Some Epistemological Ramifications of the Borel-Kolmogorov Paradox

Michael Rescorla

Abstract: This paper discusses *conditional probability* $P(A \mid B)$, or *the probability of A given B*. When $P(B) > 0$, the ratio formula determines $P(A \mid B)$. When $P(B) = 0$, the ratio formula breaks down. The Borel-Kolmogorov paradox suggests that conditional probabilities in such cases are indeterminate or ill-posed. To analyze the paradox, I explore the relation between probability and intensionality. I argue that the paradox is a *Frege case*, similar to those that arise in many probabilistic and non-probabilistic contexts. The paradox vividly illustrates how an agent's *way of representing* an entity can rationally influence her credal assignments. I deploy my analysis to defend Kolmogorov's relativistic treatment of conditional probability.

§1. The paradox of the sphere

Bertrand (1889) introduced several "problems of geometric probability" that continue to spark debate. Especially puzzling is the so-called *Borel-Kolmogorov paradox*. Imagine that we randomly pick a point ω on the surface of the Earth, considered as a perfect sphere. Say that ω falls on an arc *C* comprising half a great circle. Consider a subarc *D* occupying 1/4 of *C*. What is the probability that ω falls on *D*, given that it falls on *C*? In other words, what is the conditional probability $P(\omega \textit{ falls on } D \mid \omega \textit{ falls on } C)$? Let us notate this conditional probability more compactly as $P(D \mid C)$. An intuitively compelling argument runs as follows (see Figure 1):

> We may choose coordinates so that *C* comprises half the equator (say, the half falling in the Western Hemisphere). Since we picked ω randomly, it is equally likely to fall

> anywhere on the equator. Thus, the conditional probability that ω falls on some subarc is proportional to the subarc's length. So $P(D \mid C) = \frac{\text{length of } D}{\text{length of } C} = ¼$.

Call this *the arc length solution*. The solution seems compelling no matter where *D* is positioned within *C*. Specifically, it applies when *D* and *C* share a common endpoint. Yet, in that case, a second intuitively compelling argument favors an opposing conclusion (see Figure 2):

> We may choose coordinates so that *C* is a meridian and the North Pole is the endpoint shared by *D* and *C*. On this coordinate scheme, *D* extends from the North Pole to latitude 45°. We must therefore compute $P(\text{latitude} \geq 45° \mid C)$. By rotational symmetry around the polar axis, we should obtain the same conditional probability $P(\text{latitude} \geq 45° \mid C^*)$ no matter the meridian C^*. The particular meridian does not affect our final answer. Thus, the answer is just $P(\text{latitude} \geq 45°)$. Since we picked ω randomly, the probability that ω falls in some region is proportional to the region's surface area. So $P(D \mid C) =$
>
> $P(\text{latitude} \geq 45°) = \frac{\text{surface area of the polar cap bounded by } 45°}{\text{surface area of the whole sphere}} < ¼$.

Call this *the surface area solution*. The paradox arises because both the arc length solution and the surface area solution seem convincing, even though they yield different values for $P(D \mid C)$.

INSERT FIGURES 1 AND 2 ABOUT HERE

Mathematicians have discussed the Borel-Kolmogorov paradox extensively. Philosophers seldom address it in detail, with a few notable exceptions. This neglect is surprising, especially when compared with the large philosophical literature on other Bertrand-style problems: selecting a random chord from a circle (Bertrand, 1889); van Fraassen's (1989) box factory; von

Mises's (1957) water/wine paradox; and so on. The Borel-Kolmogorov paradox arguably has greater scientific importance than these other problems. At any rate, I will suggest that it offers important lessons for formal epistemology.

Sects. 2-3 present the paradox more systematically. Sect. 4 introduces Kolmogorov's proposed resolution of the paradox. Sect. 5 investigates the relation between probability and intensionality. Sect. 6 offers my own analysis of the paradox. Sect. 7 deploys my analysis to defend certain controversial aspects of Kolmogorov's formal treatment. Sect. 8 highlights salient unresolved issues.

§2. Conditioning on a null event

In basic undergraduate probability theory, one invariably defines conditional probability through the familiar *ratio formula*:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

The ratio formula is well-defined only when $P(B) > 0$. Yet we often want to discuss probabilities conditional upon *null events*, i.e. events B such that $P(B) = 0$. For example, suppose we randomly choose point ω on a sphere. The probability that ω falls inside some region is proportional to the region's surface area. Since meridian C has surface area 0, the probability that ω falls on C is 0. Nevertheless, we would like to discuss $P(D \mid C)$.

Unfortunately, the Borel-Kolmogorov paradox suggests that probabilities conditional upon null events are sometimes indeterminate or ill-posed. We can motivate distinct conflicting solutions for $P(D \mid C)$.

Borel (1909/1956) tries to eliminate the indeterminacy of conditional probability. He invites us to consider "a thin bundle of arcs of great circles" (p. 82). See Figure 3. Heuristically, each bundle reflects the result of measuring longitude within some margin of error. Suppose we learn that ω falls inside some bundle. Since surface area inside the bundle is concentrated towards the equator and away from the poles, ω is more likely to fall towards the equator. We may regard *C* as the limiting result of increasingly small bundles, corresponding to more precise measurements of longitude. This limiting procedure suggests a non-uniform probability distribution over *C*, biased towards the equator. Borel concludes that we should compute $P(D \mid C)$ through the surface area solution rather than the arc length solution.

INSERT FIGURE 3 ABOUT HERE

As Jaynes (2003, p. 470) emphasizes, one can just as well employ a different limiting procedure, encapsulated by Figure 4. We now consider a bundle of half-parallels. Heuristically, each bundle reflects the result of measuring latitude inside the Western Hemisphere within some margin of error. Suppose we learn that ω falls inside some bundle. Since surface area inside each bundle is distributed uniformly around the equator, the probability that ω falls between any two meridians is proportional to the distance between them. We may regard *C* as the limiting result of increasingly small bundles. This alternative limiting procedure suggests a uniform distribution over *C*, rather than the non-uniform distribution favored by Borel. Thus, limiting procedures do not help resolve the indeterminacy of conditional probability.

