In my view, this is a series advantage of of using surreal numbers to represent probabilities. I'll have more to say about this in §4.

Are surreal probabilities *countably* additive? That is, must the probability given to a countable union of disjoint propositions equal the infinite sum of the probabilities of those propositions? Of course, infinite sums of probabilities needn't be well-defined in the surreals, but we can still ask whether probabilities are countably additive *when* the relevant infinite sums are well-defined. That is, we can ask whether, if $A_i A_j = \varnothing$ for each $i \neq j$,

$$\mathbb{P}\left(\bigcup_{i=1} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad \textit{if defined}$$

Because infinite series whose partial sums are bounded can fail to be defined, the surreal version of countable additivity is less constraining than its real analogue. If probabilities are real-valued, then a countably infinite fair lottery is inconsistent with countable additivity. But there is no conflict if probabilities are surreal-valued. We are free to say that each ticket has an infinitesimal probability of ϵ of winning. For the infinite sum $\sum_{n=1}^{\infty} \epsilon$ is not well-defined. (The partial sums of the real multiples of ϵ, $\sum_{i=1}^{n} 1$, diverge to infinity.) So we are free to say that the probability of *some* ticket winning is 1.

Nonetheless, we must reject the surreal analogue of countable additivity. Consider the proposition that the coin lands heads on every flip, $A = \mathrm{h}_1\mathrm{h}_2\mathrm{h}_3 \dots$. If probabilities are surreal-valued, then this proposition should presumably receive a positive infinitesimal probability. Call that probability, whatever it is, 'x'. Then, consider the proposition that the coin *doesn't* land heads on every flip, $\neg A = \neg(\mathrm{h}_1\mathrm{h}_2\mathrm{h}_3 \dots)$. $\neg A$ says that there is at least one tails landing. Given our other assumptions about probabilities, $\neg A$ must have a probability equal to $1 - x$.[18] But $\neg A$ is equivalent to the infinite disjoint union $\mathrm{t}_1 \cup \mathrm{h}_1\mathrm{t}_2 \cup \mathrm{h}_1\mathrm{h}_2\mathrm{t}_3 \cup \dots$. And the infinite sum of the probabilities of the propositions in this disjoint union is well defined, and is equal to 1, not $1 - x$.

$$\mathbb{P}(\mathrm{t}_1) + \mathbb{P}(\mathrm{h}_1\mathrm{t}_2) + \mathbb{P}(\mathrm{h}_1\mathrm{h}_2\mathrm{t}_1) + \dots = 1/2 + 1/4 + 1/8 + \cdots + 1/2^n + \cdots = 1$$

So if we want probabilities to be surreal-valued, and we continue to understand infinite sums of surreal numbers as Conway taught us to understand them, then countable additivity must be rejected. If we want a surreal version of countable additivity, then we need a different way of evaluating infinite sums of surreal numbers.

---

18. From normalization, $\mathbb{P}(A \cup \neg A) = 1$ and $\mathbb{P}(A \cap \neg A) = 0$. So, from additivity, $\mathbb{P}(A) + \mathbb{P}(\neg A) = 1$. So $\mathbb{P}(\neg A) = 1 - \mathbb{P}(A)$.

As an aside, let me note that this example doesn't threaten a related principle known as 'countable *sub*additivity'. Let $A_1, A_2, A_3, \dots$ be an infinite sequence of propositions (which could be, but needn't be, disjoint). And suppose that the infinite sum $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ is well-defined. Then, countable subadditivity says that the probability that at least one of the propositions A_i is true cannot exceed the value of this infinite sum—if that infinite sum is well-defined. That is,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad \textit{if defined}$$

Countable subadditivity won't require us to say that the probability of the coin having at least one tails landing, $\neg A$, is 1. Instead, it will only require its probability to be *no higher* than 1. In the real numbers, countable subadditivity implies countable additivity so long as probabilities are at least finitely additive. For suppose we have an infinite collection of pairwise disjoint propositions, $A_1, A_2, \dots$. Then, finite additivity tells us that the probability of $\bigcup_{i=1}^{\infty} A_i$ can be *no less* than the infinite sum $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$. For the infinite sum is defined to be the least upper bound of the finite partial sums. By finite additivity, each finite partial sum is the probability of some finite union of the A_is. And by monotonicity, the probability of $\bigcup_{i=1}^{\infty} A_i$ cannot be less than the probability of any finite union of the A_is. So, in the reals, if we have an infinite collection of pairwise disjoint propositions, then finite additivity and countable subadditivity will together imply that

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Where the inequality on the left follows from finite additivity and the inequality on the right follows from countable subadditivity. And this implies that, whenever the A_is are pairwise disjoint, $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$, which is countable additivity. However, given the way that infinite sums are defined in the surreals, finite additivity will *not* imply that $\sum_{i=1}^{\infty} \mathbb{P}(A_i) \leqslant \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right)$ when the A_is are pairwise disjoint. For, in the surreal numbers, the infinite sum is *not* the least upper bound on the finite partial sums. There will always be surreal numbers in between all of the finite partial sums and the infinite sum. For instance, the surreal number $1 - \epsilon$ is strictly greater than each of the finite sums $\sum_{i=1}^{n} 1/2^i$, but strictly less than the infinite sum $\sum_{i=1}^{\infty} 1/2^i = 1$. If we understand infinite sums as Conway taught us to understand them, then surreal probabilities cannot be countably additive; but they may yet be countable *sub*additive.

Consider an infinite fair lottery. Because the lottery is fair, every ticket must be given the same probability of winning. There are infinitely many tickets, so the probability of any particular ticket winning must be infinitesimal. *Which* infinitesimal? As far as I can see, the probability axioms on their own don't require us to make any particular choice. But one answer naturally suggests itself: the reciprocal of the cardinality of the number of tickets. After all, given any finite number of tickets, N, a fair lottery will give a probability of N^{-1} to each ticket winning. It's natural to extend this pattern when the number of tickets is infinite, and say that, if the cardinality of the set of tickets is α, then a fair lottery will give a probability of α^{-1} to each ticket winning. This is the simplest choice, and so it's the one I will make. But I want to emphasize I am *not* endorsing any general principle about the relationship between probability and cardinality. In particular, we cannot say that a uniform measure over *any* partition of cardinality α should give probability α^{-1} to each cell in the partition.

To appreciate why, take the case of a *countably* infinite fair lottery. In this case, we can model the relevant possibilities with the positive integers, $\mathcal{W} = \{1, 2, 3, \dots\}$, with the interpretation that n is the world in which ticket n wins. The cardinality of $\mathcal{W}$ is ω. So I have proposed setting the probability of each $W \in \mathcal{W}$ to $\omega^{-1} = \epsilon$. But notice that $\mathcal{W}$ can be partitioned into the following propositions (notice how the numbers count along the anti-diagonals):[19]

$$\begin{array}{rlllllll}
A_0 = & \{ 1, & 3, & 6, & 10, & 15, & 21, & \dots \} \\
A_1 = & \{ 2 & 5 & 9 & 14 & 20 & 27 & \dots \} \\
A_2 = & \{ 4 & 8 & 13 & 19 & 26 & 34 & \dots \} \\
A_3 = & \{ 7 & 12 & 18 & 25 & 33 & 42 & \dots \} \\
A_4 = & \{ 11 & 17 & 24 & 32 & 41 & 51 & \dots \} \\
\vdots & & & & & & &
\end{array}$$

We already know how to calculate the probability that a ticket from A_m wins if there is a finite number of tickets n—call this probability '$\mathbb{P}(A_{m,n})$'. As n goes to infinity, the ratio $\mathbb{P}_n(A_{m,n}) \div \mathbb{P}_n(A_{m+1,n})$ converges to 1. So it's natural to assume that every proposition in this countable partition should be just as likely as every other. By additivity, $\mathbb{P}(A_m)$ must be greater than ϵ, greater than 2ϵ, greater than 3ϵ, and so on and so forth. So its probability must be on an Archimedean level greater than $\epsilon = \omega^{-1}$. So we cannot say that the probability of A_m is ω^{-1}, even though our probabilities are uniform over the partition $\{A_0, A_1, A_2, \dots\}$, and this partition has a cardinality of ω. So there is no general principle for determining infinitesimal probabilities on the basis of cardinality alone.

