POPPER FUNCTIONS, UNIFORM DISTRIBUTIONS AND INFINITE SEQUENCES OF HEADS

ALEXANDER R. PRUSS

Abstract. Popper functions allow one to take conditional probabilities as primitive instead of deriving them from unconditional probabilities via the ratio formula $P(A|B) = P(A \cap B)/P(B)$. A major advantage of this approach is it allows one to condition on events of zero probability. I will show that under plausible symmetry conditions, Popper functions often fail to do what they were supposed to do. For instance, suppose we want to define the Popper function for an isometrically invariant case in two dimensions and hence require the Popper function to be rotationally invariant and defined on pairs of sets from some algebra that contains at least all countable subsets. Then it turns out that the Popper function trivializes for all finite sets: $P(A|B) = 1$ for all A (including $A = \varnothing$) if B is finite. Likewise, Popper functions invariant under all sequence reflections can't be defined in a way that models a bidirectionally infinite sequence of independent coin tosses.

1. A problem for Popper functions

Classical probability theory defines conditional probabilities in terms of unconditional probabilities via the ratio formula: $P(A|B) = P(A \cap B)/P(B)$. There are at least three kinds of motivations not to do this (cf. [6, 7]). One is metaphysical: it may seem plausible to think that conditional probabilities are actually more fundamental than the unconditional ones. And two are more technical. One is that intuitively the conditional probability $P(A|B)$ can make sense even when $P(B)$ is undefined, for instance because B is a nonmeasurable set or B represents an event where it doesn't seem to make sense to assign a precise numerical probability, such as maybe the event of the cosmos exhibiting lawlike regularities. The second is that the conditional probability $P(A|B)$ often makes sense even though $P(B) = 0$, while the ratio formula then yields $0/0$. In both of these technical cases, we want to be able to say things like $P(\varnothing|B) = 0$ and $P(B|B) = 1$, as well as to be able to say or deny[1] things like $P(\{1/2\}|\{1/3, 1/2\}) = 1/2$ in the case of a uniform distribution on the interval $[0, 1] = \{x : 0 \leq x \leq 1\}$.

Popper functions are primitive conditional probabilities designed to address all three problems. I shall give some negative formal results showing that Popper functions fail to satisfy natural symmetry or invariance

[1]The arguments in [13] suggest that sometimes denial is the right move.

conditions. The mathematics here mostly comes from the paradoxical decomposition results behind the Hausdorff and Banach-Tarski paradoxes, but unlike the Hausdorff and Banach-Tarski theorems, our results do not need the Axiom of Choice.[2] For simplicity of formalism, we will work with probabilities and Popper functions in the context of probabilities for sets rather than sentences. A Popper function P can be defined as a function that assigns a real number $P(A|B)$ to every pair A and B of members of a field $\mathcal{F}$ of subsets of some set Ω (fields are non-empty, and closed under complements and finite unions), satisfying the following axioms, where $A^c = \Omega - A = \{y \in \Omega : y \notin A\}$:

(1) $0 \leq P(A|B) \leq P(B|B) = 1$
(2) if $P(B^c|B) \neq 1$, then $P(-|B)$ is a finitely additive probability function
(3) $P(A \cap B|C) = P(A|C)P(B|A \cap C)$
(4) if $P(A|B) = P(B|A) = 1$, then $P(C|A) = P(C|B)$. [19]

Following [19], we say that B is abnormal provided that $P(B^c|B) = 1$. The crucial things to observe here is that (a) when B is normal, $P(-|B)$ is a finitely additive probability and by (3) we have the product formula (b) $P(E|F)P(F|G) = P(E|G)$ whenever $E \subseteq F \subseteq G$ (just let $B = E$, $A = F$ and $C = G$). As [19] notes, beyond the condition that $P(-|B)$ be a finitely additive probability for normal B and the product formula, the rest is just detail for handling abnormal sets. We could reaxiomatize the above by distinguishing a non-empty set $\mathcal{G} \subseteq \mathcal{F}$ of normal subsets, with the condition that $P(A|B)$ is defined for all $A \in \mathcal{F}$ but only for $B \in \mathcal{G}$, and then simply requiring the finitely-additive probability and product conditions, as well as specifying that if $P(A|B) > 0$ for a normal B, then A is normal as well. [19, p. 420].

The following are some standard facts about abnormal sets.

Proposition 1. *The following conditions on a set B in $\mathcal{F}$ are equivalent given a Popper function P:*

(i) $P(B^c|B) = 1$
(ii) $P(\varnothing|B) = 1$
(iii) $P(A|B) = 1$ *for all* $A \in \mathcal{F}$.

Moreover, if B and C are in $\mathcal{F}$, with B abnormal and $C \subseteq B$, then C is abnormal. And if B and C in $\mathcal{F}$ are both abnormal, so is $B \cup C$.

The straightforward proof is given in Section 2 for completeness.

[2] That said, the methods of [16, Section 4] can be used to show that if there is a Popper function on some set of cardinality equal to the continuum for which all countable subsets are normal (in a sense that will be defined shortly), then that yields a weak version of the Axiom of Choice that is nonetheless strong enough to generate the Hausdorff and Banach-Tarski paradoxes. The proof in that paper is given for the case of a Popper function where all subsets are normal, but works just as well if only the countable ones are assumed to be normal.