INSERT FIGURE 4 ABOUT HERE

In contrast with Borel, de Finetti (1972, pp. 203-204) insists that we should compute

$P(D \mid C)$ through the arc length solution. He pinpoints a key assumption underlying the surface area solution. We first establish that there is a constant k such that

$$P(\ latitude \geq 45° \mid C^*) = k,$$

for all meridians C^*. We then infer that

$$P(\ latitude \geq 45°\ ) = k.$$

This inference implicitly assumes a principle that de Finetti calls *conglomerability*. A weak version of conglomerability runs as follows: for any event E and any partition $\{H_\beta\}$ of possible outcomes,

$$\text{If } P(E \mid H_\beta) = k \text{ for all } \beta, \text{ then } P(E) = k.$$

de Finetti urges us to reject conglomerability, thereby defusing the paradox. Hill (1980) concurs.[1]

A basic problem facing de Finetti and Hill is that conglomerability seems very plausible (Easwaran, 2008), (Jaynes, 2003, pp. 450-489). Countable additivity --- a central element of orthodox probability theory --- entails that conglomerability holds for countable partitions. Countable additivity does not entail conglomerability for uncountable partitions (such as the partition of the sphere into meridians).[2] Nevertheless, conglomerability is plausible even in the uncountable case. If I know in advance that I will obtain the same answer no matter where the outcome falls in some partition, then why can't I choose that answer now without bothering to learn where the outcome falls in the partition? Whatever one thinks of conglomerability *as a general principle,* the particular application of it in the sphere paradox seems compelling. For a satisfying resolution of the paradox, de Finetti must explain why the surface area solution looks

[1] Kadane, Schervish, and Seidenfeld (1986) and Arntzenius, Elga, and Hawthorne (2004) also attack conglomerability.

[2] As Easwaran (2008) observes, decomposition of the sphere into meridians is not technically a partition, because any two meridians share the North and South Pole. There are various technical fixes here, such as associating the North and South Poles with one privileged meridian. I will ignore this issue. Taking it into account would muddy the exposition without affecting any essential features of my argument.

so convincing. Neither de Finetti nor any other proponent of the arc length solution provides the required explanation.

A more fundamental challenge confronts Borel, de Finetti, and anyone else who wants to resolve the paradox by promoting one definite solution for *P*(*D* | *C*). Mainstream probability theory offers techniques for computing conditional probabilities given certain null events. These techniques are a standard element of the advanced undergraduate curriculum. Working statisticians routinely employ them. Yet, as Kolmogorov (1933/1956, p. 51) emphasizes, one can reproduce the paradox *using those standard techniques*. Let me explain.

§3. A paradox within mathematical practice?

Suppose that *X* and *Y* are random variables with a *joint probability density function* (pdf) *p*(*x*, *y*), so that:

$$P[(X,Y)\in A]=\iint_A p(x,y)dxdy\,, \qquad A\subseteq\mathbb{R}^2 \textit{ is any Borel set.}$$

The *conditional pdf of X given Y = y*, or *p*(*x* | *Y* = *y*), is defined as follows:

$$p(x\,|\,Y=y)=_{df}\ \frac{p(x,y)}{p_Y(y)}\,,$$

where $p_Y(y)$ is the *marginal pdf for Y*:

$$p_Y(y)=\int_{-\infty}^{\infty} p(x,y)dx\,.$$

When $p_Y(y)$ = 0, we allow *p*(*x* | *Y* = *y*) to assume any value. In standard undergraduate probability theory, one integrates the conditional pdf *p*(*x* | *Y* = *y*) over a Borel set $A\subseteq\mathbb{R}$ to compute the probability that *x* falls within *A* given that *Y* = *y*:

$$P(X\in A\,|\,Y=y)=\int_A \frac{p(x,y)}{p_Y(y)}dx\,.$$

In this manner, one handles many important cases of conditioning upon null events.

Let us apply these standard techniques to the sphere example. We fix a system of spherical coordinates for the unit sphere:

$$x = \cos\theta\cos\varphi$$

$$y = \cos\theta\sin\varphi$$

$$z = \sin\theta$$

where $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $\varphi \in [0, 2\pi)$. θ is latitude and φ is longitude. For example, the equator corresponds to $\theta = 0$, the North Pole corresponds to $\theta = \frac{\pi}{2}$, and each meridian corresponds to some particular φ. Within this coordinate system, the surface area element has magnitude $\cos\theta\, d\theta\, d\varphi$. If A contains all points such that $\theta_1 \leq \theta \leq \theta_2$ and $\varphi_1 \leq \varphi \leq \varphi_2$, then the probability that a randomly chosen point falls inside A is given by

$$P(A) = \int_{\varphi_1}^{\varphi_2}\int_{\theta_1}^{\theta_1}\frac{\cos\theta}{4\pi}\, d\theta d\varphi\,,$$

where 4π is a normalizing constant to ensure that probabilities sum to 1.

We now introduce random variables Θ and Φ, where Θ measures latitude and Φ measures longitude. For example, each equation $\Phi = \varphi$ determines the meridian corresponding to longitude φ. The joint pdf for Θ and Φ is $p(\theta, \varphi) = \frac{\cos\theta}{4\pi}$. We compute the conditional pdfs $p(\theta \mid \Phi = \varphi)$ and $p(\varphi \mid \Theta = \theta)$:

$$p(\varphi \mid \Theta = \theta) = \frac{p(\theta,\varphi)}{p_{\Theta}(\theta)} = \frac{\frac{\cos\theta}{4\pi}}{\int_0^{2\pi}\frac{\cos\theta}{4\pi}\,d\varphi} = \frac{\frac{\cos\theta}{4\pi}}{\frac{\cos\theta}{2}} = \frac{1}{2\pi}$$

$$p(\theta \mid \Phi = \varphi) = \frac{p(\theta,\varphi)}{p_\Phi(\varphi)} = \frac{\frac{\cos\theta}{4\pi}}{\int_{-\pi/2}^{\pi/2} \frac{\cos\theta}{4\pi} d\theta} = \frac{\frac{\cos\theta}{4\pi}}{\frac{1}{2\pi}} = \frac{\cos\theta}{2}.$$

The first pdf corresponds to the arc length solution, while the second corresponds to the surface area solution. The first pdf is uniform, while the second is smallest at the poles and largest at the equator. We can equally well describe arc *C* through a coordinate system that treats it as half the equator or through a coordinate system that treats it as a meridian. Yet the two coordinate systems yield different conditional probability distributions along the same arc *C*. The contrast between these probability distributions seems paradoxical.