Nonetheless, I will assume that, if α is the cardinality of $\mathcal{W}$, then a uniform distribution over $\mathcal{W}$ assigns a probability of α^{-1} to each $W \in \mathcal{W}$. I'm inclined to regard

19. A_0 contains all the triangular numbers $\{T(n) \mid n \in \mathbb{Z}^+\}$, where $T(n) = \sum_{i=1}^{n} i$. And for $m \geqslant 1$, A_m is $\{T(n+m) - m \mid n \in \mathbb{Z}^+\}$.

this as no more than a matter of convention. After all, for any positive x and y, the interval of surreal numbers $[\omega^{-x}, 1]$ is isomorphic to $[\omega^{-y}, 1]$. These intervals are distinguished by the simplicity of their endpoints; they cannot be distinguished by the Field operations $+$ and $\times$, for there is an isomorphism of the surreal numbers which maps one of these intervals into the other.[20] So any collection of surreal probabilities living in one of these intervals has a matching collection of surreal probabilities living in the other. In the field of real numbers, we represent probabilities by their relative positions in the interval $[0, 1]$. Singling this interval out involves two conventional choices: which number to use as the lowest possible probability and which to use as the greatest possible probability. There's no reason we *have* to use zero and one, though it is convenient to do so. In the Field of surreal numbers, there's an additional conventional choice to be made: which infinitesimal value to use for a uniform distribution. There's no reason we *have* to use the reciprocal of the cardinality of $\mathscr{W}$, but it is convenient and simple to do so.

3 | SURREAL PROBABILITIES FOR COIN FLIPPING

3.0 | Assumptions about probabilities for coin flipping.

Suppose we will flip a fair coin infinitely many times. We can model this with a set of infinite sequences of 'h's and 't's. Let '$\mathscr{W}$' be the set of infinite sequences like this, and call each sequence in $\mathscr{W}$ a 'world'. Propositions about how the coin lands are sets of worlds. For instance, the proposition that the coin lands heads on the first flip, h_1, is the set of all sequences from $\mathscr{W}$ which begin with an h, and the proposition that the coin lands heads on every third flip, $\mathrm{h}_1\mathrm{h}_4\mathrm{h}_7\mathrm{h}_{11}\ldots$ is the set of all sequences from $\mathscr{W}$ that have an 'h' in every third position.

Since the coin is fair, every sequence in $\mathscr{W}$ should be just as likely as every other. That is, we should have $\mathbb{P}(W) = \mathbb{P}(W^*)$ for any $W, W^* \in \mathscr{W}$. There are as many sequences of 'h's and 't's in $\mathscr{W}$ as their are real numbers,[21] so the cardinality of $\mathscr{W}$ is $2^\omega = \mathfrak{c}$. Here, '2^ω' is not surreal exponentiation, but rather the cardinality of all functions from ω to $2 = \{0, 1\}$; that is, it is the cardinality of the powerset of ω, which is the cardinality of the set of real numbers. If we assume the continuum hypothesis, $\mathfrak{c} = \omega_1$, the first uncountable cardinal; but if the continuum hypothesis is false, $\mathfrak{c}$ could be greater than ω_1. Since it won't matter for our purposes whether the continuum

20. See Hamkins (2024).

21. To appreciate this, note that each sequence can be associated with a real number between zero and one. Just swap out each 'h' with a '1' and swap out each 't' with a '0', and interpret the resulting sequence as the decimal expansion of a real number in $[0, 1]$ in binary. Since there are uncountably many real numbers in $[0, 1]$, there are uncountably many ways for the coin to land.

hypothesis is true or false, I'll just use '$\mathfrak{c}$'. But note that $\mathfrak{c}$, whichever ordinal it is, is the leader of its Archimedean level. For no infinite cardinal is any real multiple of a smaller infinite cardinal. So $\mathfrak{c}^{-1}$ is some power of ω. That is, there's some surreal number x such that $\mathfrak{c}^{-1} = \omega^{-x}$. And thus, for any surreal number y, $\mathfrak{c}^{-y} = \omega^{-xy}$, by the rules for powers of ω. So if we say that the probability of some $A \subseteq \mathscr{W}$ is $r \cdot \mathfrak{c}^{-z}$, for some real number r and some surreal number z, we are writing this probability in Conway normal form, and we can continue to use the arithmetic rules from §1.2.

For the case of flipping a fair coin infinitely often, the standard approach to defining probabilities starts with propositions about how the coin lands on some finite number of flips. We then say that the probability of a proposition specifying how the coin lands on n flips is 2^{-n}, and we appeal to common 'extension' theorems showing that these probabilities can be (uniquely) extended to a (countably additive) probability distribution over a σ-algebra containing all the propositions about how the coin lands on the first n flips.[22] On the orthodox approach, propositions like 'the coin lands heads on every odd flip', $\mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\mathrm{h}_7\ldots$, end up getting defined by the extension theorems. Because orthodoxy only uses real-valued probabilities, all propositions like this are assigned probability zero. Since we want to assign non-zero probability to propositions like these, we cannot rely on the standard extension theorems to define surreal probabilities. So I will start with a different set of propositions.

These propositions will specify how the coin lands on particular flips. So they will be (perhaps infinite) conjunctions of propositions of the form 'h_n'. I won't start off assigning probabilities to disjunctive propositions like 'the coin either lands heads on every odd flip or it lands heads on each flip which is a multiple of four'. (Though once we've assigned a probability to both 'the coin lands heads on every odd flip', 'it lands heads on each flip which is a multiple of four', and 'it lands heads on any flip which is either odd or a multiple of four', we will be able to determine the probability of this disjunctive proposition.) I assume that, since the coin is fair, making changes to *how* a proposition specifies the outcome of certain flips without changing *whether* it specifies those outcomes won't make any difference to the proposition's probability. For instance, the proposition $\mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\mathrm{h}_7\mathrm{h}_9\ldots$, which says that the coin lands heads on every odd flip, will have the same probability as the proposition $\mathrm{t}_1\mathrm{h}_3\mathrm{t}_5\mathrm{h}_7\mathrm{t}_9\ldots$, which says that the coin lands tails on every $4n-3$rd flip and heads on every $4n-1$st flip. So I'll limit attention to propositions which say that the coin lands *heads* on certain flips. Once we've given probabilities to these propositions, it's straightforward to extend

22. A σ-algebra $\mathscr{A}$ is a set of propositions containing $\mathscr{W}$ which is closed under complementation and countably union. That is, $\mathscr{W} \in \mathscr{A}$, $A^c \in \mathscr{A}$ whenever $A \in \mathscr{A}$, and $\bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$ whenever $A_1, A_2, \cdots \in \mathscr{A}$. For details, see Rosenthal, 2006.

those probabilities to propositions which call tails landings.

Say that a proposition about how the coin lands is *periodic* whenever the flips it calls eventually settle into a repeating pattern. That is, a proposition is periodic whenever there's some N and some p (the *period* of the proposition) such that, for every $n > N$, the proposition calls flip n if and only if it calls flip $n + pk$, for every positive integer k. If a proposition is periodic, then we can specify what that proposition says by saying how it calls some finite number of flips before it settles into its repeating pattern and then saying how the pattern repeats from there. For instance, the proposition that the coin lands heads on every flip is periodic with a period of 1. The proposition that the coin lands heads on every odd flip is periodic with a period of 2. For a more interesting example, take the proposition that the coin lands heads on every prime flip less than 100, and then on flips 101, 110, 200, 201, 210, 300, 301, 310 and so on. This proposition is periodic with a period of 100. The repeating pattern starts after flip 100; from there on out, it calls the first, tenth, and hundredth of every hundred flips.

Notice that the set of periodic propositions is closed under finite intersection. If A is periodic with a period of p and B is periodic with a period of q, then AB will be periodic with a period of pq. So the set of periodic propositions is what's known as a meet semi-lattice. In the appendix, I show that probabilities defined over a meet semi-lattice may be uniquely extended to a probability over a full algebra of propositions. So if we define surreal probabilities for every periodic proposition, we will have thereby defined surreal probabilities for any Boolean combination of the periodic propositions.

I will make two assumptions about the surreal probabilities of coin flipping. These assumptions leverage what we already know about the probabilities of propositions calling a finite number of flips to tell us something about the probabilities of propositions calling an infinite number of flips. Firstly, I will assume that periodic propositions about different coin flips are independent. This will hold for any proposition which calls only finitely many heads. For instance, $\mathrm{h}_1\mathrm{h}_5\mathrm{h}_{117}$ is independent of $\mathrm{h}_9\mathrm{h}_{110}\mathrm{h}_{200}\mathrm{h}_{201}$. I will assume that the same is true even when a proposition calls infinitely many heads. So, for instance, the proposition that the coin lands heads on every odd flip, $O = \mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\mathrm{h}_7\ldots$, is independent of the proposition that the coin lands heads on every even flip, $E = \mathrm{h}_2\mathrm{h}_4\mathrm{h}_6\mathrm{h}_8\ldots$.