We see that if B is abnormal, $P(-|B)$ is a trivial constant function. Popper functions only solve our two technical problems about conditional probability if the sets we want to be able to condition on are normal. For instance, a Popper function where all finite sets are abnormal offers no help with conditioning on finite sets.

Now, nonmeasurability on the real line $\mathbb{R}$ arises for Lebesgue measure from the requirement that the measure be invariant under translations: $m(A) = m(\tau A)$ for any translation τ. Similarly, abnormalcy will be seen to arise for Popper functions in dimensions two and higher from similar symmetry conditions. Suppose that Ω is a subset of some space X and that G is a group of bijective transformations of X onto X. In the cases we will consider, X will be $\mathbb{R}^n$ and G will be a group of isometries. We then say that a Popper function on $\mathcal{F}$, where $\mathcal{F}$ is a field of subsets of Ω, is weakly G-invariant provided that

(5) if $g \in G$ and $A \in \mathcal{F}$ is such that $gA \subseteq \Omega$, then $gA \in \mathcal{F}$
(6) if $g \in G$ and A, B, gA, gB are in $\mathcal{F}$, then $P(gA|gB) = P(A|B)$,

where $gA = \{ga : a \in A\}$. Observe that then if B is abnormal and $gB \in \mathcal{F}$, we have $1 = P(\varnothing|B) = P(g\varnothing|gB) = P(\varnothing|gB)$ and so gB is abnormal as well. Note too that our invariance condition does not require that $g\Omega \subseteq \Omega$. This allows us, for instance, to talk about translation and rotation invariance for subsets of, say, the cube.

For strong invariance, we add the condition:

(7) if $g \in G$ and A, B, gA are in $\mathcal{F}$ with $A \cup gA \subseteq B$, then $P(gA|B) = P(A|B)$.

It is shown in [17, Theorem 1] that weak invariance does not imply strong invariance, even if Ω is all of X and there are no abnormal sets in $\mathcal{F}$.[3]

An isometry in a metric space is a transformation ρ that preserves distances: $d(\rho x, \rho y) = d(x, y)$. In n-dimensional Euclidean space $\mathbb{R}^n$, isometries are combinations of translations, rotations and reflections. Isometric invariance is a very natural condition to put on probabilities for outcomes of certain random processes taking values in a subset of $\mathbb{R}^n$. But it turns out that isometrically invariant Popper functions in dimensions 2 and higher have *many* abnormal sets, so many that too many of the technical problems that Popper functions were supposed to solve cannot be solved by them.

Winfried Just [9] has shown, by an elaborate explicit construction without the use of the Axiom of Choice, that there is a bounded (i.e., contained

[3]In [1, p. 5] there is a claimed proof that weak invariance is equivalent to strong invariance in this case, but the putative proof is incomplete. Note also that while the Axiom of Choice is assumed at the outset of [17], it is not needed for the proof that not-(b) implies not-(a) in Theorem 1 of that paper, and that implication applied in the case where $G = \Omega = \mathbb{Z}$ (acting on itself by addition) will generate a weakly but not strongly invariant Popper function on an algebra $\mathcal{F}$ of subsets of $\mathbb{Z}$ with all members of $\mathcal{F}$ normal, contrary to [1].

within a finite disc) countably infinite subset J of $\mathbb{R}^2$ such that J can be decomposed into four pieces which can, in turn, be rigidly moved to construct two copies of J. More precisely, J can be partitioned into disjoint subsets J_1, J_2, J_3, J_4, and there are rigid motions g_1, g_2, g_3, g_4 such that $g_1 J_1 = J$ and $g_2 J_2, g_3 J_2, g_4 J_4$ is also a disjoint partition of J. (Rigid motions are combinations of translations and rotations, without reflection.) By scaling and translation, we may assume that $J \subseteq [0,1]^2$. Just's result improves on the Sierpiński-Mazurkiewicz paradox [20, pp. 9–10] which gives a simpler paradoxical decomposition in the same spirit, but of an unbounded set. A simple exposition of the Sierpiński-Mazurkiewicz paradox is given in [18].

It follows that any *strongly* isometrically invariant Popper function P on a field $\mathcal{F}$ of subsets of $[0,1]^2$ such that $\mathcal{F}$ contains all countable subsets is such that P makes J abnormal. For if J were normal, then $P(-|J)$ would define a G-invariant finitely-additive probability function on all subsets of J (i.e., $P(gA|J) = P(A|J)$ if $A, gA \subseteq J$). But that's impossible. For we would then have:

$$\begin{aligned}
1 &= P(J|J) \\
&= P(J_1|J) + P(J_2|J) + P(J_3|J) + P(J_4|J) \\
&= P(g_1 J_1|J) + P(g_2 J_2|J) + P(g_3 J_3|J) + P(g_4 J_4|J) \\
&= P(J|J) + P(g_2 J_2 \cup g_3 J_3 \cup g_4 J_4|J) \\
&= P(J|J) + P(J|J) = 2.
\end{aligned}$$

Moreover, since any subset of an abnormal set is abnormal, there will be an abnormal singleton, and by G-invariance, every singleton will be abnormal. And since finite unions of abnormal sets are abnormal, it follows that every finite subset of $[0,1]^2$ will be abnormal.

However, strong isometric invariance is somewhat less intuitive than weak isometric invariance, so the above result is perhaps not so damaging to the Popper function project. But it turns out that with some more work, a similar result can be proved for weak isometric invariance in dimensions two and higher.

