

More Trouble for Regular Probabilities

Abstract. In standard probability theory, probability zero is not the same as impossibility. However, many have suggested that it should be—that only impossible events should have probability zero. In cases where infinitely many outcomes have equal probability, regularity requires that some probabilities are infinitesimal, but merely introducing infinitesimals does not solve all of the problems with regularity. We will see that regular probabilities are not invariant over rigid transformations, even for simple, bounded, countable, constructive, and disjoint sets of possible outcomes. Hence, regular chances cannot be determined by space-time invariant physical laws, and regular credences cannot satisfy seemingly reasonable symmetry principles. Moreover, the examples here are immune to the objections against Williamson’s infinite coin flips.

- 1 *Introduction*
- 2 *Groundwork*
 - 2.1 *Assumptions about probabilities*
 - 2.2 *Regular probabilities*
 - 2.3 *Euclidean probabilities*
 - 2.4 *Uniform distributions*
- 3 *Failures of invariance*
 - 3.1 *Rotations*
 - 3.2 *Translations*

- 3.3 *Reflections*
- 3.4 *Disjoint images*
- 4 *Bearing on the regularity debate*
 - 4.1 *Williamson's coin flips*
 - 4.2 *Weintraub's criticism*
 - 4.3 *Restricting invariance to finite events*
 - 4.4 *Physicality and meaning*
- 5 *Conclusion*

1 Introduction

In standard probability theory, events with probability zero are not necessarily impossible. Suppose for example that a dart will land on a dartboard in one of infinitely many positions. (The dart does not have to be infinitely fine or perfectly symmetrical.) Assume also that these outcomes all have the same real-valued probability. If that probability were not zero, then the probabilities of the possible outcomes would add up to more than one, violating the Kolmogorov probability axioms. In fact, some *finite* number of them would add up to more than one. So in fact the probability of each such outcome must be zero. Yet one of them must occur. So on the standard theory, probability-zero events are not only possible, they occur all the time.

Many have suggested that this should not be so—that probabilities should be *regular*, i.e., only impossible events should be assigned probability zero (Carnap [1950], [1963]; Kemeny [1955], [1963]; Shimony [1955], Jeffreys [1961], Edwards, Lindman,

and Savage [1963], De Finetti [1964], Stalnaker [1970], Lewis [1980], Skyrms [1980], Appiah [1985], Jackson [1987], Jeffrey [1992]). To those schooled in the standard theory, this may seem like a naïve misconception, but it is not unmotivated. For objective chances, regularity encodes the seemingly sensible principle that the possible is more likely than the impossible. For credences it represents a reasonable willingness to update one's expectations in light of evidence, since credences of zero cannot be modified by Bayesian updating. There is also a betting argument: On standard definitions of credence, assigning credence zero to an event means you will accept a bet in which, if the event occurs, you sacrifice something of value, and if it doesn't, you gain nothing. Such an agreement seems irrational if there is any possibility that the event could occur (Shimony [1955], Kemeney [1955], [1963]).¹

And regularity is achievable, even in cases where infinitely many possible outcomes have the same probability, provided we are willing to permit not only real-valued probabilities but *hyperreals*, which include real numbers, infinitesimals, and their sums. (A number of authors have taken this approach, e.g., Bernstein and Wattenberg [1967], Lewis [1980], [1981], Nelson [1987], Benci, Horsten, and Wenmackers [2013].) Then we can assign a non-zero *infinitesimal* probability to each of the equally probable dart throw outcomes, so that a sum of finitely many such probabilities will not be greater than one. The sum of a countable infinity of infinitesimals is not in general well defined, so to allow for hyperreal probabilities we must also relax the axiom of countable additivity to merely finite additivity. That is, we do not require that the probability of a disjunction of a countable infinity of mutually exclusive events is the sum of the probabilities of the disjuncts, since that doesn't always make sense; we only require that

the probability of a disjunction of *finitely* many mutually exclusive events is the sum of the probabilities of the disjuncts. And having so relaxed that requirement, we are free to assign a probability of one to the disjunction of all our dart throw outcomes. So if we are willing to allow such relaxations of the standard definition of probability, our dartboard probabilities can be regular.

However, Timothy Williamson ([2007]) has shown that, even if we allow infinitesimals and relax the additivity axiom, regularity still raises problems. He considered the probability that an infinite sequence of coin flips will come out all heads—an extremely unlikely outcome, but strictly possible, so advocates of regularity would like to assign it non-zero probability. But Williamson showed that if the probability such an event is determined by time-invariant laws, then it must be strictly zero, not a non-zero infinitesimal. There are objections against Williamson (e.g., Weintraub [2008]), but here we will see a more general problem that is immune to the main objections. Briefly, regular probabilities must vary under rigid transformations (rotations, translations, and reflections), not only for infinite sequences of events, but for individual events that occur in bounded space and time, such as dart throws and vacuum fluctuations. Consequently, regular chances cannot be determined by space-time invariant laws, and regular credences cannot reflect epistemic indifference between symmetric events. These are heavy prices to pay for regularity. In light of such costs, the arguments for regularity should be closely scrutinised, but we will not do that here. We will only show that there are high costs not previously noted.