Independence If A and B are propositions about different coin flips, then A and B are independent.

With the periodic propositions, it's easy to specify what we mean by the proposition being *about* a particular coin flip. If A is periodic, then A is about flip n iff it includes h_n as a conjunct. With other propositions, matters get murkier. Consider the propositions 'the coin lands heads on only finitely many flips' and 'the coin lands heads on only finitely many flips after the first'. It can seem that the first proposition is about the first flip in a way that the second proposition is not—but these are one and the same proposition.[23]

Given any periodic proposition A, let 'A_n' say whatever A says about the first n flips, and no more. For instance, if O is the proposition that the coin lands heads on all *o*dd flips, $\mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\ldots$, then $O_{10} = \mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\mathrm{h}_7\mathrm{h}_9$. And if H is the proposition that the coin lands heads on every *h*undredth flip, then $H_{320} = \mathrm{h}_{100}\mathrm{h}_{200}\mathrm{h}_{300}$. It could be that A says nothing about the first n flips. In that case, A_n will be the tautology, true of every sequence, $A_n = \mathscr{W}$.[24] My second assumption will be that we can use the limiting behavior of the probabilities of A_n and B_n to infer something about the Archimedean levels of $\mathbb{P}(A)$ and $\mathbb{P}(B)$. In particular, suppose that the ratio $\mathbb{P}(A_n) \div \mathbb{P}(B_n)$ diverges to infinity as n goes to infinity. Then, as the number of flips n gets larger, A_n gets increasingly more likely than B_n. There's some number of flips after which A_n is always at least twice as likely as B_n; some number of flips after which A_n is always at least thrice as likely as B_n, and so on and so forth. In this case, we should say that A is infinitely more likely than B. That is, we should put $\mathbb{P}(A)$ on a higher level than $\mathbb{P}(B)$. By the same token, if the ratio $\mathbb{P}(A_n) \div \mathbb{P}(B_n)$ converges to zero, then we should say that A is infinitely less likely than B, and we should put $\mathbb{P}(A)$ on a lower level than $\mathbb{P}(B)$. ($\mathbb{P}(A_n) \div \mathbb{P}(B_n)$ will converge to zero iff $\mathbb{P}(B_n) \div \mathbb{P}(A_n)$ diverges to infinity.) However, if as n goes off to infinity, the ratio remains within an interval $(1/r, r)$, for some real number r, then we should say that neither A nor B is infinitely more likely than the other, and we should put both $\mathbb{P}(A)$ and $\mathbb{P}(B)$ on the same level.

Let $\#A_n$ be how many of the first n flips the proposition A says something about. Then, the ratio $\mathbb{P}(A_n) \div \mathbb{P}(B_n)$ will be $2^{\#B_n - \#A_n}$. This number will diverge to infinity as n goes to infinity iff the difference $\#B_n - \#A_n$ grows without bound, going off to ∞ in the limit. So another way of putting this assumption is to say that, if $\#B_n - \#A_n$ diverges to ∞, then $\mathbb{P}(A)$ is on a higher level than $\mathbb{P}(B)$. If, however, as n goes off to infinity, $\#B_n - \#A_n$ remains bounded between $-N$ and N, for some natural number N, then $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are on the same level.

Probabilities and Limits Take any two periodic propositions, A and B. If, as n goes to infinity, the difference between the number of heads called by B by the nth flip and the number of heads called by A by the nth flip, $\#B_n - \#A_n$, is bounded, then $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are on the same Archimedean level. If, on the other hand, the difference $\#B_n - \#A_n$ diverges to ∞ as n gets larger, then $\mathbb{P}(A)$ is on a higher level than $\mathbb{P}(B)$.

23. Cf. Builes' paradox (Builes, 2020, and Dorr *et al.*, 2021).

24. In general, if W_n is the set of sequences which match the first n outcomes of W, then $A_n = \bigcup_{W \in A} W_n$.

n	1	2	3	4	5	6	7	8	9	10	$\cdots$
$\mathbb{P}(O_n)$	1/2	1/2	1/4	1/4	1/8	1/8	1/16	1/16	1/32	1/32	$\cdots$
$\mathbb{P}(E_n)$	1	1/2	1/2	1/4	1/4	1/8	1/8	1/16	1/16	1/32	$\cdots$
$\mathbb{P}(O_n) \div \mathbb{P}(E_n)$	1/2	1	1/2	1	1/2	1	1/2	1	1/2	1	$\cdots$

TABLE 1: As n goes off to infinity, the ratio $\mathbb{P}(O_n) \div \mathbb{P}(E_n)$ will fluctuate between one half and one forever.

3.1 | The Archimedean levels of surreal probabilities

With these assumptions, we can begin to calculate the Archimedean levels of the probabilities of periodic propositions. Consider A, which says that the coin lands heads on *a*ll flips, $A = \mathrm{h}_1\mathrm{h}_2\mathrm{h}_3\mathrm{h}_4\ldots$; E, which says that the coin lands heads on all *e*ven flips, $E = \mathrm{h}_2\mathrm{h}_4\mathrm{h}_6\mathrm{h}_8\ldots$; and O, which says that the coin lands heads on all *o*dd flips, $O = \mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\mathrm{h}_7\ldots$. Notice that, as n gets larger, the ratio of $\mathbb{P}(O_n) \div \mathbb{P}(E_n)$ will flip back and forth between one half and one forever. For odd n, the ratio is one half, and for even n, the ratio is 1. (See table 1.) So our assumptions tell us that O and E must be on the same level. Call that level, whatever it is, $\mathfrak{c}^{-x_0}$. So we can write that $\mathbb{P}(O) = \sum_{i\geqslant 0} r_i \cdot \mathfrak{c}^{-x_i}$ and $\mathbb{P}(E) = \sum_{j\geqslant 0} s_j \cdot \mathfrak{c}^{-x_j}$, for some increasing sequence of surreal numbers $x_0, x_1, x_2, \ldots$, and some real numbers $r_0, r_1, r_2, \ldots$ and $s_0, s_1, s_2, \ldots$ such that both r_0 and s_0 are non-zero. Next, notice that O and E are independent, and that their conjunction is just A. And we already know that the probability of A is $\mathfrak{c}^{-1}$. So we must have

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(O) \cdot \mathbb{P}(E) \\
\mathfrak{c}^{-1} &= \left(\sum_{i\geqslant 0} r_i \cdot \mathfrak{c}^{-x_i}\right) \cdot \left(\sum_{i\geqslant 0} s_i \cdot \mathfrak{c}^{-x_i}\right) \\
\mathfrak{c}^{-1} &= \sum_{i\geqslant 0} \sum_{j\geqslant 0} r_i s_j \cdot \mathfrak{c}^{-(x_i+x_j)}
\end{aligned}$$

Since the left-hand-side has just a single Archimedean part, the right-hand-side must also have just a single Archimedean part. So we must have all but one of the products $r_i s_j = 0$. Since we already know that $r_0 s_0$ is non-zero,

$$\mathfrak{c}^{-1} = r_0 s_0 \mathfrak{c}^{-2x_0}$$

So $x_0 = 1/2$, and $r_0 s_0 = 1$. So the Archimedean level of both $\mathbb{P}(O)$ and $\mathbb{P}(E)$ is $\mathfrak{c}^{-1/2}$, or $1/\sqrt{\mathfrak{c}}$. And their real multiples are reciprocals of each other. (I'll return to these real multiples in the next section, but let's put them to the side for now.)