Why do $p(\varphi \mid \Theta = \theta)$ and $p(\theta \mid \Phi = \varphi)$ differ so dramatically? We can trace the difference to the disparate roles that θ and φ play within the formula $\cos\theta \, d\theta \, d\varphi$. The formula varies with θ but not φ. Intuitively, surface area elements "squish together" as one moves further from the equator. See Figure 5. The contrast between θ and φ entails that a randomly chosen point is more likely to be near the equator than the poles but is equally likely to be located at any longitude. It follows that the conditional probability of longitude given latitude differs markedly from the conditional probability of latitude given longitude.

The root phenomenon here does not depend upon special features of spherical geometry. It arises whenever one uses standard undergraduate techniques to condition upon a null event. Here is a simpler example (Lindley, 1982). Consider two independent random variables *X* and *Y* uniformly distributed over the unit square minus the origin: $(0, 1] \times (0, 1]$. What is the probability that $X > ½$, conditional on the event $X = Y$? We can motivate equally compelling answers depending upon how we describe the event $X = Y$. We can introduce a new variable $V = Y - X$, in which case we compute conditional probabilities given $V = 0$. Alternatively, we can introduce a variable $W = Y/X$, in which case we compute conditional probabilities given $W = 1$.

These two procedures correspond respectively to Figures 6 and 7. In Figure 6, we approximate the event $X = Y$ by a narrow bundle of values near $V = 0$. Heuristically, this bundle reflects the result of measuring $Y - X$ within some margin of error. In Figure 7, we approximate $X = Y$ by a narrow bundle of values near $W = 1$. Heuristically, this bundle reflects the result of measuring Y/X within some margin of error. The first procedure suggests that $P(X > ½ \mid X = Y) = ½$. The second procedure suggests that $P(X > ½ \mid X = Y) > ½$, since area is concentrated towards $X = 1$.

INSERT FIGURES 6 AND 7 ABOUT HERE

Once again, we can replicate the intuitive paradox using conditional pdfs. X and Y have joint pdf $p(x, y) = 1$ over $(0, 1] \times (0, 1]$. If we condition on the event $V = 0$, we obtain:

$$p\ (\ x \mid V = 0) = \left. \frac{p(x,v)}{p_V(v)} \right|_{v=0} = 1\ ,$$

where the second identity follows through standard calculus techniques (Gorroochurn, 2012, pp. 199-203). If we condition on the event $W = 1$, we obtain:

$$p(\ x \mid W = 1) = \left. \frac{p(x,w)}{p_W(w)} \right|_{w=1} = 2x\ ,$$

where the second identity again follows through standard calculus techniques. Yet the event $V = 0$ *is* the event $W = 1$! Thus, we derive two opposing answers, depending upon the coordinate system through which we represent the same event. The X-V coordinate system yields a uniform distribution over X, while the X-W coordinate system favors larger values of X.

The sphere and diagonal line examples illustrate a general phenomenon: *conditional probability density is not invariant under coordinate transformation*. Standard techniques for computing conditional pdfs yield different solutions, depending upon our choice of coordinates.

Apparently, then, the coordinate system through which I describe a null event impacts how I should condition on the null event. This dependence upon a coordinate system looks paradoxical. Since the null event remains the same, shouldn't I obtain the same answer either way?

We can generalize the diagonal line example, producing an even more extreme paradoxical conclusion. Consider random variables X and Y with joint pdf $p(x, y)$. Suppose that $range(X)$, $range(Y)$, and $range(X) \cap range(Y)$ are non-empty intervals (possibly of infinite length). Say we want to condition on the null event $X = Y$. Let h be any strictly increasing function on the real numbers. We introduce a new random variable

$$Z = h(X) - h(Y)$$

and compute $p(x \mid Z = 0)$. By varying h, one varies the resulting conditional pdf. Arnold and Roberston (2002) show that the pdf may be almost anything one pleases, given a suitable choice for h. More precisely, let $g(x)$ be any integrable function that is positive on $range(X) \cap range(Y)$. Then we can choose h so that

$$p(x \mid Z = 0) \propto g(x)\, I_{range(X) \cap range(Y)}$$

where $I_{range(X) \cap range(Y)}$ is the characteristic function for $range(X) \cap range(Y)$. So the conditional distribution of X given $X = Y$ can be almost anything, given a suitable choice of coordinates.

Crucially, these paradoxical consequences arise *within standard undergraduate probability theory*. We require no extraneous assumptions or intuitions. We only require the familiar apparatus of conditional pdfs, used daily by undergraduates and professional statisticians. Thus, the paradox is firmly embedded within existing mathematical practice. In this respect, it differs significantly from many Bertrand-style cases widely discussed by philosophers. The random chord, box factory, and water/wine examples generate evident problems only if we assume something like the Principle of Indifference. We can avoid these problems simply by rejecting any such principle. It is much less clear how to handle the Borel-Kolmogorov paradox, which emerges entirely within the standard mathematics of modern probability theory.

Perhaps we should reject the standard undergraduate techniques for computing conditional probabilities? This suggestion is quite radical, since it mandates a substantial revision in contemporary mathematical practice. It also seems rather unpromising, since the notion of conditional pdf looks like a compelling extension of the standard ratio definition. One would need to explain why the extension looks compelling, why it yields fruitful results in so many applications, and why we should nevertheless reject it.

§4. The relativity of conditional probability

Kolmogorov's (1933/1956) diagnosis of the paradox serves as a touchstone for all subsequent discussions. He insists that conditioning upon a null event is *underdetermined*. Even if we fix all unconditional probabilities, and even if we specify the null event upon which we wish to condition, we do not yet fix unique conditional probabilities. As he puts it (p. 51): "the concept of a conditional probability with regard to an isolated given hypothesis whose probability equals 0 is inadmissible." One cannot condition on a null event *simpliciter*. One can only condition upon a null event *relative to other possible events that might have occurred*.