Theorem 1. *Let Ω be any subset of $\mathbb{R}^n$, $n \geq 2$, that contains a solid disc or ball of non-zero radius. Suppose P is an isometrically invariant Popper function on Ω such that $\mathcal{F}$ contains all countable subsets. Then there is a countably infinite abnormal $J \subset \Omega$, and hence every finite subset of Ω is abnormal.*

The claim about finite subsets follows from the existence of the countably infinite abnormal J as before. Moreover, we only need to prove the theorem for $n = 2$, since if $n > 2$, we can just apply the $n = 2$ case to the Popper function restricted to pairs of subsets of a plane passing through the center of the solid ball in Ω. Note that the assumption that all countable subsets are in $\mathcal{F}$ parallels the fact that ordinary measures like the Lebesgue or Hausdorff measures are defined for all Borel sets, and all countable sets are Borel sets.

The proof will be given in Section 2. It is not known if isometries can be replaced with rigid motions in Theorem 1.[4]

But in dimensions 3 and higher, it turns out that all we need to generate abnormality is invariance under rotations about the origin. Let $B_r = \{z : |z| < r\}$ be the interior of the ball of radius r centered on the origin and let $S_r = \{z : |z| = r\}$ be its surface.

Theorem 2. *Let $\Omega \subseteq \mathbb{R}^n$, $n \geq 3$, contain S_r for some $r > 0$. Suppose P is a Popper function on Ω invariant under rotations about the origin and such that $\mathcal{F}$ contains all countable subsets. Then there is a countably infinite abnormal $J_r \subseteq S_r$. Hence if $\Omega = B_R$ for some $R > 0$, then every finite subset of Ω not containing the origin is abnormal.*

Again, the last sentence follows trivially from what came before. For if $J_r \subseteq S_r$ is abnormal for an r strictly between 0 and R, so will some singleton in S_r be, and hence so will every singleton in S_r be (they are all rotationally equivalent), and hence so will be every finite subset of B_R not containing the origin. The proof is also in Section 2 and uses one of the constructions involved in the Banach-Tarski paradox.

Note that Theorem 2 is false for $\mathbb{R}^2$, at least given the Axiom of Choice, since [12] shows that then there is a rotationally-invariant Popper function on all subsets of any circle (the unit interval with translation modulo 1 that they work with is equivalent to the circle) with only the empty set abnormal (in our terminology).

The situations in the theorems include paradigmatic cases where we would like to have well-defined Popper functions. Imagine, for instance, that we have no information about where a dart with a perfectly defined tip hit a circular planar target. It is reasonable to take the distribution of the dart location over the target to be uniform and the corresponding Popper function P to be weakly invariant with respect to isometries. Let x and y be any two points on the target. We will want to be able to say that given that one hit either x or y, the conditional probability of having x is $1/2$, or at least less than one. But each finite subset will be abnormal by Theorem 1, so absurdly $P(\{x\}|\{x, y\}) = P(\{y\}|\{x, y\}) = P(\varnothing|\{x, y\}) = 1$.

Or consider a particle released at at the center of a sphere, and engaging in Brownian motion until it hits the surface of the sphere. Let X be the point where the particle impacts the surface of the sphere. We expect the probabilities, conditional and unconditional, concerning X to be symmetric

[4] The difference may seem small, namely whether one also requires invariance under reflections—and, in fact, under any single fixed reflection, since all reflections in $\mathbb{R}^2$ can be generated by a single fixed reflection combined with appropriate rigid motions. But nonetheless the difference may be significant. The proof of Theorem 1 uses a trick similar to that used in [15] to show that there is no isometrically-invariant preordering on the subsets of the circle that extends strict inclusion. But [15] also showed that there is such a preordering when isometries are replaced with rigid motions. So we know that sometimes there is a significant difference between isometries and rigid motions in respect of the generation of invariant functions or relations.

under rotations about the center. We want to be able to say things like: given that the particle impacts either on the North or the South Pole, the probability that it impacts on the North Pole is 1/2. But we can't say this, given that all countable, and hence all finite, subsets of the sphere are abnormal by Theorem 2, and hence all probabilities conditional on them are 1.

In [6] it was argued that the way to solve the problems with the ratio formula in the case of null probability events such as a dart's hitting a particular point is to move to primitive conditional probabilities of the Popper variety. This requires singletons and hence all finite subsets to be normal in our terminology, and thus in situations exhibiting appropriate isometric symmetries it will not work.

More recently, [5] has convincingly argued that assigning hyperreal infinitesimal probabilities to null probability events posits too fine-grained a probabilistic structure—after all, we would need to specify the particular hyperreal system to be used. Instead, [5] seems to suggest that Popper functions have a better level of granularity. Here, one can think of classical probability assignments as having the lowest level of granularity: they lump all measure zero sets, including all countable subsets of a Euclidean space in isometrically invariant cases, together. Popper functions allow for a finer level of granularity. For instance, when two subsets A and B have measure zero, as long as both sets are normal, we can numerically compare their sizes, for instance by comparing $P(A|A \cup B)$ with $P(B|A \cup B)$ as in [2] or even more finely by comparing $P(A - B|(A - B) \cup (B - A))$ with $P(A - B|(A - B) \cup (B - A))$ as in [4]. Hyperreal probability assignments, on the other hand, allow for even finer-grained comparisons, and it is these comparisons that [5] has argued to posit excessive probabilistic structure. But our results show that Popper functions already posit too much structure in two- and three-dimensional isometrically invariant cases, since the extra structure they posit in the case of countable subsets must fail to have the requisite invariance (just as the hyperreal probabilities have too much structure already in one-dimensional isometrically invariant cases [14]).