Besides contributing to the case against a regularity requirement on probabilities, this paper has a second purpose. Recent work has developed new alternatives to Cantor's

theory of set size (Katz [1981], Benci and Di Nasso [2003], Benci, Di Nasso, and Forti [2007]; see Mancosu [2009] and Parker [forthcoming] for discussion and further references). Unlike Cantor’s theory of cardinals, “Euclidean” theories satisfy the *Elements’* Common Notion 5, “The whole is greater than the part,” in the sense that, in Euclidean theories, a set is always larger than its proper subsets. It is shown in Parker [forthcoming] that the sizes treated by such theories must be largely arbitrary, due in part to the fact that, in geometric contexts, they cannot be invariant under rigid transformations. But it is suggested there that Euclidean sizes may nonetheless have useful applications in areas such as non-standard probability theory, for several authors have connected Euclidean notions of size with regular and hyperreal probabilities (McCall and Armstrong[1989], Gwiazda [2008], [2010], Wenmackers and Horsten [2013], Benci, Horsten, and Wenmackers [2013]). Here we will see that, in fact, the limitations of Euclidean set sizes are also significant limitations for regular probabilities. The failures of rigid transformation invariance, which carry over from Euclidean sizes to regular probabilities, imply that the attractive and plausible symmetries normally afforded by uniform distributions are not available. So the second purpose of this paper is to convey the bad news that, even in application to probabilities, Euclidean sizes have serious disadvantages.

2 Groundwork

2.1 Assumptions about probabilities

As suggested above, we need to relax the standard axioms of probability a little in order to make regularity possible. We need to permit infinitesimal values in addition to real numbers and substitute finite additivity for countable additivity. But for our purposes,

the axioms can be relaxed even more. We only need a few assumptions about probabilities in order to show that regular probabilities are not invariant under rigid transformations. Hence our conclusions will be quite general. Even if we adopt a *much* more relaxed notion of probability, we can't have a regular probability function that applies to the simple examples described below and is invariant under translations, rotations, or reflections.

We will still assume that a *probability space* consists, as usual, of three components, a *sample space* S , an *algebra* D , and a *probability function* P .² The sample space S is just any non-empty set. In applications, it represents the mutually exclusive possible outcomes of some experiment or chance occurrence, such as the places where a dart could hit a dartboard. The *algebra* D is a non-empty set of subsets of S that is closed under complementation as well as finitary union and intersection. That is,

(i) D contains at least one set,

and if A and B are sets in D , then

(ii) the complement $S \setminus A$ (consisting of all elements of S that are not in A) is in D , and

(iii) $A \cap B$ is in D .

It follows that for any A and B in D , $A \cup B$ is also in D , as are S and \emptyset .

Normally, the function P is a map from the algebra D into the real numbers. Here we will require only that P maps D into some set F equipped with a relation ' $>$ ', an operation '+', and an element '0' such that

(iv) for any x and y in F , if $x > 0$ then $y + x > y$,

and

(v) (finite additivity) if A and B are disjoint sets in D then $P(A \cup B) = P(A) + P(B)$.

These properties are all implied by the standard probability axioms. There is a long tradition of arguments supporting those axioms, to the effect that, holding credences or ascribing chances that violate those axioms would be irrational, or at the very least disadvantageous (e.g., Ramsey [1931], De Finetti [1964], Joyce [1998], [2009], Leitgeb and Pettigrew [2010a], [2010b]). However, such arguments usually attempt to justify finite additivity rather than the more controversial axiom of countable additivity (see Adams [1962–1964] and Williamson [1999] for noted exceptions), and most take the assumption that probabilities are real numbers, rather than hyperreals or something else, entirely for granted. So it is not unreasonable to at least consider relaxing those assumptions, as we will here. But our assumptions (i)–(v) have strong support, and are also quite weak. Besides following from the axioms of probability, they are implied by the definition of a measure or a finitely additive measure. So our conclusions here apply not only to probabilities, but to a broad class of functions and notions of size.

2.2 Regular probabilities

Regularity can take strong and weak forms. Let's say that a probability function P is *weakly regular* if every non-empty set of possible outcomes that has a probability at all has probability greater than zero. That is, if $A \in D$ and $A \neq \emptyset$ then $P(A) > 0$. We'll say P

is *strongly regular* if it is are weakly regular and P is defined on *all* subsets of S , i.e., if D is the entire power set $\mathcal{A}(S)$. In applications, if P is only weakly regular but not strongly, this means there are some sets of possible outcomes that do not have positive probability, because they do not have probabilities at all. Hájek [unpublished] gives several arguments that rational credences need not always be strongly regular, and in some cases shouldn't be.³ But it is weak regularity that mainly concerns us here, for weak regularity is enough to contradict rigid transformation invariance, provided that P is defined for just a few simple, bounded, Lebesgue-measurable⁴ point sets.

2.3 Euclidean probabilities

To see why regular probabilities don't have rigid transformation invariance, we will use the following fact: Given finite additivity or better, weak regularity implies *Euclideanism*. As noted above, Euclidean set sizes satisfy Common Notion 5, 'The whole is greater than the part'. Similarly we will call a probability function Euclidean if a set of possible outcomes always has greater *probability* than any of its proper subsets (provided both sets have probabilities). So P is Euclidean if for all sets A, B in D such that A is a proper subset of B , $P(B) > P(A)$.

It's easy to see why weakly regular probabilities are always Euclidean: If A is a proper subset of B then $B = A \cup (B \setminus A)$ and $B \setminus A$ is non-empty. So if A and B have probabilities at all, then by additivity, $P(B) = P(A) + P(B \setminus A)$, and by regularity, $P(B \setminus A) > 0$. By (iv), then, $P(B) > P(A)$. Thus regularity implies that whenever A is a proper subset of B , $P(B) > P(A)$, which is Euclideanism.⁵

One can draw stronger connections between Euclidean set sizes and regular probabilities. Benci, Horsten, and Wenmackers [2013] have introduced a theory of “Non-Archimedean Probabilities” (NAPs), which are regular probabilities closely related to a family of Euclidean set sizes called *numerosities* (Benci and di Nasso [2003], Benci, Di Nasso, and Forti [2007]). If a NAP P is “fair”—if it assigns the same probability to each single-element subset of a sample space—then $P(A)$ is just the numerosity of A divided by the numerosity of the sample space.