Next, consider the proposition $T_1 = \mathrm{h}_1\mathrm{h}_4\mathrm{h}_7\mathrm{h}_{10}\mathrm{h}_{13}\ldots$, which says that the coin

n	1	2	3	4	5	6	7	8	9	10	$\cdots$
$\mathbb{P}(T_{1,n})$	1/2	1/2	1/2	1/4	1/4	1/4	1/8	1/8	1/8	1/16	$\cdots$
$\mathbb{P}(T_{2,n})$	1	1/2	1/2	1/2	1/4	1/4	1/4	1/8	1/8	1/8	$\cdots$
$\mathbb{P}(T_{3,n})$	1	1	1/2	1/2	1/2	1/4	1/4	1/4	1/8	1/8	$\cdots$
$\mathbb{P}(T_{1,n}) \div \mathbb{P}(T_{2,n})$	1/2	1	1	1/2	1	1	1/2	1	1	1/2	$\cdots$
$\mathbb{P}(T_{2,n}) \div \mathbb{P}(T_{3,n})$	1	1/2	1	1	1/2	1	1	1/2	1	1	$\cdots$

TABLE 2: As n goes off to infinity, the ratios $\mathbb{P}(T_{1,n}) \div \mathbb{P}(T_{2,n})$ and $\mathbb{P}(T_{2,n}) \div \mathbb{P}(T_{3,n})$ will fluctuate between one half and one forever.

lands heads on the first of every three flips, $T_2 = \mathrm{h}_2\mathrm{h}_5\mathrm{h}_8\mathrm{h}_{11}\mathrm{h}_{14}\ldots$, which says that the coin lands heads on the second of every three flips, and $T_3 = \mathrm{h}_3\mathrm{h}_6\mathrm{h}_9\mathrm{h}_{12}\mathrm{h}_{15}\ldots$, which says that the coin lands heads on the third of every three flips. Our assumptions tell us that T_1, T_2, and T_3 are all on the same level. For, as n gets larger, the ratios $\mathbb{P}(T_{1,n}) \div \mathbb{P}(T_{2,n})$ and $\mathbb{P}(T_{2,n}) \div \mathbb{P}(T_{3,n})$ will fluctuate back and forth between one half and one forever. (See table 2.) So they will be bounded, and none of T_1, T_2, and T_3 are infinitely more likely than any of T_1, T_2, and T_3. These three propositions are about different flips, so they are independent.[25] And their conjunction is $A = \mathrm{h}_1\mathrm{h}_2\mathrm{h}_3\mathrm{h}_4\ldots$. So we must have $\mathbb{P}(A) = \mathbb{P}(T_1)\cdot\mathbb{P}(T_2)\cdot\mathbb{P}(T_3)$. The kind of reasoning we went through in the previous paragraph shows that, in general, two numbers whose product has a single Archimedean part cannot have more than one Archimedean part. So we know that each of $\mathbb{P}(T_1)$, $\mathbb{P}(T_2)$, and $\mathbb{P}(T_3)$ have just one Archimedean part, and we can write that $\mathbb{P}(T_1) = r \cdot \mathfrak{c}^{-y}$, $\mathbb{P}(T_2) = s \cdot \mathfrak{c}^{-y}$ and $\mathbb{P}(T_3) = t \cdot \mathfrak{c}^{-y}$. Then,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(T_1) \cdot \mathbb{P}(T_2) \cdot \mathbb{P}(T_3) \\ \mathfrak{c}^{-1} &= (r \cdot \mathfrak{c}^{-y}) \cdot (s \cdot \mathfrak{c}^{-y}) \cdot (t \cdot \mathfrak{c}^{-y}) \\ \mathfrak{c}^{-1} &= rst \cdot \mathfrak{c}^{-3y}\end{aligned}$$

So $y = 1/3$, and $rst = 1$. So the level of $\mathbb{P}(T_1), \mathbb{P}(T_2)$, and $\mathbb{P}(T_3)$ is $\mathfrak{c}^{-1/3}$. By similar reasoning, $\mathbb{P}(T_1T_2)$, $\mathbb{P}(T_1T_3)$, and $\mathbb{P}(T_2T_3)$ are on level $\mathfrak{c}^{-2/3}$, with real multiples rs, st, and rt, respectively.

We can apply this kind of reasoning more generally. Pick any rational number q/p with both q and p integers greater than zero and q less than p. Then, let P_1 say that the coin lands heads on the first of every p flips, let P_2 say that it lands heads on the second of every p flips, and so on and so forth. The sequence of ratios $\mathbb{P}(P_i) \div \mathbb{P}(P_{i+1})$, for each $i < p$, will be bounded, so each $\mathbb{P}(P_i)$ must be on the same level. Each P_i is

25. Not only are T_1, T_2, and T_3 independent, but also T_1T_2 and T_3 are independent, since T_1T_2 and T_3 are about different flips.

independent of every other, since they each call different flips.[26] And $P_1 P_2 \dots P_p$ is the proposition $A = \mathrm{h}_1\mathrm{h}_2\mathrm{h}_3\mathrm{h}_4\dots$. So

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(P_1)\cdot\mathbb{P}(P_2)\cdot\dots\cdot\mathbb{P}(P_p)\\ \mathfrak{c}^{-1} &= (r_1\cdot\mathfrak{c}^{-z})\cdot(r_2\cdot\mathfrak{c}^{-z})\cdot\dots\cdot\left(r_p\cdot\mathfrak{c}^{-z}\right)\\ \mathfrak{c}^{-1} &= r_1 r_2\dots r_p\cdot\mathfrak{c}^{-pz}\end{aligned}$$

So $z = 1/p$, and $r_1 r_2 \dots r_p = 1$. By similar reasoning, we will get that $\mathbb{P}(P_1 P_2 \dots P_q)$ must live on the level $\mathfrak{c}^{-q/p}$.

So for every rational number between 0 and 1, there is a corresponding Archimedean level of surreal probabilities about how the coin lands. And since $\mathfrak{c}^{-x} > \mathfrak{c}^{-y}$ whenever $y > x$, the ordering of these levels is the reverse of the ordering of the rationals. So the levels of surreal probabilities are dense—between any two are infinitely many others.

3.2 | Shifted propositions

Given any proposition A, let $A \oplus m$ (read '*A shift m*') say about the $n+m$th flip whatever A says about the nth flip. For instance, if we let A be the proposition that the coin lands heads on *a*ll flips, $\mathrm{h}_1\mathrm{h}_2\mathrm{h}_3\mathrm{h}_4\dots$, then $A \oplus 1 = \mathrm{h}_2\mathrm{h}_3\mathrm{h}_4\mathrm{h}_5\dots$, $A \oplus 2 = \mathrm{h}_3\mathrm{h}_4\mathrm{h}_5\mathrm{h}_6\dots$, and so on. $A \oplus m$ must be 2^m times as likely as A. For A itself is just the conjunction of a proposition saying that the first m flips land heads and $A \oplus m$. So we must have $\mathbb{P}(A) = \mathbb{P}(\mathrm{h}_1\mathrm{h}_2\dots\mathrm{h}_m)\cdot\mathbb{P}(A\oplus m)$ and we know that $\mathbb{P}(\mathrm{h}_1\mathrm{h}_2\dots\mathrm{h}_m) = 2^{-m}$. Therefore, $\mathbb{P}(A \oplus m) = 2^m \cdot \mathbb{P}(A)$.

Williamson (2007) objects to non-Archimedean probabilities because he believes that $A \oplus 1$ should be just as likely as A. After all, any infinite sequence of heads landings is qualitatively indistinguishable from any other. Qualitatively indistinguishable outcomes should have the same probability. So the probability of $\mathbb{P}(A \oplus 1)$ should be equal to the probability of $\mathbb{P}(A)$. Moreover, we must have $\mathbb{P}(A) = \mathbb{P}(\mathrm{h}_1)\cdot\mathbb{P}(A \oplus 1)$, and $\mathbb{P}(\mathrm{h}_1) = 1/2$, so we must have $\mathbb{P}(A) = 0$.

What this shows us is that, if we think that there are non-Archimedean probabilities, then we must deny that qualitatively indistinguishable outcomes have the same probabilities. This is a cost, to be sure, but no way out is cost-free. The following claims are all independently plausible but jointly inconsistent given normalization: (a) $A \oplus 1$ is just as likely as A; (b) $A \oplus 1$ is twice as likely as A; and (c) A is more likely than a contradiction. Williamson's reasoning shows us that (a) and (b) imply that the probability of A must be zero. But if the probability of A is zero, then A is just as likely

26. Moreover, any conjunction of the P_i's is about different flips than any conjunction of the rest of the P_i's. So those propositions are independent, too.

as a contradiction. Williamson decides to reject (c) and say that A is no more likely than a contradiction; non-Archimedean approaches to probability reject (a) and say that qualitatively indistinguishable outcomes can have different probabilities.

Talk of one probability being twice another only makes sense given the multiplicative structure of some field of numbers. But as Williamson notes, we can formulate a similar argument using only qualitative comparative relations like 'X is more likely than Y'. Assume that this relation obeys a version of additivity according to which, if Z is disjoint from X and Y, then X is more likely than Y only if $X \cup Z$ is more likely than $Y \cup Z$. Then, the following two claims are inconsistent:

(1) $A \oplus 1$ is no more likely than A.