An optimistic thought would be that much as we have learned to think of some sets as non-measurable, we should also accept some sets as "non-conditioning"—sets like Just's paradoxical set J. But if we want to be able to condition on finite sets, these non-conditioning sets cannot simply be Popper-style abnormal sets, since every subset of an abnormal set is abnormal, while we want finite subsets of non-conditioning sets to be something we can condition on.

Perhaps we should thus modify the definition of Popper functions so that instead of $P(A|B)$ being defined for all A and B in a single field $\mathcal{F}$, there a field $\mathcal{F}$ and a set or field $\mathcal{G}$ such that $P(A|B)$ is defined whenever $A \in \mathcal{F}$ and $B \in \mathcal{G}$. But it is not clear, however, what requirements should be put on $\mathcal{G}$ that will not be *ad hoc*. This is left for further investigation.

We should note that the above problems come from requiring *rotational* invariance. For we have:

Theorem 3. *Given the Axiom of Choice, there is a strongly translation- and coordinate-reflection-invariant Popper function defined on all pairs of subsets of $\mathbb{R}^n$, $n \geq 1$, with every non-empty subset normal.*

Here, a coordinate-reflection is a function $f : \mathbb{R}^n \to \mathbb{R}^n$ of the form $f(x_1, \ldots, x_n) = (\alpha_1 x_1, \ldots, \alpha_n x_n)$ for a sequence of $\alpha_i \in \{-1, 1\}$. The proof will be a simple application of a result of [1] and will given in Section 2. Note that by Theorems 1 and 2, we cannot require invariance under all reflections in Theorem 3, since compositions of reflections can be used to generate rotations.

Given that the problem comes from rotational invariance, instead of taking away the lesson that Popper functions are a flawed tool, one might conclude that space cannot really be rotationally isotropic or that there cannot be genuinely rotationally isotropic random processes (cf. [8]). The naturalness of Brownian motion makes this an unattractive way out.

Moreover, we have one final problem for Popper functions that does not depend on rotation in higher dimensions but merely one-dimensional reflection. This problem is a variant on Williamson's elegant argument [21] that the probability of an infinite sequence of heads cannot be given by an infinitesimal: Let H_n be the event that the independent fair coin tosses numbered $n, n+1, n+2, \ldots$ all result in heads, and let X_n be the result of the nth toss. Then by independence $P(H_1) = P(X_1 = H)P(H_2) = (1/2)P(H_2)$. But $P(H_1) = P(H_2)$, since both are the probability of an equivalent[5] infinite sequence of fair coin tosses and so $P(H_1) = (1/2)P(H_1)$, which implies $P(H_1) = 0$.

Given the correspondence between Popper functions and hyperreal probabilities [10, 11], we have reason to expect a corresponding negative result for Popper functions. Here is one such result, though for technical reasons in the context of a bidirectionally infinite sequence. A bidirectionally infinite sequence of coin tosses can be represented as a function from the integers $\mathbb{Z}$ to $\{H, T\}$. Thus, let $\Omega = \{H, T\}^{\mathbb{Z}}$ be the set of such functions. Let H_n be the event of getting heads on tosses $n, n+1, n+2, \ldots$, i.e.,

$$H_n = \{\omega \in \Omega : \forall k (k \geq n \to \omega(k) = H)\}.$$

Let X_n be the random variable giving result of the nth toss, i.e., $X_n(\omega) = \omega(n)$. Let G be the group of transformations of Ω generated by (i.e., writeable as a finite product of) sequence reflections. A sequence reflection is a transformation R_x, for $x \in \mathbb{Z}/2$ (the integers and half-integers), such that $R_x(\omega)(n) = \omega(2x - n)$. Thus, $R_x(\omega)(n)$ reflects the sequence about the point

[5]Though see [8] for an interesting critique of this equivalence condition, and in general of symmetry-based reasoning. Nonetheless, abandoning symmetry in this way appears a very high cost.

x. Observe that every sequence translation can be written as the product of two sequence reflections.

A natural symmetry condition, then, on our Popper function on Ω is weak invariance under sequence reflections or, equivalently, under G (and hence automatically also under translations). For instance, the probability of getting heads on tosses $n, n+1, n+2, \ldots$ conditionally on some event E is the same as the probability of getting heads on tosses $n, n-1, n-2, \ldots$ conditionally on $R_n E$. A second natural condition is that $P(H_1|H_2) = P(X_1 = H|H_2) = 1/2$ since the first toss is independent of H_2.

Theorem 4. *If G is the transformations of $\Omega = \{H,T\}^{\mathbb{Z}}$ generated by reflections, then there is no G-invariant field $\mathcal{F}$ of subsets of Ω containing all the events H_n, $n \in \mathbb{Z}$, and weakly G-invariant Popper function P on $\mathcal{F}$ such that $P(H_1|H_2) = 1/2$.*

I do not know if there is a version of this result for unidirectionally infinite sequences.