NAPs have some nice features. For one thing, they apply to *all* subsets of a continuum, not just certain “measurable” subsets, as in standard probability theory, and one can even obtain a total, finitely additive probability function with *standard* real values by taking the real parts of NAPs. Such totality may be surprising to those familiar with non-measurable sets and the Banach-Tarski paradox, which seem to show that a reasonable measure cannot be defined over all point sets. But totality comes at a cost, namely rigid transformation invariance. The paradoxical appearance of non-measurable sets derives from the intuition that natural measures are translation- and rotation-invariant. For certain highly complex, non-constructive sets, such invariance turns out to be incompatible with additivity. But here we will see that *regular* measures are even more limited: A *regular* rigid transformation-invariant measure can’t even be defined on certain simple, countable, constructive, Lebesgue-measurable point sets in one-dimensional space, even if we replace real-valued probabilities with hyperreals and countable additivity with finite additivity.

Defenders of Euclideanism may feel that rigid transformation invariance should not be expected where a set includes a rigid transformation of itself as a proper subset.

After all, in their eyes, proper subsets are *smaller*. But we will see examples where this argument doesn't apply—where a set and its rotation or translation are *not* related by proper inclusion, and do not even overlap, but still cannot have the same probability if regularity holds.

2.4 Uniform distributions

A standard example in conventional real-valued probability theory is the *uniform distribution* over a bounded region of Euclidean space. It's given by the probability function $P_{uniform}(A) = L(A) / L(S)$, where L is Lebesgue measure, on the Lebesgue-measurable subsets A of a region S . If S is a subset of the real line, for example, then $P_{uniform}(A)$ is just the total length of A in units chosen so that $P_{uniform}(S) = 1$ (provided A is Lebesgue measurable). Under this probability function, all sets with Lebesgue measure zero have probability zero, and sets that are not Lebesgue measurable get no probability at all. The term 'uniform' is justified by the fact that, for any real number r , $P_{uniform}$ assigns the same probability to all open balls of radius r , and more generally it assigns the same probability to any two geometrically congruent point sets (if those sets are Lebesgue measurable). For any rigid transformation T on a subset of Euclidean space and any Lebesgue measurable $A \subseteq S$, if $TA \subseteq S$ then $P_{uniform}(TA) = P_{uniform}(A)$.

Uniform distributions are especially popular as prior probabilities, since they are thought by some to represent complete ignorance. This is expressed by the much maligned Principle of Indifference (Keynes [2004]) and the more sophisticated Maximum Entropy Principle (Jaynes [1968]). However, these principles do not pick out the uniform distribution uniquely unless we presuppose a privileged coordinate system or background

measure (Uffink [1995]). Jaynes ([1973]) argues that the choice of prior probabilities can be narrowed or entirely determined by symmetry considerations. For example, if a problem concerning a probability distribution over a dartboard does not state the position or orientation of the dartboard, then in order for the problem to have a unique solution, we must assume that that doesn't matter, i.e., that the solution is invariant under translations and rotations. Jaynes also argues that such considerations are likely to yield accurate predictions in physical applications, because distributions that have such symmetries require less "skill" (Jaynes's scare quotes). Jaynes's views warrant serious doubts (see Gyenis and Rédei [forthcoming], §7, for a concise criticism), but they nonetheless exemplify the special applications that uniform distributions and rigid transformation invariance have.

In any case, we normally expect it to be at least *possible* to assign rational credences or set up objective chances in such a way that, under some coordinate system, congruent point sets get the same value. The well known Bertrand-type paradoxes have caused considerable trouble for the supposed *privilege* of symmetric credences. They show that the Principle of Indifference can't be applied without presupposing a privileged background measure or coordinate system. In this sense symmetric credences are underdetermined. But if regularity holds, symmetric credences aren't just underdetermined or arbitrary, they're ruled out altogether; we *can't* rationally accept, and the world *can't* exhibit, any translation- or rotation-invariant distribution, even *given* a privileged frame.

It is also worth noting that one of the motivations for regular probabilities is to make certain "fair" games possible, such as a fair infinite lottery (McCall and Armstrong

[1989], Wenmackers and Horsten [2013], Benci, Horsten, and Wenmackers [2013]). But on one conception (e.g., Skyrms [1995]), fairness implies translation invariance, and we will now see that such invariance is incompatible with regularity.

3 Failures of transformation invariance

3.1 Rotations

The following example is central to this paper, for the others all derive from it, though they each have specific points to make. Take a radial line segment—a spoke if you will—from centre of a dartboard to the rightmost point on its edge. Call that spoke r_0 . So in polar coordinates, the spoke $r_0 = \{(r, 0) : r \in [0, 1]\}$, assuming the dartboard has radius one. We're going to rotate this spoke around the board to obtain infinitely many spokes, and then take their union—the set containing all the points in those spokes. So let T be an anti-clockwise rotation of one radian about the centre of the board, i.e., $T(r, \theta) = (r, \theta + 1)$. Now let R be the union of r_0 and all its images by T , i.e., $R = \bigcup_{n \in \mathbb{N}} T^n r_0$. Since one radian is incommensurable with a full revolution of 2π radians, the spokes $T^n r_0$ will never coincide; we will never have $T^n r_0 = T^m r_0$ for $n \neq m$. Thus R is a union of a countable infinity of spokes. Now, if we rotate this set R by one radian, we obtain a proper subset $TR = \bigcup_{n > 1} T^n r_0$. It's a proper subset because the first spoke r_0 in R is not included in TR .⁶ Ergo, if a probability function P is regular, and therefore Euclidean, then P must assign a larger probability to R than to the rotation TR , if it assigns them probabilities at all. So rotation invariance fails for any regular probability defined on such sets.