Kolmogorov develops his position using the abstract language of measure theory.[3] The basic object of study is a *probability space* $(\Omega, \mathfrak{F}, P)$, where Ω is any set, $\mathfrak{F}$ is a σ-field over Ω, and P is a probability measure on $\mathfrak{F}$. Probabilities attach to members of $\mathfrak{F}$. Conditional probability is relativized to a sub-σ-field $\mathfrak{G}\subseteq\mathfrak{F}$. Intuitively, $\mathfrak{G}$ serves as an "information filter." When outcome $\omega\in\Omega$ occurs, one does not learn everything about ω. Rather, one learns whether

[3] I follow Billingsley's (1995, pp. 427-440) exposition of Kolmogorov's theory.

ω belongs to each $G \in \mathcal{G}$. On that basis, one updates the probability assigned to $A \in \mathcal{F}$. Thus, we want a function that assigns a probability to event *A given outcome* ω *as filtered through* $\mathcal{G}$.

What properties should this function have? Let us first consider an easy case. Suppose that Ω admits a countable partition $G_1, G_2, \ldots, G_n, \ldots$ of non-null sets drawn from $\mathcal{F}$. Thus, $\Omega = \bigcup_{n=1}^{\infty} G_n$, $G_n \cap G_m = \varnothing$, $G_n \in \mathcal{F}$, and $P(G_n) > 0$. The ratio formula yields

$$P(A \mid G_n) = \frac{P(A \cap G_n)}{P(G_n)}.$$

Define a function

$$\eta(\omega) =_{df} P(A \mid G_n), \qquad \textit{if } \omega \in G_n.$$

Intuitively, η assigns A an appropriate new probability, given outcome ω as filtered through the partition $G_1, G_2, \ldots, G_n, \ldots$ Let $\mathcal{G}$ be the sub-σ-field generated by the partition $G_1, G_2, \ldots, G_n, \ldots$, i.e. the minimal σ-field containing all the sets G_n. Any $G \in \mathcal{G}$ is the union of elements belonging to the partition, so we may write $G = \bigcup_{k=1}^{\infty} G_{n_k}$. Using the standard axioms of probability theory, one can easily show that

$$P(A \cap G) = \sum_{k=1}^{\infty} P(A \mid G_{n_k}) P(G_{n_k}).$$

By definition of η, we have

$$P(A \cap G) = \int_G \eta(\omega) dP(\omega), \qquad \textit{for any } G \in \mathcal{G}.$$

This identity, called *the integral formula*, generalizes the law of total probability.

Kolmogorov's basic idea is to take the integral formula not as a theorem but as a *defining property* of conditional probability. If $(\Omega, \mathcal{F}, P)$ is probability space and $\mathcal{G} \subseteq \mathcal{F}$ is a sub-σ-field, then a *conditional probability for A given* $\mathcal{G}$ is a function $\eta: \Omega \to \mathbb{R}$ such that:

(1) η is $\mathcal{G}$-measurable.

(2) η satisfies the *integral formula*

$$P(A \cap G) = \int_G \eta(\omega) dP(\omega), \qquad \textit{for any } G \in \mathcal{G}.$$

(1) ensures that conditionalization depends only upon information about the outcome ω as filtered through $\mathcal{G}$. Since η is $\mathcal{G}$-measurable, it must attain the same value for any ω and ω^* that belong to the same elements of $\mathcal{G}$. Intuitively: the only differences among outcomes that can inform conditionalization are differences that $\mathcal{G}$ notices. Clause (2) ensures that conditional probabilities satisfy an appropriately generalized law of total probability. It is customary to notate $\eta(\omega)$ as something like $P_{\mathcal{G}}(A \mid \omega)$ so that the integral formula becomes

$$P(A \cap G) = \int_G P_{\mathcal{G}}(A \mid \omega) dP(\omega), \qquad \textit{for any } G \in \mathcal{G}.$$

This notation is potentially misleading, since it suggests that there exists a single unique conditional probability satisfying (1)-(2). In fact, any $\mathcal{G}$-measurable function that agrees with $P_{\mathcal{G}}(A \mid \omega)$ on a set of P-measure 1 is also a conditional probability for A given $\mathcal{G}$. Nevertheless, I follow standard practice in using the potentially misleading notation.

Kolmogorov's definition extends the ratio formula and the theory of conditional pdfs. Suppose again that Ω admits a countable partition $G_1, G_2, \ldots, G_n, \ldots$ with $P(G_n) > 0$, and let $\mathcal{G}$ be the sub-σ-field generated by the partition. One can show that the function

$$P_{\mathcal{G}}(A \mid \omega) =_{df} \frac{P(A \cap G_n)}{P(G_n)}, \qquad \textit{if } \omega \in G_n$$

is a conditional probability for A given $\mathcal{G}$. Similarly, consider random variables $X: \Omega \to \mathbb{R}$ and $Y: \Omega \to \mathbb{R}$ over $(\Omega, \mathcal{F}, P)$. Suppose X and Y have joint pdf $p(x, y)$. Let $\sigma(Y)$ be the σ-field generated by Y (i.e. the smallest σ-field relative to which Y is measurable). For any Borel set B, let

$$P_{\sigma(Y)}(X\in B \mid \omega) =_{df} P(X\in B \mid Y = Y(\omega)),$$

where the right-hand side is defined as in Sect. 3. Then one can show that $P_{\sigma(Y)}(X\in B \mid \omega)$ is a conditional probability for $X\in B$ given $\sigma(Y)$.[4]