2. Proofs

Proof of Proposition 1. Trivially, (iii) implies (ii). If (ii) is true then $P(-|B)$ is not a probability function, and so by (2) we must have (i). So it remains to see that (i) implies (iii). First, observe that $1 = P(\varnothing|\varnothing) = P(A \cap A^c|\varnothing) = P(A|\varnothing)P(A^c|A \cap \varnothing)$ by (1) and (3). But the only way two numbers between 0 and 1 have a product equal to 1 is if both are 1, so in particular $P(A|\varnothing) = 1$ for all A. Supposing $P(B^c|B) = 1$, by (3) we have

$$P(\varnothing|B) = P(B^c \cap B|B) = P(B^c|B)P(B|B^c \cap B) = P(B|\varnothing) = 1.$$

Thus, by (4), $P(A|B) = P(A|\varnothing) = 1$ for all A.

Next suppose B is abnormal and for *reductio* suppose $C \in \mathcal{F}$ is a subset of B and is normal. Then $P(B|C) \geq P(C|C) = 1$ by (1) and (2). But by (iii) we have $P(C|B) = 1$, and so by (4) we have $P(\varnothing|C) = P(\varnothing|B) = 1$, which contradicts the assumption of normalcy.

Finally, suppose B and C are abnormal. Then by (3) we have

$$P(B \cap C|B \cup C) = P(B|B \cup C)P(C|B \cap (B \cup C)) = P(B|B \cup C)P(C|B).$$

But $P(C|B) = 1$ by abnormalcy. Thus, $P(B \cap C|B \cup C) = P(B|B \cup C)$. For a *reductio*, suppose $B \cup C$ is normal. Then by (2) we have

$$\begin{aligned} P(B|B \cup C) &= P((B \cap C) \cup (B - C)|B \cup C) \\ &= P(B \cap C|B \cup C) + P(B - C|B \cup C) \end{aligned}$$

and so $P(B - C|B \cup C) = 0$. Now,

$$\begin{aligned} 1 &= P(B \cup C|B \cup C) = P(C \cup (B - C)|B \cup C) \\ &= P(C|B \cup C) + P(B - C|B \cup C) = P(C|B \cup C). \end{aligned}$$

But of course $P(B \cup C|C) = 1$ by abnormalcy of C. It follows from (4) that $B \cup C$ is abnormal, which contradicts the assumption that it's normal. $\square$

Both Theorems 1 and 2 immediately generalize to higher dimensions as soon as they are proved for some dimension n. That is because a Popper function on a higher-dimensional space induces Popper functions on its subsets and hence on lower-dimensional spaces. Thus, we only need to prove the theorems in the cases where $n = 2$ and $n = 3$, respectively.

Now, for a Popper function P and $A, B \in \mathcal{F}$, define $c_P(A, B) = P(A|A \cup B)/P(B|A \cup B)$, where $x/0 = \infty$ for $x > 0$. Note that $c_P(A, B)$ is well-defined, since if $A \cup B$ is abnormal, both its numerator and denominator are 1, and if $A \cup B$ is normal, then by finite additivity the numerator and denominator cannot both be zero. Also, $c_P(B, B) = 1$ for all B. One can think of c_P as a kind of relative probability (cf. [3]) or local exchange rate [1, 2]. The following summarizes some easy facts and will be useful for the proof of Theorem 1.

Lemma 1. *Let P be a Popper function. Then:*

(i) *If B is normal, then $c_P(-, B)$ is finitely additive.*
(ii) *For all $A, B, C \in \mathcal{F}$ we have $c_P(A, B)c_P(B, C) = c_P(A, C)$ if the left-hand-side is well defined (i.e., is neither $0 \cdot \infty$ nor $\infty \cdot 0$).*
(iii) *If P is weakly G-invariant, then for any $g \in G$ and $A, B \in \mathcal{F}$ such that $gA, gB \in \mathcal{F}$, we have $c_P(A, B) = c_P(gA, gB)$.*

Proof. Claim (i) is obvious.

Now let $T = A \cup B \cup C$. Suppose first that $P(A \cup B|T) > 0$ and $P(B \cup C|T) > 0$. Then if $X \subseteq A \cup B$, we have

$$\begin{aligned} P(X|T) &= P((A \cup B) \cap X|T) = P(A \cup B|T)P(X|(A \cup B) \cap T) \\ &= P(A \cup B|T)P(X|A \cup B) \end{aligned}$$

by (3), and so $P(X|A \cup B) = P(X|T)/P(A \cup B|T)$. Similarly, $P(Y|B \cup C) = P(Y|T)/P(B \cup C|T)$ for any $Y \subseteq B \cup C$. Thus if $c_P(A, B)c_P(B, C)$ is well-defined:

$$\begin{aligned} c_P(A, B)c_P(B, C) &= \frac{P(A|T)/P(A \cup B|T)}{P(B|T)/P(A \cup B|T)} \cdot \frac{P(B|T)/P(B \cup C|T)}{P(C|T)/P(B \cup C|T)} \\ &= \frac{P(A|T)}{P(C|T)}. \end{aligned}$$

If $P(A \cup C|T) > 0$, then $P(X|A \cup C) = P(X|T)/P(A \cup C|T)$ for $X \subseteq A \cup C$ as before, and the right hand side of the displayed equation equals $c_P(A, C)$ as desired.

If $P(A \cup C|T) = 0$, then T is normal and so $P(A|T) = P(C|T) = 0$ while $P(B|T) = 1$ as $T = A \cup B \cup C$. Thus, $c_P(A, B) = P(A|T)/P(B|T) = 0$ and $c_P(B, C) = P(B|T)/P(C|T) = \infty$ and so $c_P(A, B)c_P(B, C)$ is not well-defined, and so we trivially have (ii).