Consider what this means: It's impossible to have a regular distribution on the dartboard that's indifferent to rotations. Suppose that the possible outcomes of a dart throw correspond in a natural physical way to the exact points in the plane of the dartboard. (Again, we do not need any silliness about infinitely fine or perfectly symmetric darts; it is enough if the possible positions of the dart form a continuum.) Let that $P(A)$ be the probability that the dart lands in a position corresponding to a point in A . Regularity implies that $P(TR) < P(R)$. So if physical chances are regular, then the dart must somehow discriminate between point sets that are exactly alike except for an angle of rotation. No matter how *badly* one throws the dart, no matter how *little* control is exercised, the darts will still discriminate in this incredibly fine way. That's physically implausible. And if rational credences must be regular, they too must so discriminate; no matter how symmetric the dartboard is or how little we know about the thrower, we can't rationally give the same credence to the dart hitting a point in R as we do for a rotation of R . (In contrast, the standard uniform distribution on the dartboard is rotation invariant, so it assigns the same probability to R and TR , namely zero.)

Again, defenders of regularity and Euclideanism might dismiss such rotation dependence as perfectly natural. After all, R contains all of the points in TR as well as others, so from a Euclidean viewpoint, R *should* be assigned higher probability. But as we have just seen, what seems natural from a regularist or Euclidean perspective has strange and limiting consequences. It implies that we can't have the sort of symmetry we expect from space-time invariant laws or unprejudiced beliefs. Objective Bayesians like Jaynes would say that if we lack other information, then rationally, we *must* adopt rotationally symmetric probabilities for our dart experiment. But even if we don't accept

Jaynes’s arguments, we can surely imagine situations where rotationally symmetric probabilities are appropriate. Suppose a needle rotates at constant speed for many revolutions. Elsewhere, in a small vacuum, pairs of particles spontaneously appear and annihilate each other. Given a set A of angles, what is the probability that one such vacuum fluctuation begins exactly when our needle is at an angle $\theta \in A$? Other things being equal, shouldn’t it be the same as the probability that a fluctuation begins when the needle is at an angle $\theta \in TA$, where T is a rotation? But given regularity, this *can’t* be so for all sets of angles.

And it gets worse.

3.2 Translations

Not only rotations on a disk, but also translations and reflections within an interval fail to preserve regular probabilities. To see this, we will construct a set of points on the real line analogous to our set of spokes on the dartboard. Let’s say a *translation mod 1* is a transformation T on the half-open unit interval $[0, 1)$ equivalent to first performing a translation and then taking the fractional part of the result. That is, for some $c \in [0, 1)$,

$$Tx = x + c \pmod{1} = \begin{cases} x + c & \text{if } x + c < 1, \\ x + c - 1 & \text{otherwise.} \end{cases}$$

So a translation mod 1 is a “piecewise translation”, so to speak, made up of two translations: T_1 , mapping $[0, 1 - c)$ rightwards to $[c, 1)$, and T_2 , mapping $[1 - c, 1)$ leftwards to $[0, c)$.

Like rotations of the circle, translations mod 1 fail to preserve regular probabilities. For suppose c is irrational. Then the points $T^n 0$ never coincide for different whole numbers n . Now let $X = \{T^n 0: n \in \mathbf{N}\}$. Then $TX = \{T^n 0: n \geq 1\}$ is a proper subset of X . So if X is regular, then it's Euclidean, and hence if $P(X)$ and $P(TX)$ are defined, $P(TX)$ is less than $P(X)$. Thus translations mod 1 that are defined on such sets don't preserve regular probabilities.

Now, translations mod 1 aren't translations per se, but the fact that translations mod 1 don't preserve regular probabilities implies that true translations don't either. Assume that $P(X)$, $P(TX)$, and $P([0, 1 - c])$ are defined. Then P is also defined for intersections and complements of these sets, by our assumptions (ii) and (iii). Now let $X_1 = X \cap [0, 1 - c)$ and $X_2 = X \cap [1 - c, 1)$. Then $T_1 X_1$ and $T_2 X_2$ are disjoint. Since $P(X) \neq P(TX)$, it follows by additivity that either $P(X_1) \neq P(T_1 X_1)$ or $P(X_2) \neq P(T_2 X_2)$. So at least one of the translations T_1 or T_2 fails to preserve P . Thus translations do not preserve weakly regular probabilities that are defined on simple sets like X , TX , and $[0, 1 - c)$. A fortiori, *strongly* regular probabilities on an interval are *never* preserved by all translations.

3.3 Reflections

As a quick corollary to the above, we can infer that reflections on the real line don't preserve regular probabilities either. This is because every translation is a composition of two reflections. We know that the translation T above does not preserve regular probabilities, and T can be written as a composition of two reflections R_1 and R_2 , so at least one of these reflections must fail to preserve regular probabilities.

3.4 Disjoint images

Notice that we are no longer talking about cases where TB is a subset of B , and hence where the intuitions supporting regularity and Common Notion 5 directly justify a difference in probabilities. We can't say, "Of course the translated image is smaller than the original, because it's a proper subset"; our T_1X_1 in §3.2 is *not* a subset of X_1 , and T_2X_2 is not a subset of X_2 . In fact, if P is defined on sufficiently small intervals in $[0, 1)$ as well as the set X , then we can just split X up into smaller sets X_i such that each TX_i is disjoint from X_i . By additivity, then, there is at least one such set X_i such that $P(X_i) \neq P(TX_i)$.⁷ A similar argument applies to reflections.⁸ So if a regular probability measure is defined over sufficiently small intervals (as probability measures on a continuous space normally are) then for any translation or reflection T there are disjoint sets A and TA that differ in probability. This inconvenient inequity cannot be made more palatable by pointing out that TA is proper subsets of A , because it isn't.