(2) Anything possible is more likely than anything impossible.

To appreciate the inconsistency, consider the proposition B, which says that the coin lands tails on the first flip and heads thereafter, $B = \mathrm{t}_1\mathrm{h}_2\mathrm{h}_3\mathrm{h}_4\ldots$. Since B is possible, (2) tells us that B is more likely than a contradiction, $\perp$. So the additivity assumption tells us that $A \cup B = A \oplus 1$ is more likely than $A \cup \perp = A$, contradicting (1).

It's no doubt counterintuitive to deny (1); but it is likewise counterintuitive to deny (2). Deciding between these counterintuitive conclusions requires a more systematic investigation of the total theories which endorse (1) and those which endorse (2). Shifting the flips you call forward doesn't seem like it should make a difference to the probability of your claim. But nor does it seem that shifting the guests in your hotel over one room should make a difference to whether you have any vacancies. Actual infinities require us to deny intuitive claims like these. Of course, denying intuitive claims is still a *cost*; but it appears to be the cost of admission to the realm of the infinite. You can seek solace in finitism; but if you wish to theorize about the probabilities of infinite sequences, there is a price to be paid. Either probabilities distinguish between qualitatively identical outcomes or else they do not distinguish between the possible and the impossible. You must pick your poison.

Consider again the proposition that the coin lands heads on every *o*dd flip, $O = \mathrm{h}_1\mathrm{h}_3\mathrm{h}_5\mathrm{h}_7\ldots$. We learnt in the previous subsection that O and $E = O \oplus 1$ are on the same Archimedean level, $\mathfrak{c}^{-1/2}$, and that their real multiples are reciprocals of each other. Moreover, since O is just the conjunction of h_1 and $O \oplus 2$, we know that $O \oplus 2$ is twice as likely as O. And since $O \oplus 1$ is just the conjunction of h_2 and $O \oplus 3$, we know that $O \oplus 3$ is twice as likely as $O \oplus 1$. This pattern continues. So we know that, for some real multiple r,

$$\mathbb{P}(O) = r \cdot \mathfrak{c}^{-1/2} \qquad \mathbb{P}(O \oplus 1) = r^{-1} \cdot \mathfrak{c}^{-1/2}$$
$$\mathbb{P}(O \oplus 2) = 2r \cdot \mathfrak{c}^{-1/2} \qquad \mathbb{P}(O \oplus 3) = 2r^{-1} \cdot \mathfrak{c}^{-1/2}$$

$$\mathbb{P}(O \oplus 4) = 4r \cdot \mathfrak{c}^{-1/2} \qquad \mathbb{P}(O \oplus 5) = 4r^{-1} \cdot \mathfrak{c}^{-1/2}$$
$$\vdots \qquad\qquad \vdots$$

Assuming that these probabilities do not get smaller as we shift O's calls further to the right, we must have r between $2^{-1/2}$ and 1, inclusive. As far as I can see, there's nothing forcing r to be any particular value in the interval $[2^{-1/2}, 1]$. We could let $r = 1$. Then, we would say that, while $O \oplus 1$ is just as likely as O, $O \oplus 2$ is twice as likely as $O \oplus 1$. This sudden change in probability strikes me as unmotivated. If shifting O's calls to the right twice makes those calls more likely, shifting them to the right once should also make them more likely. Moreover, it is natural to expect that the factor by which the probability increases should be the same with each rightward shift. So there should be some fixed multiple λ such that $\mathbb{P}(O \oplus m) = \lambda^m \cdot \mathbb{P}(O)$. Since $\mathbb{P}(O \oplus 2) = 2 \cdot \mathbb{P}(O)$, this implies that $\lambda = 2^{1/2}$. So $O \oplus 1$ should be $\sqrt{2}$ times as likely as O, and $O \oplus 2$ should be $\sqrt{2}$ times as likely as $O \oplus 1$. In general, shifting O's calls to the right m places makes it $2^{m/2}$ times more likely. Recall that, as n went off to infinity, the ratio $\mathbb{P}(O_n) \div \mathbb{P}(E_n)$ fluctuated back and forth between 2^{-1} and 2^0 forever. If we assume that shifting O's calls to the right makes them $2^{1/2}$ times as likely, then we assume that the ratio $\mathbb{P}(O) \div \mathbb{P}(E)$ is $2^{-1/2}$, which is the geometric mean of 2^{-1} and 2^0. With this assumption, we can solve for r. For, if $\mathbb{P}(O \oplus 1) = \sqrt{2} \cdot \mathbb{P}(O)$, we must have that $r^{-1} = \sqrt{2} \cdot r$. So $r = 2^{-1/4}$. And in general, we will have $\mathbb{P}(O \oplus m) = 2^{(2m-1)/4} \cdot \mathfrak{c}^{-1/2}$.

Or consider the proposition $T = \mathrm{h}_1\mathrm{h}_4\mathrm{h}_7\mathrm{h}_{10}\ldots$, which says that the coin lands heads on the first of every *t*hree flips. We saw in the previous subsection that $\mathbb{P}(T)$, $\mathbb{P}(T \oplus 1)$, and $\mathbb{P}(T \oplus 2)$ all live on the same Archimedean level, $\mathfrak{c}^{-1/3}$, and that their real multiples have a product of 1. For some r, $\mathbb{P}(T) = r \cdot \mathfrak{c}^{-1/3}$. And $\mathbb{P}(T \oplus 3) = 2r \cdot \mathfrak{c}^{-1/3}$. Assuming that each shift to the right increases the probability by a constant factor, that factor must be $2^{1/3}$. Then, in general $T \oplus m$ will be $2^{m/3}$ times as likely as T. Recall that, as n went off to infinity, the ratio between $\mathbb{P}(T_n)$ and $\mathbb{P}(T \oplus 1_n)$ cycled through the values $2^0, 2^0$, and 2^{-1} over and over. By assuming that $\mathbb{P}(T \oplus 1)$ is $2^{1/3}$ times as likely as $\mathbb{P}(T)$, we assume that $\mathbb{P}(T) \div \mathbb{P}(T \oplus 1) = 2^{-1/3}$, which is the geometric mean of $2^0, 2^0$, and 2^{-1}. If we assume that shifting T's calls to the right makes them $2^{1/3}$ times as likely, then we then we can solve for r in $\mathbb{P}(T) = r \cdot \mathfrak{c}^{-1/3}$. For we must have $\mathbb{P}(T \oplus 1) = r2^{1/3} \cdot \mathfrak{c}^{-1/3}$ and $\mathbb{P}(T \oplus 2) = r2^{2/3} \cdot \mathfrak{c}^{-1/3}$. We saw in the previous subsection that the product of these real multiples must be one. So we must have

$$\begin{aligned} r \cdot 2^{1/3} r \cdot 2^{2/3} r &= 1 \\ 2r^3 &= 1 \\ r &= 2^{-1/3} \end{aligned}$$

And, in general, we will have $\mathbb{P}(T \oplus m) = 2^{(m-1)/3} \cdot \mathfrak{c}^{-1/3}$.

We can generalize. Pick any positive integer p and let $P = \mathrm{h}_1\mathrm{h}_{p+1}\mathrm{h}_{2p+1}\mathrm{h}_{3p+1}\ldots$ say that the coin lands heads on the first of every p flips. We learnt in the previous subsection that, for any m, $\mathbb{P}(P \oplus m)$ is on the Archimedean level $\mathfrak{c}^{-1/p}$, and that the real coefficients of $\mathbb{P}(P), \mathbb{P}(P \oplus 1), \ldots$, and $\mathbb{P}(P \oplus p-1)$ (written in Conway normal form) have a product of 1. Assume that shifting P's calls to the right increases their probability by a constant multiple, λ. Then, since $\mathbb{P}(P \oplus p) = 2 \cdot \mathbb{P}(P)$, we must have $\lambda = 2^{1/p}$. So, in general, $P \oplus m$ will be $2^{m/p}$ times as likely as P. If $\mathbb{P}(P) = r \cdot \mathfrak{c}^{-1/p}$, then $\mathbb{P}(P \oplus m) = 2^{m/p} r \cdot \mathfrak{c}^{-1/p}$. We can therefore use the fact that the real coefficients of the probabilities of the first $p-1$ shifts of P multiply to one to solve for the value of r:

$$\begin{aligned} \prod_{m=0}^{p-1} 2^{m/p} \cdot r &= 1 \\ 2^{(p-1)/2} \cdot r^p &= 1 \\ r &= 2^{(1-p)/2p} \end{aligned}$$