Next suppose $P(A \cup B|T) = 0$. Again, it follows that T is normal so $P(A|T) = P(B|T) = 0$ and $P(C|T) = 1$. Thus, $P(B \cup C|T) > 0$ and so $c_P(B, C) = P(B|T)/P(C|T)$ as before and thus $c_P(B, C) = 0$. In the same way $P(A \cup C|T) > 0$ and so $c_P(A, C) = 0$. Thus, $c_P(A, B)c_P(B, C) = c_P(A, C)$ if the left hand side is well-defined.

Finally, suppose $P(B \cup C|T) = 0$. Then $P(B|T) = P(C|T) = 0$ and $P(A|T) = 1$. We thus have $P(A \cup B|T) > 0$ and as before $c_P(A, B) = P(A|T)/P(B|T) = \infty$. Likewise $P(A \cup C|T) > 0$ and so $c_P(A, C) = P(A|T)/P(C|T) = \infty$. Thus, again $c_P(A, B)c_P(B, C) = c_P(A, C)$ if the left hand side is well-defined, and the proof of (ii) is complete.

Finally if P is weakly G-invariant, g is in G and A and B are such that A, B, gA, gB are in $\mathcal{F}$, then $c_P(gA, gB) = P(gA|gA \cup gB)/P(gB|gA \cup gB) = P(A|A \cup B)/P(B|A \cup B) = c_P(A, B)$, and hence (iii) holds. □

Let e be the identity transformation. Suppose $\Omega \subseteq \mathbb{R}^2$.

Lemma 2. *Suppose P is a weakly G-invariant Popper function on all subsets of Ω. Suppose that $g \in G$ is such that $g^2 = e$. Suppose A, B, gA are subsets of Ω. Then $c_P(A|B) = c_P(gA|B)$.*

Proof. We have $c_P(gA|A) = c_P(g^2A|gA) = c(A|gA)$ by Lemma 1(iii). Thus, $c_P(gA|A)c_P(A|gA) = (c_P(gA|A))^2$. Thus, the product on the left-hand-side is defined, and so hence must equal $c_P(gA|gA) = 1$ by Lemma 1(ii). Thus, $c_P(gA|A) = 1$. But then $c_P(gA|A)c_P(A|B) = c_P(gA|B)$. Thus, $c_P(A|B) = c_P(gA|B)$ by a final application of Lemma 1(ii). □

Now observe that every isometry on $\mathbb{R}^2$ can be written as a composition of reflections. For, every isometry is a composition of reflections, translations and rotations. And every translation or rotation is a composition of reflections: a translation by a vector v is the result of first reflecting in any line L at right angles to v and then reflecting in the line $v + L$, while a rotation $\rho_{x,\theta}$ by an angle θ about a point x is the result of first reflecting about any line L through x and then reflecting in $\rho_{x,\theta/2}L$.

Given a sequence of isometries $g_1, g_2, \ldots, g_n$ of $\mathbb{R}^n$, let $R(r; g_1, g_2, \ldots, g_n)$ be the infimum of those values of R such that every point of $\bar{B}_r \cup g_1\bar{B}_r \cup g_2g_1\bar{B}_1 \cup \cdots \cup (g_n \cdots g_2g_1\bar{B}_r)$ is contained in $\bar{B}_R$, where $\bar{B}_r = \{z : |z| \leq r\}$. Since isometries map bounded sets to bounded sets, $R(r; g_1, g_2, \ldots, g_n) < \infty$ if $r < \infty$.

By the construction of [9], let J be a countably infinite subset of $\bar{B}_1$ such that J can be partitioned into disjoint subsets J_1, J_2, J_3, J_4, and there are rigid motions g_1, g_2, g_3, g_4 such that $g_1J_1 = J$ and g_2J_2, g_3J_2, g_4J_4 is also a disjoint partition of J. We can now write $g_i = g_{i,n_i} \ldots g_{i,1}$ where the $g_{i,j}$ are self-inverse. Let $R = \max_{1 \leq i \leq 4} R(1; g_{i,1}, \ldots, g_{i,n_i})$. Let G be all isometries of $\mathbb{R}^2$.

Lemma 3. *Let P be any weakly G-invariant Popper function on all subsets of $\bar{B}_R$. Then J is abnormal.*

Proof of Lemma 3. By repeated applications of Lemma 2 (we can put this explicitly in terms of induction) and choice of R, we have $c_P(g_iJ_i, J) = c_P(g_{i,n}g_{i,n-1} \cdots g_{i,1}J_i, J) = c_P(g_{i,n-1} \cdots g_{i,1}J_i, J) = \cdots = c_P(J_i, J)$.

If J is normal, then by Lemma 1(i):

$$\begin{aligned}1 &= c_P(J|J) = c_P(J_1|J) + c_P(J_2|J) + c_P(J_3|J) + c_P(J_4|J)\\ &= c_P(g_1J_1|J) + c_P(g_2J_2|J) + c_P(g_3J_3|J) + c_P(g_4J_4|J)\\ &= c_P(J|J) + c_P(g_2J_2 \cup g_3J_3 \cup g_4J_4|J) = 2c_P(J|J),\end{aligned}$$

which is absurd. Thus, J is abnormal. □

Proof of Theorem 1. We only need to prove the existence of the countably infinite abnormal J, and only for $n = 2$. Suppose that Ω contains some ball of non-zero radius. By translation and rescaling, we may assume that Ω contains $\bar{B}_R$, and the existence of the desired countable abnormal set J now follows from Lemma 3. □

Now, it is time to prove Theorem 2. Famously, the Banach-Tarski Theorem states that, given the Axiom of Choice, a ball can be cut up into a finite number of pieces which can be reassembled to form two balls of the same size. Our proof will use completely standard methods employed in proving to Banach-Tarski types of paradoxicality results [20], but since we are not proving the full Banach-Tarski theorem, we can avoid the Axiom of Choice.