Hence, on the regularist or Euclidean view, it is impossible to choose a random number in the interval so that no set is privileged over any of its disjoint translations, nor over its disjoint reflections. We cannot throw a dart at a rectangular dartboard in such a way that it is as likely to hit a point with x -coordinate in a set A as in a disjoint translation or reflection TA . Likewise, if quantum fluctuations occur in some vacuous region, there will be bounded sets A of points such that a fluctuation is slightly more likely to occur at a point in A than in certain disjoint translations and reflections of A , and similarly there will be bounded sets B of *times* such that a fluctuation is more likely to occur at a time $t \in B$ than in certain disjoint translations and reflections of B . Furthermore, if regularists

about rational credence are right, the certain sets of events must be given higher credence than some of their disjoint translations and reflections.

4 Bearing on the regularity debate

4.1 Williamson's coin flips

Let us now relate these observations to the ongoing debate over regular probabilities. Williamson [2007] considers a countably infinite sequence of independent coin flips. Let's call this sequence of flips s . He then considers the proper subsequence s' beginning with the second flip in s . Let us write $H(x)$ for the proposition that a given sequence x of coin flips comes out all heads. For the regularist, $0 < P(H(s)) < P(H(s'))$, since (i) $H(s)$ is possible and (ii) the outcomes where $H(s')$ occurs are a proper subset of those in which $H(s)$ occurs. Williamson argues very effectively (though in a tentative, open-minded spirit) that this is mistaken. Assume the two sequences of flips are identical in their qualitative physical properties. If physical circumstances determine chances (and by the Principal Principle, well informed rational credences), then $P(H(s))$ should not differ *at all* from $P(H(s'))$, even by an infinitesimal. Likewise, if an entirely separate sequence t of coin flips begins at the same time as s' and carries on in parallel, then $P(H(t))$ should be the same as $P(H(s))$ and $P(H(s'))$, provided the physical circumstances are the same in all three cases.

4.2 Weintraub's criticism

Weintraub [2008] responds to Williamson by pointing out that the coin flips in s' and t occur at different times from those in s . The former times are a proper subset of the

latter. In this way the physical circumstances *are* different, so there is no paradox if the respective probabilities differ.

Stated that way, such a response misses an important point. It is true that there is a physical difference between the two sequences of coin flips, but the difference seems to concern only the times at which they occur, and it is a time-honoured principle that the laws of physics do not depend on or change over time. If we find that physical systems behave differently at different times, we look for some other change in the physical circumstances besides the time itself. This may be only a methodological convention (Poincaré [1911]) vulnerable to eventual rejection (Quine [1951]), but it is a usefully simplifying one, and to drop it would dramatically change our picture of how the world works. It would amount to saying that the way things behave changes over time for no underlying reason. So while Williamson's argument doesn't show that regularity is *logically* paradoxical, it puts the regularist in a very awkward dilemma: She must either abandon the standard and sensible principle that physical laws are unchanging and time invariant, or deny that probabilities are determined by physical circumstances and laws.⁹

But Weintraub might argue that this is not so, on the grounds that the difference between s , s' , and t is not *just* a difference of in when the flips occur. The asymmetric relations of *proper inclusion* between s and s' and between the times of the flips in s and t , can themselves be seen as differences of physical circumstance, and might be responsible for the differences in probabilities. Weintraub has further suggested (personal communication) that we might understand such differences in terms of duration. If a rod is heated for a longer duration than another rod, the resulting effects are

different. And Williamson's sequences s' and t can be thought of as shorter in duration than s , since they start later. Of course, this is only true under a very non-standard Euclidean notion of duration. Conventionally we would say that the durations of all three sequences are infinite, or perhaps undefined, but not unequal. So Weintraub's suggestion implies Euclideanism not only for probabilities but for durations too, but regularists might be perfectly comfortable with this.

We might reply to Weintraub that, even if the sequences differ in physical respects other than merely occurring at different times, they still differ only in respects that are entirely *due* to occurring at different times. Regularity implies that we could alter the probability that a sequence of coin flips comes out all heads just by beginning the sequence a little earlier or later, and doing nothing else differently. This already conflicts with the spirit of the principle of time-invariance.

But we need not press that point. The examples presented above render Weintraub's objections irrelevant. Regularity implies that a set of possible outcomes for a *single finite-time* trial will differ in probability from *disjoint* translations and rotations of that set of outcomes. In the case of vacuum fluctuations, for example, we considered two sets of times, exactly alike in structure and duration, where neither contains or overlaps the other, and yet the probability that a fluctuation occurs at a time in one set must differ from the probability that a fluctuation occurs at a time in the other set. We cannot explain this away by appeal to the subset relation or a difference in duration; the sets *only* differ in their times. We also considered sets of points in space, exactly alike in structure, where neither contains or overlaps the other, and even for a single instantaneous event (the very beginning of the first vacuum fluctuation to occur in either

set within a given time span), the corresponding probabilities must differ, if regularity holds. This runs against another principle: The laws of physics don't vary across space, either. Both examples contradict the relativistic principle of space-time invariance, unless probabilities simply are not determined by physical law.

4.3 Restricting invariance to finite events

Another move the regularist might make in response to Williamson is to suggest that the *local* laws governing individual finite-time processes are time invariant, as tradition would have it, but the laws (if any) governing the probabilities of *infinite sequences* of events are not. In fact, it is an assumption of Williamson's examples that the outcomes of the individual coin tosses are mutually independent and all have exactly the same probability, namely one half. So the regularist must hold that the probabilities of infinite sequences of independent events are not determined by the probabilities of the individual outcomes in the sequence. In particular, probabilities are partly determined by the principle of regularity itself, and under regularity, $P(H(s))$ *must* be smaller than $P(H(s'))$.