Given a probability space $(\Omega, \mathfrak{F}, P)$, a sub-σ-field $\mathfrak{G}\subseteq\mathfrak{F}$, and an event $A\in\mathfrak{F}$, there exists a conditional probability for A given $\mathfrak{G}$ (Billingsley, 1995, p. 430). In pathological cases, conditional probabilities corresponding to various A may not satisfy the normal probability calculus axioms. More precisely, consider a collection of conditional probabilities $P_{\mathfrak{G}}(A \mid .)$, for each $A\in\mathfrak{F}$. For any fixed $\omega\in\Omega$, these conditional probabilities induce a function $P_{\mathfrak{G}}(. \mid \omega): \mathfrak{F} \rightarrow \mathbb{R}$. However, the induced function may not be a probability measure (Billingsley, 1995, p. 443). Conditioning upon new evidence should carry one to a probability measure, so we would like to avoid such pathologies. Luckily, general existence theorems ensure that we can avoid pathological cases in a wide variety of circumstances. Say that a function $P_{\mathfrak{G}} : \mathfrak{F} \times \Omega\rightarrow \mathbb{R}$ is a *regular conditional distribution (rcd) for P given* $\mathcal{G}$ when

$P_{\mathfrak{G}}(A \mid .)$ is a conditional probability for A given $\mathcal{G}$, *for all* $A\in\mathcal{F}$

$P_{\mathfrak{G}}(. \mid \omega)$ is a probability measure, *for all* $\omega\in\Omega$.

One can show that, under fairly general assumptions, rcds always exist (Billingsley, 1995, p. 439), (Rao, 2005, pp. 125-182).

A key aspect of Kolmogorov's framework is that one never conditions upon an isolated event. Rather, one conditions upon an outcome *as filtered through a sub-σ-field* $\mathcal{G}$. In that sense, conditional probabilities are relative to sub-σ-fields. Consider the diagonal line example. The sample space Ω is $(0, 1] \times (0, 1]$, the σ-field comprises the Borel subsets of $(0, 1] \times (0, 1]$, and

[4] Following standard practice, I use "$X\in B$" as shorthand for "$\{\omega : X(\omega)\in B\}$" and "$\{X = x\}$" as shorthand for "$\{\omega : X(\omega) = x\}$".

the probability measure is Lebesgue measure. $X: \Omega \to \mathbb{R}$ and $Y: \Omega \to \mathbb{R}$ map each outcome $\omega = (x, y)$ to the x coordinate or y coordinate, respectively. We can take the conditioning sub-σ-field to be $\sigma(V)$, where $V = Y - X$. Or we can take the conditioning sub-σ-field to be $\sigma(W)$, where $W = Y/X$. The null event $\{V = 0\}$ (i.e. the null event $\{W = 1\}$) belongs to both $\sigma(V)$ and $\sigma(W)$. These different conditioning sub-σ-fields yield different rcds $P_{\sigma(V)}$ and $P_{\sigma(W)}$, corresponding respectively to the conditional pdfs $p(x \mid V = 0) = 1$ and $p(x \mid W = 1) = 2x$ from Sect. 3. A similar diagnosis applies to the sphere example: the conditioning sub-σ-field corresponding to latitude yields a different rcd than the conditioning sub-σ-field corresponding to longitude. The first rcd corresponds to the arc length solution, the second to the surface area solution.

Kolmogorov's relativization of conditional probabilities to conditioning sub-σ-fields troubles many researchers, including Hill (1980), Howson (2014), and Myrvold (forthcoming). Why should conditioning on a null event require relativization? When we condition on a non-null event, the ratio formula yields a single definite answer. Our intuitive concept of conditional probability does not seem to involve relativity to sub-σ-fields. Doesn't this show that Kolmogorov departs substantially from our pre-theoretic starting point? Perhaps most distressingly, Kolmogorov's theory may seem useless for practical statistical inference. Suppose one learns that a null event has occurred. How should one update one's probabilities? Kolmogorov provides no answer, except "It depends on the conditioning sub-σ-field." This answer leaves conditional probabilities almost completely arbitrary. As Kadane, Schervish, and Seidenfeld (1986, pp. 224-225) put it, Kolmogorov's relativistic viewpoint "is unacceptable from the point of view of the statistician who, when given the information that [a null event] has occurred must determine the conditional distribution."

I seek to assuage these worries. I will argue that we should happily embrace relativization of conditional probabilities to conditioning sub-σ-fields. In Sects. 5-6, I analyze the Borel-Kolmogorov paradox using tools drawn from philosophy of mind. In Sect. 7, I deploy my analysis to defend Kolmogorov's relativistic viewpoint.

The literature offers several alternatives to Kolmogorov's theory of conditional probability, including theories due to Dubins (1975), Jaynes (2003), Popper (1959), and Rényi (1955). For various reasons, philosophers often prefer one or another of these alternative frameworks.[5] However, Kolmogorov's theory retains orthodox status within contemporary mathematics. It informs every standard graduate level textbook. So there is considerable philosophical interest in examining how well Kolmogorov's framework *taken on its own terms* handles the Borel-Kolmogorov paradox. That is the task I undertake here.

§5. The objects of credence

The Borel-Kolmogorov paradox arises within the mathematics of probability, whether one interprets probabilities as *credences*, *frequencies*, *propensities*, *logical probabilities*, *objective chances*, or whatever else. One must address the paradox however one interprets the probability calculus. But we should not necessarily expect the same treatment for each interpretation, since a satisfying treatment may cite non-mathematical aspects of our interpretation. I focus exclusively upon the interpretation of probabilities as credences (i.e. *subjective probabilities*). How the paradox bears upon alternative interpretations is a question that I leave to proponents of those interpretations.

Since credences are "degrees of belief," it seems likely that they attach in the first

[5] Hájek (2003, 2011) and Kadane, Schervish, and Seidenfeld (2001) criticize Kolmogorov on various grounds. Easwaran (2008, 2011) responds to some of these criticisms. Hájek favorably discusses Popper's (1959) framework, while Kadane, Schervish, and Seidenfeld recommend Dubins's (1975) approach.

instance to the same entities that serve as objects of belief: *propositions* (Jaynes, 2003), or *natural language sentences* (Davidson, 2004, pp. 151-166), or *mental representations* (Field, 2001, pp. 72-82, pp. 153-156), or *sets of epistemically possible scenarios* (Chalmers, 2011), or something else along these lines. For lack of a better term, I will describe the objects of credence as *propositions*. I remain fairly neutral about the nature of propositions. I also allow that we may choose different entities to serve as propositions, depending upon our theoretical purposes.