And, in general, we will have

$$\mathbb{P}(P \oplus m) = 2^{(1-p+2m)/2p} \cdot \mathfrak{c}^{-1/p}$$

We've now said enough to determine the probability of every periodic proposition. For any periodic proposition can be decomposed into two parts: some finite number of initial flips, before the proposition settles into its repeating pattern of calls, and its periodic repeating pattern. We can refer to the first part of the proposition as its 'initial calls', and the second part as its 'periodic calls'. By independence, we can factor the proposition's probability into the probability of these two parts,

$$\mathbb{P}(A) = \mathbb{P}(A\text{'s initial calls}) \cdot \mathbb{P}(A\text{'s periodic calls})$$

If it doesn't have any initial calls, then this first factor will be 1; and if it doesn't have any periodic calls, the second factor will be 1. If A settles into a repeating pattern after N flips, and the repeating pattern has a period of p, then A's periodic calls can be expressed as a conjunction $\bigcap_{j=1}^{k} P \oplus N + m_j$, for some collection of integers $m_1 m_2, \ldots, m_k < p$. Each of these these conjuncts is about different flips, so by independence, we can get the probability of the conjunction by multiplying together the probability of the conjuncts. So, if A makes i initial calls before flip N and then, after flip N, settles into a repeating pattern of calling the m_1th, m_2th, …, and m_kth of every

p flips thereafter, then A's probability will be

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(i \textit{ initial calls}) \cdot \prod_{j=1}^{k} \mathbb{P}(P \oplus N + m_j) \\
&= 2^{-i} \cdot \prod_{j=1}^{k} 2^{(1-p+2N+2m_j)/2p} \cdot \mathfrak{c}^{-1/p} \\
&= 2^{[(k-kp+2kN+2\sum_{i=1}^{k} m_j)/2p]-i} \cdot \mathfrak{c}^{-k/p}
\end{aligned}$$

3.3 | *Further vistas*

The periodic propositions are an easily managed class of propositions. But there are many, many more propositions about how the coin lands than these. And their surreal probabilities lie on many, many more Archimedean levels of probability.

Take any propositions which says that the coin lands heads on the mth of every p flips, $P \oplus m$. Notice that this proposition is associated with a certain function, $f_{P\oplus m}(n) = pn + m$. This function lists the calls made by $P \oplus m$. First, it calls a heads on the $p(1) + m$th flip; second, it calls a heads on the $p(2) + m$th flip; and so on. It's natural to ask about higher-degree polynomials. What about the proposition $S = \text{h}_1\text{h}_4\text{h}_9\text{h}_{16}\text{h}_{25}\ldots$ which says that the coin lands heads on every square flip? This proposition is associated with the function $f_S(n) = n^2$. Or what about the proposition $Q = \text{h}_{17}\text{h}_{382}\text{h}_{1947}\text{h}_{6152}\ldots$ associated with the quartic function $f_Q(n) = 24n^4+5n-12$? Let's call a proposition with an associated polynomial listing the flips it calls a 'polynomial proposition'.

Suppose f_A lists A's calls and f_B lists B's calls. And suppose that, for any m, eventually $f_A(n) > f_B(n + m)$.[27] Then, for any m, eventually, B will call the $n + m$th head before A has a chance to call its nth head. So the difference between the number of heads called by B and the number called by A must grow without bound. In that case, our assumptions tell us that A is infinitely more likely than B, and $\mathbb{P}(A)$ is on a higher level than $\mathbb{P}(B)$. In the other direction, if there's some m such that $f_A(n)$ is always less than $f_B(n + m)$, then B will never get more than m calls ahead of A, so the difference between the number of heads called by B and the number called by A will not grow arbitrarily high. So, given our assumptions, A will not be infinitely more likely than B.

So we can say: if, for any m, eventually $f_A(n) > f_B(n+m)$, then A is infinitely more likely than B and $\mathbb{P}(A)$ is on a higher Archimedean level than $\mathbb{P}(B)$. On the other hand, if there's some m such that $f_A(n)$ is always less than $f_B(n + m)$, then A is not infinitely more likely than B. This allows us to read information about the Archimedean level

27. By 'eventually, $f_A(n) > f_B(n+m)$', I mean that there's some N such that, for all $n > N$, $f_A(n) > f_B(n+m)$.

of a proposition's probability off of its associated polynomial. For illustration, take the propositions S and T, corresponding to the polynomials n^2 and $n^2 + 10n$, respectively. S calls a head on every square flip, $S = \mathrm{h}_1\mathrm{h}_4\mathrm{h}_9\mathrm{h}_{16}\mathrm{h}_{25}\ldots$, and T calls a head $10n$ flips after the nth square, $T = \mathrm{h}_{11}\mathrm{h}_{24}\mathrm{h}_{39}\mathrm{h}_{56}\mathrm{h}_{75}\ldots$. For $m \geqslant 5$, $n^2 + 10n \leqslant (n+m)^2$ and $n^2 \leqslant (n+m)^2 + 10(n+m)$. So neither S nor T is infinitely more likely than the other, and $\mathbb{P}(S)$ and $\mathbb{P}(T)$ live on the same level.

As this example suggests, propositions with higher degree polynomials are infinitely more likely than those with lower degree polynomials. (The *degree* of a polynomial $f(n)$ is just the greatest power of n that appears in the polynomial.) If $d > c$, then for large n, $n^d \gg n^c$, and for fixed m, $n^d \gg (n+m)^c$, too. Similarly, if f_A and f_B have the same degree but f_A's highest degree coefficient is greater than f_B's, then A will be infinitely more likely than B. For large n, $5n^3 \gg 4n^3$, and for fixed m, $5n^3 \gg 4(n+m)^3$, too. And the same goes whenever one polynomial's highest degree coefficient is greater than another's. However, if f_A and f_B have the same degree and the same highest-degree coefficient, then neither will be infinitely more likely than the other. For $(n+m)^d = \sum_{i=0}^{d} \binom{d}{i} m^{d-i} n^i$ will include a term n^i for every degree $i \leqslant d$, and we can always choose m high enough that the coefficient of the degree n^i term is greater than the coefficient in the other polynomial's degree n^i term. So if f_A and f_B have the same degree and the same highest-degree coefficient, then there will always be some m such that $f_A(n) \leqslant f_B(n+m)$ and $f_B(n) \leqslant f_A(n+m)$. So neither A nor B will be infinitely more likely than the other.

We've seen that each 'linear' proposition (each proposition corresponding to a linear polynomial $pn + m$) lives on an Archimedean level $\mathfrak{c}^{-1/p}$ indexed by a rational number between -1 and 0. The simplest non-linear proposition is $S = \mathrm{h}_1\mathrm{h}_4\mathrm{h}_9\mathrm{h}_{16}\ldots$, which calls a heads on every square flip. S is infinitely more likely than every linear polynomial proposition. So it must live on an Archimedean level higher than $\mathfrak{c}^{-1/2}$, $\mathfrak{c}^{-1/4}$, $\mathfrak{c}^{-1/8}$, and so on. The simplest such level is $\mathfrak{c}^{-\epsilon}$ so, just to give it a home, let's place $\mathbb{P}(S)$ there. Then, notice that S is a conjunction of the propositions S^2_1 and S^2_2, corresponding to $(2n-1)^2 = 4n^2 - 4n + 1$ and $(2n)^2 = 4n^2$, respectively. Because the polynomials associated with S^2_1 and S^2_2 have the same degree and the same highest-degree coefficient, their probabilities must live on the same Archimedean level. They are about different flips, so the probability of their conjunction must equal the product of their probabilities. So they must live on the level $\mathfrak{c}^{-\epsilon/2}$. Likewise, S is a conjunction of S^3_1, S^3_2, and S^3_3, corresponding to the polynomials $(3n-2)^2 = 9n^2 - 12n + 4$, $(3n-1)^2 = 9n^2 - 6n + 1$, and $(3n)^2 = 9n^2$. These propositions have probabilities on the same level, since their associated polynomials have the same degree and the same highest-degree coefficients. And they are about different flips, so the probability of their conjunction is the product of their probabilities. So those probabilities must live on the Archimedean level $\mathfrak{c}^{-\epsilon/3}$. The pattern suggests that a polynomial proposition whose leading term is pn^2 will have a probability on the level $\mathfrak{c}^{-\epsilon/\sqrt{p}}$.