Proof of Theorem 2. Again, we only need to prove the existence of J_r and only in the case $n = 3$. Rescaling, we may assume $r = 1$.

A standard route to proving the Banach-Tarski Theorem is first to show that there is a countably infinite group H of rotations about the origin which has a paradoxical decomposition into four disjoint pieces H_i, $1 \le i \le 4$, such that H is the union of the pieces and there are members σ and τ in H such that $H_1 \cup \sigma H_2 = H$ and $H_3 \cup \tau H_4 = H$. While the full proof of the Banach-Tarski paradox uses the Axiom of Choice, the proof of the existence of H follows immediately from [20, Thms. 2.1 and 4.5], which do not use Choice.

It immediately follows that there is no finitely additive probability function ν on all subsets of H that is left-invariant under H, i.e., such that $\nu(hA) = \nu(A)$ for $h \in H$ and $A \subseteq H$. For then we would have $1 = \nu(H) \le \nu(H_1) + \nu(\sigma H_2) = \nu(H_1) + \nu(H_2)$ and $1 \le \nu(H_3) + \nu(\tau H_4) = \nu(H_3) + \nu(H_4)$. But then we would have $\nu(H) = \nu(H_1 \cup H_2 \cup H_3 \cup H_4) = \nu(H_1) + \nu(H_2) + \nu(H_3) + \nu(H_4) \ge 1 + 1$, which is impossible.

Now, every member of H is a rotation about the origin. Thus every member of H other than the identity e fixes some axis through the origin but moves all other points around. There are uncountably many lines through the origin in $\mathbb{R}^3$, and only countably many of them are axes of non-identity rotations in H. Let z_0 be a point on the unit sphere S_1 that does not lie on any of the axes of non-identity rotations in H. Then $hz_0 \ne z_0$ for all $h \in H - \{e\}$. Let $J_1 = \{hz_0 : h \in H\}$ be the orbit of z_0. This is a countably infinite subset of S_1.

For a *reductio*, suppose J_1 is normal. Then $P(-|J_1)$ is a finitely additive H-invariant probability function on all subsets of J_1, since $P(hA|J_1) =$

$P(hA|hJ_1) = P(A|J_1)$ for $h \in H$, as $hJ_1 = J_1$ and as P is weakly G-invariant and H is a subgroup of G.

Now let $\nu(A) = P(\{hz_0 : h \in A\}|J_1)$. Let us check that ν is a finitely additive left H-invariant probability function. We have $\nu(H) = P(J_1|J_1) = 1$. If C and D are disjoint subsets of H, then $\{hz_0 : h \in C\}$ and $\{hz_0 : h \in D\}$ are disjoint subsets of J_1. For suppose that z is a member of both, so that $z = cz_0 = dz_0$ for $c \in C$ and $d \in D$. Then $c^{-1}dz_0 = z_0$, and by choice of z_0 it follows that $c^{-1}d = e$, and hence $c = d$, contradicting the disjointness of C and D. Thus, $\{hz_0 : h \in C\}$ and $\{hz_0 : h \in D\}$ are disjoint and so

$$\begin{aligned}\nu(C \cup D) &= P(\{hz_0 : h \in C \cup D\}|J_1)\\ &= P(\{hz_0 : h \in C\} \cup \{hz_0 : h \in D\}|J_1)\\ &= P(\{hz_0 : h \in C\}|J_1) + P(\{hz_0 : h \in D\}|J_1) = \nu(C) + \nu(D).\end{aligned}$$

and so indeed ν is a finitely additive probability function. Moreover, for $h \in H$,

$$\begin{aligned}\nu(hA) &= P(\{h'z_0 : h' \in hA\}|J) = P(\{hh'z_0 : h' \in A\}|J)\\ &= P(h\{h'z_0 : h' \in A\}|J) = P(\{h'z_0 : h' \in A\}|J) = \nu(A),\end{aligned}$$

and so ν is left H-invariant. But we've already seen that there is no left H-invariant finitely additive probability function on H. Thus, J_1 cannot be normal. □

It remains to prove Theorem 3. For any finite subset S of a group G, let $\gamma_S(n)$ be the number of elements in G that can be written in the form $g_1g_2\ldots g_n$, where each of the g_i is either in S or is the inverse of a member of S. Then G is *exponentially bounded* provided that $\lim_{n\to\infty}(\gamma_S(n))^{1/n} = 1$ for every finite $S \subseteq G$. Given two groups G and H, their *direct product* is the set $G \times H$ with the operation $(a, b)(c, d) = (ac, bd)$.