But this move too is rendered irrelevant by our new examples, for they do not involve infinite sequences of events. Rather they concern individual events in finite space-time. Granted, these events—the dart hitting a point in a given set, or a vacuum fluctuation occurring at a time in a given set, for example—are countably infinite unions of individual micro-events, such as the dart hitting a particular exact point. Given the fact that we are not insisting on countable additivity, one might suggest that the probabilities of these individual outcomes do not determine those of the unions. But to hold that the probabilities of these unions might vary across time and space despite being

bounded and exactly alike in structure flies in the face of space-time invariance as it is normally understood. After all, the events that the discipline of physics predicts and for which it gives probabilities are normally themselves infinite unions of micro-events. Quantum mechanics for example tells us about the probability that a particle measurement will read ‘spin-up’ under certain circumstances. But the event of a ‘spin-up’ outcome is a union of many micro-events in which the particle might have any of various positions, velocities, and quantum states, and of course the experimenter, other nearby objects, and the rest of the universe might also be in various states. So to conclude that infinitary unions of exact states are not determined by space-time invariant laws would be to throw out the very sort of space-time invariance that physics and philosophy of physics have long accepted. Of course, those disciplines might just have it wrong, but to imply this is an enormous burden for regularity arguments to bear.

4.4 Physicality and meaning

Yet another possible objection to Williamson is that his infinite sequences of coin flips are unrealistic or even physically impossible. One cannot flip the same coin under the same circumstances infinitely many times, and arguably, one cannot perform *any* experiment infinitely many times. Our examples have also circumvented this objection. They show that regular probabilities violate rigid transformation invariance even for a single experiment, conducted in a small region of space-time.

But one might raise a related objection against our new examples, to the effect that they too are unrealistic, or that any probabilities applied to them are meaningless, since darts do not strike exact points and vacuum fluctuations do not occur at exact times,

and even if they did, we would never be able to measure such things with sufficient accuracy to determine whether or not they fell within sets like our R and X above. After all, those sets are nowhere dense—riddled with gaps in such a way that membership in such a set depends critically on the *exact* outcome. And it may be that there are no exact continuous magnitudes at all in the real world, or no circumstance where we can determine whether a physical magnitude lies in a given nowhere dense set.

To this I will concede uncertainty. I do not know whether there are exact physical quantities whose ranges of possible values include nowhere dense sets like R and X . Nor am I certain that chances or credences of such events are meaningful. That would seem to depend on the details of what exactly chances and credences are taken to be.¹⁰ But regularists themselves have adopted infinitesimal and hyperreal probabilities precisely in order to save regularity in the face of such idealised and experience-transcending examples. If it were not for examples like the dart throw described in the introduction (which assumes a continuum of possible outcomes), infinite sequences of coin tosses, and lotteries with infinitely many equally likely outcomes, regularists would have no need for hyperreal probabilities. Indeed, much of the literature supporting regular or Euclidean probabilities concerns just such cases (e.g., exact dart throws: Bernstein and Wattenberg [1969]; infinite coin tosses: Lewis [1980], p. 270, Gwiazda [2008]; infinite lotteries: McCall and Armstrong [1989], Gwiazda [2010], Wenmackers and Horsten [2013], Benci, Horsten, and Wenmackers [2013]), and one of the chief benefits of Benci, Horsten, and Wenmackers' NAP theory, as discussed above, is to define probabilities even for highly abstract non-measurable sets in the continuum. So if we are to dismiss such examples as unrealistic or meaningless, then the benefits of hyperreal probabilities go with them.

Regularists might then avoid many problems, but not by expanding the range of probability values. Rather they would do so by dismissing the entire discipline of continuous probability theory as unphysical or meaningless, despite the important role that it plays in the treatment of many real-world problems.

5 Conclusion

Requiring probabilities to be regular would be very costly. It would mean that in some cases we could not assign the same probability to perfectly symmetrical sets of outcomes, even when these sets are simple, countable, and constructive. If the events corresponding to such sets have objective physical chances, then regularity implies that those chances are not determined by space-time invariant laws. So given regularity, either (i) the laws of physics are not space-time invariant, (ii) chances are not determined by the laws of physics, or (iii) such events have no chances at all. Any of these is a difficult consequence to accept.

For rational subjective credences, regularity creates another dilemma. The regularist claims that it is irrational to be so convinced of a contingent proposition that one is willing to bet everything for an arbitrarily small return (or none at all), or to be unswayed by any amount of statistical evidence (as zero credence and Bayesianism would imply). But the objective Bayesian claims that in certain situations it is irrational to assign credences that are not translation- or rotation-invariant. Both cannot be right. And even if we accept that objective Bayesianism is false, regularity implies more. It means that in some cases we *can't* assign the same probability to perfectly symmetric events, no matter how symmetric our knowledge of them may be.

To these remarks the devoted regularist might reply, “So be it. The arguments for regularity are ironclad, however inconvenient their consequences may be. We must simply accept them, and re-think everything else.” But given the consequences, the arguments for regularity should be scrutinised closely, and I suspect we will find chinks in their armour.

If indeed those arguments can be undone, or they are not strong enough to bear the burden of their awkward consequences, then this also tells us something about the new Euclidean theories of set size. It means that in one of the key applications claimed for them—non-standard probability measures—they are as limited and uninformative as they are for set size itself.

References

- Adams, E. [1962–1964] ‘On Rational Betting Systems’, *Archiv für mathematische Logik und Grundlagenforschung*, **6**, pp. 7–29 and 112–128.
- Appiah, A. [1985]: *Assertion and Conditionals*, New York: Cambridge University Press.
- Benci, V. and Di Nasso, M. [2003]: ‘Numerosities of Labeled Sets: A New Way of Counting’, *Advances in Mathematics*, **173**, pp. 50–67.
- Benci, V., Di Nasso, M., and Forti, M. [2007]: ‘An Euclidean Measure of Size for Mathematical Universes’, *Logique et Analyse*, **50**, pp. 43–62.
- Benci, V., Horsten, L., and Wenmackers, S. [2013]: ‘Non-Archimedean Probability’, *Milan Journal of Mathematics*, **81**, pp. 121–151.