However we construe the objects of credence, we must accommodate cases where a thinker represents a single entity "in different ways." As Chalmers (2011) emphasizes, one can readily construct probabilistic analogues to the Fregean *paradoxes of identity* widely discussed within philosophy of language and mind. Consider Frege's most famous example: a thinker who does not realize that Hesperus is Phosphorus. This thinker could attach quite different credences to the proposition *Hesperus has craters* and the proposition *Phosphorus has craters*. For example, she might have credences:

$P(\textit{Hesperus has craters}) = .9$

$P(\textit{Phosphorus has craters}) = .2$

The contrast in credence would manifest in various ways: her propensity to assert the relevant propositions; the gambles she would willingly accept; and so on. She might also have conditional probabilities along the following lines:

$P(\textit{ Hesperus has craters} \mid \textit{Hesperus has craters}) = 1$

$P(\textit{ Hesperus has craters} \mid \textit{Phosphorus has craters}) = .2$

These credal assignments seem perfectly coherent. There is nothing rationally or cognitively defective about our hypothetical thinker. She simply does not know that Hesperus is Phosphorus.

In a sense, the propositions *Hesperus has craters* and *Phosphorus has craters* describe the same "state of affairs." A thinker can represent this state of affairs *in different ways* --- as Burge (2009) puts it, from different *representational perspectives*. A good theory must allow differences in representational perspective to yield different rational credal assignments.

The Hesperus-Phosphorus example involves *empirical* ignorance. As Frege emphasized, one can also construct non-empirical Frege cases. Consider the propositions *The box has sides of length 27 inches* and *The box has volume 19683 cubic inches*. In some sense, these propositions describe the same "state of affairs." Still, a thinker might not realize that $27^3 = 19683$. She might have conditional probabilities

P(*The box has volume 19683* | *The box has volume 19683*) = 1

P(*The box has volume 19683* | *The box has sides of length 27 inches*) = .2

There does not seem to be anything irrational or incoherent about this hypothetical thinker. True, she is not mathematically or logically omniscient. She does not know that $27^3 = 19683$, which we may assume is a straightforward logical consequence of her other arithmetical beliefs. But mathematical and logical omniscience are not requirements of rationality.[6]

Philosophers have intensely debated non-probabilistic Frege cases. Obviously, theories that postulate *Fregean senses* are explicitly tailored to these cases. How well do rival frameworks fare? The answer is contentious. As Chalmers (2011) notes, many familiar maneuvers from philosophy of language and mind will also arise in the debate over probabilistic Frege cases. I do not want to presuppose any specific position in these debates. I just want to emphasize a very basic point: a complete theory of credence must accommodate probabilistic Frege cases, including the Hesperus-Phosphorus example and the length-volume example. Perhaps a good theory will postulate Fregean senses. Perhaps it will proceed along quite different

[6] The example recalls van Fraassen's box factory. For related discussion, see (Bangu, 2010).

lines. One way or another, a complete theory must respect *intensional distinctions*, i.e. distinctions among ways of representing entities.

§6. Analyzing the paradox

In my view, the Borel-Kolmogorov paradox occurs because one can condition upon the same event *as represented in different ways*. Credences are rationally sensitive to the way that one represents an event, so it is not surprising that different conditional probabilities arise when one represents the conditioning event differently. The Borel-Kolmogorov paradox is a Fregean "paradox of identity."

To illustrate, let *C* be half a great circle on the unit sphere. We can introduce coordinates that treat *C* as a meridian, or we can introduce coordinates that treat *C* as half the equator. Let random variable Φ measure longitude in the first coordinate system, with

$$\omega \text{ falls on arc } C \text{ iff } \Phi(\omega) = \varphi_0$$

for some φ_0. Let random variable Ψ measure latitude in the second coordinate system, with

$$\omega \text{ falls on arc } C \text{ iff } \Psi(\omega) = \psi_0$$

for some ψ_0.[7] Φ partitions the sphere into meridians $\{\Phi = \varphi\}$, while Ψ partitions the sphere into half-parallels $\{\Psi = \psi\}$. The two partitions share the null event $\{\Phi = \varphi_0\}$ (i.e. the null event $\{\Psi = \psi_0\}$). When we regard this null event as belonging to the partition generated by Φ, we represent

[7] Random variable Ψ differs from the random variable Θ from Sect. 3. Θ measures latitude in the same coordinate system as Φ, while Ψ measures latitude in a second coordinate system. Arc *C* is a meridian in the first coordinate system, while *C* occupies half the equator in the second coordinate system. Lines of constant Θ are parallels, while lines of constant Ψ are half-parallels. $range(\Theta) = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, while $range(\Psi) = (-\pi, \pi]$. These changes are to ensure that some equation $\Psi = \psi_0$ defines *C*. It is natural to introduce an additional random variable Γ with $range(\Gamma) = [0, \pi)$, where each equation $\Gamma = \gamma$ defines a great circle through the North Pole of the second coordinate system. I will describe Γas measuring longitude, although strictly speaking a point that satisfies $\Gamma = \gamma$ may have either longitude γ or longitude $\gamma+\pi$. One can show that $p(\gamma \mid \Psi = \psi) = \frac{1}{\pi}$.

it differently than when we regard it as belonging to the partition generated by Ψ. Thus, it is hardly surprising that conditioning on the null event *as represented through* Φ yields a different result than conditioning on that same null event *as represented through* Ψ. The proposition that $\Phi = \varphi_0$ is distinct from the proposition that $\Psi = \psi_0$.

Intuitively speaking, there is a large difference between representing *C* through the equation $\Phi = \varphi_0$ versus the equation $\Psi = \psi_0$. In the first case, one represents *C* as one among many meridians through a fixed North Pole. In the second, one represents *C* as one among many half-parallels (relative to a different fixed North Pole). These two modes of representation presuppose different coordinate systems. The difference in coordinate systems generates numerous additional disparities:

- Different coordinates systems are canonically associated with different procedures for determining position on the sphere's surface.
- Different coordinate systems are canonically associated with different symmetry groups. Each coordinate system is canonically associated with the group of rotations around its polar axis.
- Different coordinate systems are canonically associated with different procedures for calculating surface integrals, as encapsulated by different surface area elements.