The next simplest polynomial proposition is $C = \mathrm{h}_1\mathrm{h}_8\mathrm{h}_{27}\mathrm{h}_{64}\ldots$, which calls a head on every cube flip. This proposition's probability must lie on an Archimedean level higher than $\mathfrak{c}^{-\epsilon/2}$, $\mathfrak{c}^{-\epsilon/4}$, $\mathfrak{c}^{-\epsilon/8}$, and so on. The simplest such level is $\mathfrak{c}^{-\epsilon^2}$. Let us place $\mathbb{P}(C)$ there. Notice that C is a conjunction of C_1^2 and C_2^2, corresponding to $(2n-1)^3 = 8n^3 - 12n^2 + 6n - 1$ and $(2n)^3 = 8n^3$, respectively. These propositions are characterized by polynomials with the same degree and the same highest-degree coefficients, so their probabilities must live on the same level; since they are about different flips, the probability of their conjunction is the product of their probabilities. So their probabilities must live on the level $\mathfrak{c}^{-\epsilon^2/2}$. The general pattern suggests that a polynomial proposition whose leading term is pn^3 will have a probability on the level $\mathfrak{c}^{-\epsilon^2/\sqrt[3]{p}}$.

Generalizing, the pattern suggests that a polynomial proposition whose leading term is pn^m will have a probability on the Archimedean level $\mathfrak{c}^{-\epsilon^{m-1}/\sqrt[m]{p}}$.

We can climb higher. There are propositions more likely than any polynomial proposition. Consider the proposition $\mathrm{h}_1\mathrm{h}_4\mathrm{h}_{27}\mathrm{h}_{256}\mathrm{h}_{3125}\ldots$, whose calls are listed by the function $f(n) = n^n$. This proposition is more likely than any polynomial proposition. So it must live on an Archimedean level higher than $\mathfrak{c}^{-\epsilon}$, $\mathfrak{c}^{-\epsilon^2}$, $\mathfrak{c}^{-\epsilon^3}$, and so on. And of course we can keep climbing from there. We can build a hierarchy of functions f_α, indexed by the countable ordinals. We start with $f_\omega(n) = n$, and let $f_{\alpha+1} = f_\alpha(n) + 1$. If α is a limit ordinal, we choose an ω sequence of ordinals less than α, $\beta_1, \beta_2, \ldots$ whose limit is α, and let $f_\alpha(n) = f_{\beta_n}(n)$.[28] For instance, we can let $f_{2\omega}(n) = f_{\omega+n}(n) = 2n$, and $f_{3\omega}(n) = f_{2\omega+n}(n) = 3n$. Proceeding in this way, we can let the polynomial $a_m n^m + a_{m-1}n^{m-1} + \cdots + a_1 n + a_0$ be the function indexed by the ordinal $a_m\omega^m + a_{m-1}\omega^{m-1} + \cdots + a_1\omega + a_0$, and we can let n^n be the function indexed by ω^ω. In this hierarchy of functions, each function indexed by a limit ordinal must correspond to a proposition which is infinitely more likely than any proposition corresponding to a function indexed by a lesser ordinal. So there will be as many ascending Archimedean levels of probability in this hierarchy as there are countable limit ordinals. And there will be many, many Archimedean levels of probability lying in between the levels in this hierarchy. After all, we will need levels to house all of the conjunctions of the propositions whose calls are listed by the functions in this hierarchy.

28. A common hierarchy of functions like this known as the 'slow-growing hierarchy'.

3.4 | *In Summary*

As the discussion in §3.3 makes vivid, even in the simple case of flipping a fair coin infinitely often, the space of surreal probabilities is truly gargantuan. Even when we confine ourselves to the periodic propositions, there are at least as many Archimedean levels of probability as there are rational numbers. And the Archimedean levels of probability for propositions about how the coin lands extend much, much further. Nonetheless, as we saw in §§3.1–3.2, the structure imposed by our relatively weak assumptions is enough to allow us to calculate many precise values for these probabilities. No doubt with stronger assumptions, we could show more.

4 | CONCLUSION

Pre-theoretically, probabilities in infinite domains appear to be non-Archimedean. The proposition that the coin lands heads on the first flip, h_1, appears to be infinitely more likely than the proposition that it lands heads on every odd flip, $h_1h_3h_5h_7\ldots$, which appears in turn to be infinitely more likely than the proposition that it lands heads on every flip, $h_1h_2h_3h_4\ldots$. The standard mathematical tools cannot respect these pre-theoretic judgments, because the standard mathematical tools represent probabilities with real-valued functions, and the real numbers are Archimedean.

The standard mathematical tools also deny plausible principles like 'anything possible is more likely than anything impossible'. Call this thesis *regularity*. In uncountably infinite domains, standard real-valued probabilities will assign a probability of zero to every possible outcome. And this is exactly the probability they assign to any impossible proposition. So real-valued probabilities have a habit of telling us that an impossible outcome is just as likely as the actual outcome, in violation of regularity. Repeated exposure has inured many of us to the oddity of this claim—but take a step back, and listen to it afresh: a radon atom decays at some particular time, and you ask 'what are the chances it would decay just then?'. If we're modeling the chances with real-valued functions, the model will tell us: *It was exactly as likely to decay just then as it was to both decay and not decay, which is just as likely as it being more massive than itself*. It seems to me that these violations of regularity are merely artifacts of our mathematical model—not to be taken too seriously. Of course, real-valued probabilities are enormously helpful tools, and the models they provide can tell us many true things. But they also tell some falsehoods, and they don't tell all there is to tell. When they tell us that one proposition is more likely than another, they always speak truly; but they don't always tell us when one proposition is more likely than another. Sometimes, they say that two propositions are equally likely when they are not. This is an expressive limitation of the model; there are more probabilistic distinctions to be drawn than the real numbers allow us to draw.

These probabilistic distinctions are relevant to rational choice. Suppose I offer you a choice between two bets—one of which pays out \$1 if the radon atom decays at t and the other of which pays out \$1 if the radon atom is more massive than itself. You should prefer the former. Likewise if one of the bets pays out \$1 if the radon atom decays at a time *other than t* and the other pays out \$1 if it decays at all. You should prefer the guaranteed \$1 to the bet which pays out \$1 with a probability which can only be represented in the real numbers with 100%.

The standard mathematical tools are not the only possible tools. Here, I've tried to put new tools on the table. In my view, these tools have much to be said for them. They not only respect the pre-theoretic judgments suggesting that probabilities are non-Archimedean. They additionally allow us to rigorously investigate the extent to which they are non-Archimedean in particular cases—how many distinct Archimedean levels of probability there are, and where on those levels the infinitely unlikely propositions sit. Surreal probabilities also evade some of the objections which have been levied against hyperreal probabilities. In particular, they are not susceptible to Hájek's complaint of ineffability. Hájek complains: "when we are dealing with sets on this tiny scale, where minute differences count for a lot, we want things pinned down *very* precisely. 'ϵ' is a nice name for an infinitesimal probability, but what is it *exactly*?"[29] In the surreal numbers, 'ϵ' is not just a name for any old infinitesimal number, never mind which—it is the name of a *specific*, *precise* infinitesimal number. It is the simplest number greater than zero but less than every dyadic fraction. You reach it in the binary tree to taking one step right to 1, then infinitely many steps to the left. It corresponds to the ω length sequence of one plus followed by infinitely many minuses, $(+, -, -, -, \dots)$.

Hájek, ms, and Pruss, 2013, note that even hyperreal-valued probabilities will have to give possible outcomes a probability of zero if the space of possibilities is large enough. For instance, Hájek imagines throwing an infinitely precise dart, not at the *real* number line, but instead at the *hyperreal* number line. Then, there is no way to give positive probability to every possible outcome. Hájek writes: "I envisage a kind of arms race: we scotched regularity for real-valued probability functions by canvassing sufficiently large domains: making them uncountable. The friends of regularity fought back, enriching their ranges: making them hyperreal-valued. I counter with a still larger domain: making its values hyperreal-valued. Perhaps regularity can be preserved over that domain by enriching the range again, as it might be, making it hyper-hyperreal-valued. I counter again with a yet larger domain: making its values hyper-hyperreal-valued." Unlike hyperreal-valued probabilities, surreal-valued probabilities are able to give positive probability to every individual outcome in *any* set-sized

29. Hájek, 2003, p. 293.

domain, no matter its cardinality. If you have a set of possibilities of cardinality α, a uniform surreal probability distribution will assign a probability of $\alpha^{-1} > 0$ to each possibility in the domain. With surreal probabilities, the only way to get the arms race going is to consider a proper class of possibilities too large to have a cardinality. There cannot be any regular surreal-valued probability over a proper class of possibilities—but that's because there cannot be any *function* over a proper class of possibilities; it's irrelevant what values we want the function to take on.