- Bernstein, A. R. and Wattenberg, F. [1969]: ‘Non-standard Measure Theory’, in W. A. J. Luxemburg (ed), *Applications of Model Theory to Algebra, Analysis, and Probability*, New York: Holt, Rinehart, and Winston, pp. 171–185.
- Carnap, R. [1950]: *Logical Foundations of Probability*, Chicago: University of Chicago Press.
- . [1963]: ‘Replies and Systematic Expositions’, in Schilpp [1963], pp. 859–1013.
- De Finetti, B. [1964]: ‘Foresight: Its Logical Laws, its Subjective Sources’, in H. Kyburg and H. Smokler (eds), *Studies in Subjective Probability*, Huntington, NY: Krieger, pp. 93–158.
- Edwards, W., Lindman, H., and Savage, L. J. [1963]: ‘Bayesian Statistical Inference for Psychological Research’, *Psychological Review*, **70**, pp. 193–242.
- Gyenis, Z., and Rédei, M. [forthcoming]: ‘Defusing Bertrand’s Paradox’, *British Journal of the Philosophy of Science*.
- Gwiazda, J. [2008]: ‘The Probability of an Infinite Sequence of Heads’, PhilSci Archive, available at <http://philsci-archive.pitt.edu/eprint/4017>.
- . [2010]: ‘Probability, Hyperreals, Asymptotic Density, and God’s Lottery’, PhilSci Archive, available at <http://philsci-archive.pitt.edu/id/eprint/5527>.
- Hájek, A. [2003]: ‘What Conditional Probability Could Not Be’, *Synthese*, **137**, pp. 273–323.
- . [unpublished]: ‘Staying Regular?’, available at http://fitelson.org/few/hajek_paper.pdf. Retrieved 24 May 2012.
- Haverkamp, N., and Schulz, M. [2012]: ‘A Note on Comparative Probability’, *Erkenntnis*, **76**, pp. 395–402.

- Jackson, F. [1987]: *Conditionals*, Oxford: Blackwell.
- Jaynes, E.T. [1968]: ‘Prior Probabilities’, *IEEE Transactions on Systems Science and Cybernetics*, **4**, pp. 227–241.
- . [1973]: ‘The Well-Posed Problem’, *Foundations of Physics*, **3**, pp. 477–492.
- Jeffrey, R. [1992]: *Probability and the Art of Judgment*, Cambridge, UK: Cambridge University Press.
- Jeffreys, H. [1961]: *Theory of Probability*, 3rd Edition, Oxford: Clarendon Press.
- Joyce, J. M. [1998]: ‘A Nonpragmatic Vindication of Probabilism’, *Philosophy of Science*, **65**, pp. 575–603.
- . [2009]: ‘Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief’, in F. Huber and C. Schmidt-Petri (eds), *Degrees of Belief*, Houten: Springer Netherlands, pp. 263–297.
- Katz, F. M. [1981]: *Sets and Their Sizes*, Ph.D. Dissertation, MIT, available from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.7026>.
- Kemeny, J. G. [1955]: ‘Fair Bets and Inductive Probabilities’, *The Journal of Symbolic Logic*, **20**, pp. 263–273.
- . [1963]: ‘Carnap’s Theory of Probability and Induction’, in Schilpp [1963], pp. 711–738.
- Keynes, J.M. [2004]: *A Treatise on Probability*, New York: Dover Publications.
- Leitgeb, H., and Pettigrew, R. [2010a]: ‘An Objective Justification of Bayesianism I: Measuring Inaccuracy’, *Philosophy of Science*, **77**, pp. 201–235.
- . [2010b]: ‘An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy’, *Philosophy of Science*, **77**, pp. 236–272.

- Lewis, D. [1980]: ‘A Subjectivist’s Guide to Objective Chance’, in R.C. Jeffrey (ed), *Studies in Inductive Logic and Probability*, v. II, Berkeley and Los Angeles: University of California Press, pp. 263–293.
- Makinson, D. [2011]: ‘Conditional Probability in the Light of Qualitative Belief Change’, *Journal of Philosophical Logic*, **40**, pp. 121–153.
- Mancosu, P. [2009]: ‘Measuring the Size of Infinite Collections of Natural Numbers: Was Cantor’s Theory of Infinite Number Inevitable?’, *Review of Symbolic Logic*, **2**, pp. 612–646.
- Mayberry, J. [2000]: *The Foundations of Mathematics in the Theory of Sets*, Cambridge: Cambridge University Press.
- McCall, S., and Armstrong, D. M. [1989]: ‘God’s Lottery’, *Analysis*, **49**, pp. 223–224.
- Nelson, E. [1987]: *Radically Elementary Probability Theory*, Princeton: Princeton University Press.
- Parker, M. [2009]: ‘Philosophical Method and Galileo’s Paradox of Infinity’, in B. van Kerkhove (ed), *New Perspectives on Mathematical Practices*, Hackensack, NJ: World Scientific, pp. 76–113.
- . [forthcoming]: ‘Set Size and the Part–Whole Principle’, *Review of Symbolic Logic*.
- Poincaré, H. [1911]: ‘L’Evolution des Lois’, *Scientia*, **9**, pp. 275–92. Trans. in Poincaré [1963]: *Mathematics and Science: Last Essays*, pp. 1-13, New York: Dover Publications.
- Quine, W. [1951]: ‘Two Dogmas of Empiricism’, *The Philosophical Review*, **60**, pp. 20–43.