These disparities reflect (or perhaps help constitute) a difference in representational perspective upon *C*. The informal and formal arguments adduced in Sects. 1-3 demonstrate that such disparities bear directly upon rational credal assignment. How one represents *C* influences how one should condition upon the news that *C* contains ω.

A similar diagnosis applies to the diagonal line example. The proposition that $V = 0$ is distinct from the proposition that $W = 1$, even though the propositions specify the same null event (i.e. the event $X = Y$). It is not surprising that one obtains a different result when one conditions upon the first proposition than when one conditions upon the second.

Non-probabilistic Frege cases show that one can bear rational conflicting attitudes towards distinct propositions that represent the same "state of affairs." Probabilistic Frege cases show that representational perspective informs rational credal assignments. The Borel-Kolmogorov paradox instantiates the general sensitivity of rational credal assignment to representational perspective. An apparent contradiction arises only if we mistakenly demand that a thinker's conditional probabilities agree when she represents the conditioning event in different ways. Thus, the Borel-Kolmogorov paradox is not remotely paradoxical.

I can reformulate my analysis in more meta-linguistic terms. How does the apparent paradox arise? We describe a null event B, and we argue that

(1) $$P(A \mid B) = r.$$

We also describe a null event B^*, and we argue that

$$P(A \mid B^*) = r^*,$$

where $r \neq r^*$. We then generate a contradiction by noting that

(2) $$B = B^*$$

and inferring from (1) and (2) that

(3) $$P(A \mid B^*) = r.$$

However, the inference from (1) and (2) to (3) is fallacious, because the locution "$P(A \mid \ .\ )$" creates an opaque context. One cannot freely intersubstitute descriptions "B" and "B^*" of the same event, even when the identity "$B = B^*$" is discoverable through *a priori* means. This observation blocks derivation of (3), thereby dispelling the threat of contradiction.[8]

[8] Proschan and Presnell (1998) also condemn as fallacious the inference from (1) and (2) to (3). My discussion places this diagnosis in a broader explanatory context, by relating it to general Fregean paradoxes of identity.

We can now clarify the sense in which conditional probabilities are "relative." Most fundamentally, they are relative to an agent's representational perspective upon the conditioning event. In more familiar philosophical terminology, they are relative to the proposition through which one represents the conditioning event. If $\sigma(X) \neq \sigma(Y)$, then learning that $X = x_0$ is a different doxastic state than learning that $Y = y_0$. The difference rationally impacts credal assignments. I submit that the Borel-Kolmogorov paradox provides fresh vindication for Frege's seminal insight: mental processes are rationally sensitive to fine-grained differences in representational perspective. The paradox suggests that this rational sensitivity is deeply lodged within the mathematical structure of conditionalization.[9]

To highlight the influence that representational perspective exerts upon credal assignment, let us consider Hill's (1982) discussion of the paradox. Following de Finetti, Hill favors the arc length solution over the surface area solution. He writes (p. 44):

> Suppose the sphere is not labeled with a prespecified north pole, and you regard the point as uniformly distributed on the surface of the sphere. If you are given *only* the information that the point lies on some *exact* great circle, would you now regard the point as uniform on that great circle (since there is no north pole, no other distribution seems natural, so presumably you either regard the distribution as uniform, or else consider it as indeterminate)?

[9] Some authors claim that conditional probabilities depend not just upon the conditioning event but also upon *the way one learns* that the conditioning event occurred (Easwaran, 2008, pp. 88-89), (Lindley, 1997, p. 184), (Shafer, 1985, p. 262). In my view, the key explanatory factor is not *the way one learns* that the conditioning event occurred but rather *the way one represents* the conditioning event. One can learn through many diverse avenues that an event *as represented a single fixed way* occurred. In some cases, a fixed way of representing the conditioning event may have constitutive ties to certain canonical verification procedures. But non-canonical verification procedures can also establish that the event *as represented that same way* occurred. For example, one can learn that $\Phi = \varphi_0$ through direct measurement, deductive reasoning, testimony, abductive inference, or various other avenues. These variations in learning method do not seem relevant to conditionalization. What matters is simply that one represents the conditioning event *in a certain way*: namely as the event $\Phi = \varphi_0$. The content of one's knowledge, not its etiology nor its justificatory basis, is the relevant factor.

Hill's formulation elides a crucial issue: in what way does one represent *that the chosen point lies on some exact great circle*? The normal way to represent great circles is to specify a spherical coordinate system, which allows one to isolate an equation satisfied by precisely those points falling on the desired great circle. Hill assumes that we are not representing the exact great circle in this way, since he stipulates that there is no prespecified north pole. Thus, it is unclear how Hill imagines us representing the null event upon which we are to condition. But then one should not expect a determinate answer to Hill's question.

By analogy, consider the following question: how should one update one's credence in the proposition *Hesperus has craters*, assuming one learns *of* Hesperus that it has craters? This question has no determinate answer, because it does not specify whether one learns *that Hesperus has craters* or *that Phosphorus has craters*. Similarly, Hill stipulates that one learns *of* a great circle that it contains the chosen point, without specifying how one represents this great circle. Hill has not described a single determinate doxastic state that can serve as input for conditionalization.

Hill might demand that our theory dictate determinate probabilities even when one does not represent the relevant entities in any particular way. However, this demand strikes me as illegitimate. We finite humans are not capable of representing an entity without representing it any some particular way. As Burge puts (2009, pp. 251-252): "We lack cognitive power to perceive or think of ordinary entities in no way at all, or to incorporate them whole into perception or thought --- apart from any representational means that constitutes one of many possible perspectives on them." Perhaps an omniscient mind could incorporate entities whole into thought. But finite minds such as our own can only represent an entity from some specific representational perspective. The Borel-Kolmogorov paradox vividly illustrates how an agent's way of representing a null event helps determine how she should conditionalize upon the event.

§7. In defense of relativity

I have related the Borel-Kolmogorov paradox to Fregean paradoxes of identity. I now deploy my analysis to defend Kolmogorov's relativistic treatment of conditional probability.