Here, I've only scratched the surface of surreal probabilities. Several outstanding questions remain—both philosophical and mathematical. Philosophical: should we require surreal probabilities to be countably subadditive? Or should we perhaps find a different way of taking infinite sums of surreal numbers, and say that surreal probabilities are countably additive (or fully additive), given this alternative understanding of infinite summation? Mathematical: can the extension theorems for real-valued probabilities be generalized to cover surreal-valued probabilities?[30] Is there a generalization of Lebesgue integration for surreal-valued measures?[31] A complete evaluation of surreal probabilities would require answering these questions—and they are unfortunately beyond my reach. However, in my view, surreal-valued probabilities are promising enough that these further questions are worth investigating.

30. For an introduction to the Carathéodory extension theorem, see Billingsley, 1995 or Rosenthal, 2006. For an overview of other extension theorems—especially extension theorems for merely finitely additive measures—see Bhaskara Rao & Bhaskara Rao, 1983.

31. For an overview of Lebesgue integration, see Bartle, 1966. There are several proposals for generalizing Riemann integration, though many of them seem to give the wrong results in some cases. For instance, they evaluate $\int_0^\omega e^x dx$ as e^ω, rather than $e^\omega - 1$. Costin & Ehrlich (2022) provide a method for extending real-valued functions on the reals to surreal-valued 'functions' on the surreals. They show how to define antiderivatives for these extended 'functions', and they use these antiderivatives to define surreal integration. Their integral correctly evaluates $\int_0^y e^x dx$ as $e^y - 1$.

REFERENCES

Alling, Norman L. 1986. "Fundamentals of Analysis of Surreal Number Fields." In *Rocky Mountain Journal of Mathematics*, **19** (1). [5]

Bartle, Robert G. 1966. *The Elements of Integration and Lebesgue Measure*. Wiley Classics Library Edition. New York: John Wiley & Sons, Inc. [35]

Benci, Vieri, Horsten, Leon, & Wenmackers, Sylvia. 2013. "Non-Archimedean Probability." In *Milan Journal of Mathematics*, **81**: 121–151. [2]

Benci, Vieri, Horsten, Leon, & Wenmackers, Sylvia. 2016. "Infinitesimal Probabilities." In *British Journal for the Philosophy of Science*, **69** (2): 509–552. doi:10.1093/bjps/axw013. [2]

Bhaskara Rao, K.P.S. & Bhaskara Rao, M. 1983. *Theory of Charges: A Study of Finitely Additive Measures*. London: Academic Press. [35]

Billingsley, Patrick. 1995. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, third edition. [35]

Builes, David. 2020. "A Paradox of Evidential Equivalence." In *Mind*, **129** (513): 113–127. doi:10.1093/mind/fzz046. [23]

Chen, Eddy Keming & Rubio, Daniel. 2020. "Surreal Decisions." In *Philosophy and Phenomenological Research*, **100** (1): 54–74. doi:10.1111/phpr.12510. [4]

Conway, John H. 1976. *On Numbers and Games*. Academic Press. [5], [8], [11]

Costin, Ovidiu & Ehrlich, Philip. 2022. "Integration on the Surreals." [5], [35]

Dorr, Cian, Hawthorne, John, & Isaacs, Yoaav. 2021. "Solving a Paradox of Evidential Equivalence." In *Mind*, **130** (520): 1159–82. doi:10.1093/mind/fzaa022. [23]

Easwaran, Kenny. 2014. "Regularity and Hyperreal Credences." In *The Philosophical Review*, **123** (1): 1–41. [2], [3]

Ehrlich, Philip. 1994. "All Numbers Great and Small." In *Real Numbers, Generalizations of the Reals, and Theories of Continua*, edited by Philip Ehrlich, Dordrecht: Springer Netherlands, 239–258. doi:10.1007/978-94-015-8248-3_9. [5]

Ehrlich, Philip. 2001. "Number Systems with Simplicity Hierarchies: A Generalization of Conway's Theory of Surreal Numbers." In *The Journal of Symbolic Logic*, **66** (3): 1231–1258. ISSN 00224812. [5], [14] Erlich, Philip. 2011. "Conway names, the simplicity hierarchy and the surreal number tree." In *Journal of Logic & Analysis*, **3** (1): 1–26. [5]

Goldblatt, Robert. 1998. *Lectures on the Hyperreals*. Graduate Texts in Mathematics. New York, NY: Springer. doi:10.1007/978-1-4612-0615-6. [2]

Goldbring, Isaac. 2022. *Ultrafilters Throughout Mathematics*, volume 220 of *Graduate Studies in Mathematics*. American Mathematical Society. [3]

Gonshor, Harry. 1986. *An Introduction to the Theory of Surreal Numbers*. London Mathematical Society Lecture Note Series. Cambridge: Cambridge University Press. [5], [8]

Hájek, Alan. 2003. "Waging War on Pascal's Wager." In *Philosophical Review*, **112** (1): 27–56. doi:10.1215/00318108-112-1-27. [4]

Hájek, Alan. 2003. "What Conditional Probability Could Not Be." In *Synthese*, **137** (3): 273–323. [3], [34]

Hájek, Alan. ms. "Staying Regular?" Available at `http://hplms.berkeley.edu/HajekStayingRegular.pdf`. [2], [34]

Hamkins, Joel David. 2024. "Answer to "In surreal numbers, what exactly is ω_1?"." MathOverflow. URL:https://mathoverflow.net/q/470755. [20]

Hofweber, Thomas. 2014. "Cardinality Arguments Against Regular Probability Measures." In *Thought: A Journal of Philosophy*, **3** (2): 166–175. doi:10.1002/tht3.125. [2]

Knuth, Donald E. 1974. *Surreal numbers*. Reading, Massachusetts: Addison-Wesley Publishing Co. [4]

Lewis, David K. 1980. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability*, edited by Richard C. Jeffrey, Berkeley: University of California Press, volume II, 263–293. [2]

Lewis, David K. 1981. "Causal Decision Theory." In *Australasian Journal of Philosophy*, **59** (1): 5–30. [2]

Parker, Matthew W. 2018. "Symmetry Arguments Against Regular Probability: A Reply to Recent Objections." In *European Journal for Philosophy of Science*, **9** (1): 8. doi:10.1007/s13194-018-0229-1. [2]

Pruss, Alexander R. 2012. "Infinite Lotteries, Perfectly Thin Darts and Infinitesimals." In *Thought: A Journal of Philosophy*, **1** (2): 81–89. doi:10.1002/tht3.13. [2]

Pruss, Alexander R. 2013. "Probability, Regularity, and Cardinality." In *Philosophy of Science*, **80** (2): 231–240. ISSN 00318248, 1539767X. [34]

Rosenthal, Jeffrey S. 2006. *First Look at Rigorous Probability Theory*. Singapore: World Scientific Publishing Company, second edition. [21], [35]

Rubinstein-Salzedo, Simon & Swaminathan, Ashvin. 2014. "Analysis on Surreal Numbers." In *Journal of Logic & Analysis*, **6** (5): 1–39. [5], [15]

Skyrms, Brian. 1980. *Causal Necessity*. New Haven: Yale University Press. [2]

von Neumann, John & Morgenstern, Oskar. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ, USA: Princeton University Press. [4]

Weintraub, Ruth. 2008. "How Probable is an Infinite Sequence of Heads? A Reply to Williamson." In *Analysis*, **68** (299): 247–250. doi:10.1111/j.1467-8284.2008.00747.x. [2]

Wenmackers, Sylvia. 2016. "Infinitesimal Probabilities." In *The Open Handbook of Formal Epistemology*, edited by Richard Pettigrew & Jonathan Weisberg, PhilPapers Foundation, 199–265. [2]

Wenmackers, Sylvia & Horsten, Leon. 2013. "Fair Infinite Lotteries." In *Synthese*, **190** (1): 37–61. doi:10.1007/s11229-010-9836-x. [2]

Williamson, Timothy. 2007. "How probable is an infinite sequence of heads?" In *Analysis*, **67** (3): 173–180. [2], [4], [26], [27]