Lemma 4. *If G and H are exponentially bounded, so is their direct product.*

Proof. Let π_1 and π_2 be the projections of $G\times H$ onto G and H, respectively. Let $S \subseteq G \times H$ be finite. Then it is easy to see that $\gamma_S(n) \leq \gamma_{\pi_1[G]}\gamma_{\pi_2[G]}$ and so $\limsup_{n\to\infty}(\gamma_S(n))^{1/n} \leq 1$ by the exponential boundedness of G and H. But $\gamma_S(n) \geq 1$ for all n, so $\lim_{n\to\infty}(\gamma_S(n))^{1/n}$. □

Proof of Theorem 3. Let G be the group of isometries generated by the translations and coordinate reflections. This is a direct product of n copies of G_1, the group of all isometries of $\mathbb{R}$. Now, G_1 is exponentially bounded [20, Cor 12.12], and thus by Lemma 4 so is G. Thus, G is supramenable [20, Thm. 12.8] and the existence of a strongly G-invariant Popper function on $\mathbb{R}^n$ with all non-empty subsets normal follows from [1, p. 7]. □

Proof of Theorem 4. Suppose $\mathcal{F}$ is a G-invariant field of subsets of Ω containing all the H_n and P is a weakly G-invariant Popper function on $\mathcal{F}$. Let $c_P(A, B) = P(A|A \cup B)/P(B|A \cup B)$ as before. By Lemma 2, for all reflections R_x with $x \in \mathbb{Z}/2$, we have $c_P(R_xA, B) = c_P(A, B)$. Since these reflections generate G and in our present setting all members of G map Ω onto itself, we have $c_P(gA, B) = c_P(A, B)$ for all $g \in G$ and $A, B \in \mathcal{F}$.

Let T_1 be the translation on Ω given by $(T_1\omega)(n) = \omega(n+1)$. Observe that T_1 is a bijection of H_2 onto H_1 and $T_1 = R_1 R_{3/2} \in G$. We thus have $c_P(H_1, H_2) = c_P(T_1 H_2, H_2) = c_P(H_2, H_2) = 1$. But $c_P(H_1, H_2) = P(H_1|H_1 \cup H_2)/P(H_2|H_1 \cup H_2) = P(H_1|H_2)$ since $H_1 \subseteq H_2$, and hence under the above assumptions we cannot have $P(H_1|H_2) = 1/2$. $\square$

References

[1] Armstrong, Thomas E. (1989). Invariance of full conditional probabilities under group actions. In: R. D. Mauldin, R. M. Shortt and C. E. Silva (eds.), *Measure and Measurable Dynamics: Proceedings of a Conference in Honor of Dorothy Maharam Stone, held September 17–19, 1987*. 1–22. Providence RI: American Mathematical Society.

[2] Armstrong, Thomas E. and Sudderth, William D. (1989). Locally coherent rates of exchange. *Annals of Statistics* 17:1394–1408.

[3] Bartha, Paul and Johns, Richard (2001). Probability and symmetry. *Proceedings of the Philosophy of Science Association 2001* S109–S122.

[4] de Finetti, Bruno. 1975. *Theory of Probability*. Transl. A. Machi and A. F. M. Smith. New York: Wiley.

[5] Easwaran, Kenny. 2014. "Regularity and hyperreal credences." *Philosophical Review* 123: 1–41.

[6] Hájek, Alan (2003). What conditional probability could not be. *Synthese* 137:273–323.

[7] Hájek, Alan (2011). Conditional probability. In: P. S. Bandyopadhyay and M. R. Forster (eds.), *Handbook of the Philosophy of Science, Volume 7: Philosophy of Statistics*. 99-135. Amsterdam: Elsevier.

[8] Hofweber, Thomas (2013). Infinitesimal chances. *Philosophers' Imprint*, forthcoming.

[9] Just, Winfried (1988). A bounded paradoxical subset of the plane. *Bulletin of the Polish Academy of Sciences – Mathematics* 36:1–3.

[10] Krauss, P. H. (1968). Representation of conditional probability measures on Boolean algebras. *Acta Mathematica Academiae Scientiarum Hungarica* 19:229-241.

[11] McGee, Vann (1994). Learning the impossible. In: Ellery Eells and Brian Skyrms (eds.), *Probability and Conditionals*. Cambridge: Cambridge University Press, 1994.

[12] Parikh, Rohit and Parnes, Milton (1974). Conditional probabilities and uniform sets. In: A. Hurd and P. Loeb (eds.), *Victoria Symposium on Nonstandard Analysis*. 180–194. Berlin: Springer.

[13] Pruss, Alexander R. (2013). Uniform probabilities and Borel paradox. `http://alexanderpruss.blogspot.com/2013/07/uniform-probabilities-and-borel-paradox.html`.

[14] Pruss, Alexander R. (2013). Null probability, dominance and rotation. *Analysis* 73:682–685.

[15] Pruss, Alexander R. (2013). Linear extensions of orders invariant under abelian group actions. arXiv:1309.7295

[16] Pruss, Alexander R. (forthcoming). Regular probability comparisons imply the Banach-Tarski Paradox. *Synthese*.

[17] Pruss, Alexander R. (forthcoming). Two kinds of invariance of full conditional probabilities. *Bulletin of the Polish Academy of Sciences – Mathematics*.

[18] Su, Francis E., et al. (Undated). Sierpinski-Mazurkiewicz Paradox. *Math Fun Facts*. `http://www.math.hmc.edu/funfacts`.

[19] van Fraassen, Bas C. (1976). Representation of conditional probabilities. *Journal of Philosophical Logic* 5:417–430.

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

[21] Williamson, Thomas (2007). How probable is an infinite sequence of heads? *Analysis* 67:173–180.