- Ramsey, F. P. [1931]: ‘Truth and Probability’, in *The Foundations of Mathematics and Other Logical Essays*, London and New York: Harcourt, Brace and Co., pp. 156–198.
- Schilpp, P. A. (ed) [1963]: *The Philosophy of Rudolf Carnap*, The Library of Living Philosophers Vol. XI, Chicago: Open Court.
- Shimony, A. [1955]: ‘Coherence and the Axioms of Confirmation’, *Journal Of Symbolic Logic*, **20**, pp. 1–28.
- Skyrms, B. [1980]: *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*, Yale University Press: New Haven and London.
- . [1995]: ‘Strict Coherence, Sigma Coherence and the Metaphysics of Quantity’, *Philosophical Studies*, **77**, pp. 39–55.
- Stalnaker, R. C. [1970]: ‘Probability and Conditionals’, *Philosophy of Science* 37: 64–80.
- Uffink, J. [1995]: ‘Can the Maximum Entropy Principle Be Explained as a Consistency Requirement?’, *Studies in the History and Philosophy of Modern Physics*, **26**, pp. 223–261.
- Weintraub, R. [2008]: ‘How Probable is an Infinite Sequence of Heads? A Reply to Williamson’, *Analysis*, **68**, pp. 247–50.
- Wenmackers, S., and Horsten, L. [2013]: ‘Fair Infinite Lotteries’, *Synthese*, **190**, pp. 37–61.
- Williamson, J. [1999]: ‘Countable Additivity and Subjective Probability’, *British Journal for the Philosophy of Science*, **50**, pp. 401–416.
- Williamson, T. [2007]: ‘How Probable is an Infinite Sequence of Heads?’, *Analysis*, **67**, pp. 173–180.

¹ There is yet another motivation for regularity: To eliminate the difficulty of conditionalizing on events of probability zero. However, this can also be achieved by taking conditional probability as primitive, and there are independent arguments for that approach (Hájek [2003], Makinson [2011]).

² I am deliberately avoiding the Greek letters and script types traditionally used for the components of a probability space, as they tend to make simple matters seem more technical and intimidating.

³ Hájek also gives one argument that could be directed against weak regularity: He claims that there is a difficulty in giving Kolmogorov-style axioms of probability that permit probability functions to have arbitrarily large nonstandard ranges: “Just try providing an axiomatization along the lines of Kolmogorov’s that has flexibility in the range built into it,” he writes. “If we don’t know exactly what the range is, “we don’t know what its notion of additivity will look like.” But we have easily met this challenge with our weak requirements on the range and good old finite additivity. Benci, Horsten, and Wenmackers [2013] have met it with a stricter set of axioms and a stronger additivity property analogous to countable additivity. So it is not clear what the difficulty is supposed to be.

⁴ The Lebesgue measurable sets are those point sets in Euclidean space over which the Lebesgue measure is defined. Lebesgue measure is the standard notion of total length, width, or n -dimensional volume in Euclidean space. A full definition can readily be found on the Internet or in any analysis textbook.

⁵ In fact, the converse holds as well, given finite additivity and the further postulate

(vi) For any x and y in F , if $x < 0$ then $y + x < y$.

For by finite additivity, $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$ for any $A \in D$. By (vi), this implies that $P(\emptyset) \geq 0$. But \emptyset is a proper subset of any non-empty set, so if P is Euclidean and $\emptyset \neq A \in D$, then $P(A) > P(\emptyset)$. Thus all non-empty sets in D have positive probability, i.e., P is weakly regular.

⁶ Some respondents have questioned whether TR really is a proper subset of R , but the answer is clearly yes. The metaphorical language of rotation might lead one to think these sets are in some sense the same set, just positioned in two different ways. But these are sets of *points*, positions in a space, and these cannot be literally moved. A rotation is just a function mapping some points to others, and the set R contains all of the points in TR as well as others. So unless we adopt some radically unorthodox sort of set theory, TR is indeed a proper subset of R .

⁷ More formally, let T be a translation $Tx = x + c$. Suppose $A, TA \subseteq [0, 1)$ and $P(A) \neq P(TA)$. Choose $n \in \mathbb{N}$ so that $1/n \leq c$. For each whole number $i < n$, let $A_i = A \cap [i/n, (i + 1)/n)$. Then A_i and TA_i are disjoint, and by finite additivity, $\sum_{i \in \{0, 1, \dots, n-1\}} P(A_i) = P(A) \neq P(TA) = \sum_{i \in \{0, 1, \dots, n-1\}} P(TA_i)$. So for at least one i , $P(A_i) \neq P(TA_i)$. Hence some translation of a set B , disjoint from B , has a different probability from B .

⁸ Suppose T is a reflection about a fixed point c . Let $A_1 = A \cap [0, c)$ and $A_2 = A \cap [c, 1)$. Then $A_1 \cap TA_1 = A_2 \cap TA_2 = \emptyset$, and by additivity, $P(A_1) + P(\{c\}) + P(A_2) = P(A) \neq P(TA) = P(TA_1) + P(T\{c\}) + P(TA_2)$. Since $T(c) = c$, either $P(A_1) \neq P(TA_1)$ or $P(A_2) \neq P(TA_2)$. So in either case, we have disjoint reflection images with different probabilities.

⁹ Haverkamp and Schulz [2012] offer a related critique of Weintraub's argument. They argue that separate runs of the same physical device or set-up ought to produce outcomes with the same probabilities regardless of when it is set running. However, there are difficulties with the idea of having the same physical device implement an infinite process in two entirely separate instances. Each run could be executed in finite time if the coin flips occur at shorter and shorter intervals, as Haverkamp and Schulz suggest, but then we lose the strong parallel between a sequence of flips and a proper subsequence. The objections to Weintraub made here are more general and avoid these problems.

¹⁰ Thanks to ----- for pressing this point with regard to the notion of chance.