Kolmogorov's axioms say nothing about the entities that compose Ω. One can smuggle intensional distinctions into Kolmogorov's framework by letting Ω contain intensional entities. For example, Chalmers (2011) argues that we should sometimes take Ω to contain epistemically possible scenarios. One might also take Ω to contain mental representations, Fregean thoughts, or various other such entities.

Mathematicians eschew intensional construals of Ω. They invariably offer a purely extensional treatment, individuating Ω's elements without regard to representational perspective. For example, when mathematicians define Lebesgue measure over the Borel sets, they do not consider how one represents Borel sets. The impressive achievements of probability theory (as manifested by any standard textbook) demonstrate how much progress is possible when one ignores all intensional distinctions. One can often ignore such distinctions even when studying conditional probability, provided that one focuses solely upon non-null conditioning events.

Nevertheless, the Borel-Kolmogorov paradox indicates that one can proceed only so far while disregarding all differences in representational perspective. To analyze conditioning upon null events, one must capture certain key intensional distinctions within one's mathematical model. I suggest that we interpret Kolmogorov's theory in this light. On my interpretation, Kolmogorov captures vital intensional distinctions by relativizing conditional probabilities to conditioning sub-σ-fields. *Different sub-σ-fields reflect different ways of representing the same event*. The rest of Sect. 7 develops this interpretation. I focus exclusively upon the special case where the conditioning sub-σ-field is $\sigma(X)$, for some random variable X.[10]

§7.1 From propositions to sub-σ-fields

Suppose that we use a probability space $(\Omega, \mathcal{F}, P)$ to model some thinker's initial credences. Suppose that we use random variables $X : \Omega \rightarrow \mathbb{R}$ and $Y: \Omega \rightarrow \mathbb{R}$ to model possible news that the thinker might receive. When the thinker conditions upon news of X's value, Kolmogorov models her doxastic state through $\sigma(X)$. When she conditions upon news of Y's value, he models her doxastic state through $\sigma(Y)$. It may be that $\sigma(X) \neq \sigma(Y)$, even though there exist x_0 and y_0 such that $\{X = x_0\} = \{Y = y_0\}$. Then Kolmogorov postulates distinct doxastic states corresponding to the event $\{X = x_0\}$, i.e. the event $\{Y = y_0\}$. He thereby accommodates the Fregean insight that philosophers would typically express by saying: the proposition that $X = x_0$ and the proposition that $Y = y_0$ are distinct.

To illustrate, consider the sphere example. We may take the probability space to be $(E, \mathcal{B}(E), P)$, where E is the surface of the unit sphere, $\mathcal{B}(E)$ comprises the Borel subsets of E, and P is a uniform distribution over the sphere's surface. Suppose our thinker learns ω's longitude, as described using some spherical coordinate system. Kolmogorov models her doxastic state with a random variable Φ, which determines a sub-σ-field $\sigma(\Phi)$. In contrast, suppose our thinker learns ω's latitude, as described within a different spherical coordinate system. Kolmogorov models her doxastic state with a different random variable Ψ, which determines a different sub-σ-field $\sigma(\Psi)$.

[10] For ease of exposition, I will frequently attribute various doctrines and assumptions to Kolmogorov. However, I do not suggest that my treatment explicates Kolmogorov's intentions. I claim only that it constitutes one fruitful way of relating Kolmogorov's mathematical framework to the non-mathematical realm. Kolmogorov favored a frequentist viewpoint (1933/1956, p. 3), so presumably he would not have endorsed my subjectivist approach.

We may suppose that $\{\Phi = \varphi_0\} = \{\Psi = \psi_0\}$. Nevertheless, Kolmogorov differentiates *learning that* $\Phi = \varphi_0$ versus *learning that* $\Psi = \psi_0$ by specifying $\sigma(\Phi)$ versus $\sigma(\Psi)$ as the conditioning sub-σ-field. In this manner, Kolmogorov honors important differences in representational perspective without expanding his ontology to include mental representations, propositions, or the like.

I do not say that conditioning sub-σ-fields can do *all* the work that philosophers demand of propositions or mental representations. Classification of doxastic states through sub-σ-fields is coarser-grained than classification through any reasonable notion of propositional content. In particular, random variables X and Y often reflect different ways of representing the same event even though $\sigma(X) = \sigma(Y)$. For example, renumbering the lines of constant longitude would induce a new random variable Φ^* but the same σ-field $\sigma(\Phi^*) = \sigma(\Phi)$. Thus, we cannot backtrack from the conditioning sub-σ-field $\sigma(X)$ to the particular way that an agent represents the event $\{X = x_0\}$. For that reason, Kolmogorov's emphasis upon $\sigma(X)$ discounts many important differences in representational perspective. A more psychologically realistic model would take those differences into account.

The crucial point is that Kolmogorov honors *some* differences in representational perspective. He uses sub-σ-fields to capture vital intensional distinctions. By capturing these intensional distinctions, Kolmogorov models the rational impact of representational perspective upon credal assignment. For example, there is a large difference between representing *C as a line of constant longitude in some coordinate system* and representing *C as a line of constant latitude in some other coordinate system*. Kolmogorov marks this difference by citing distinct sub-σ-fields $\sigma(\Phi)$ and $\sigma(\Psi)$. The distinct sub-σ-fields generate distinct rcds $P_{\sigma(\Phi)}$ and $P_{\sigma(\Psi)}$, corresponding respectively to density $\frac{\cos\theta}{2}$ and to a constant density over longitude.[11] The contrast between these two rcds reflects the fact that one is conditioning upon the same null event *as represented in different ways*.

I conclude that the relativistic aspects of Kolmogorov's theory are not remotely troubling. On the contrary, relativization to sub-σ-fields is an elegant device for preserving intensional distinctions within an extensional setting. The real surprise is that probability theorists are forced to honor key intensional distinctions *only* when studying conditionalization on null events.

Easwaran (2008, 2011) also embraces Kolmogorov's relativistic approach to conditional probability. However, he interprets the relativization in a very different way than I do. He proposes that we relativize conditional probability to a partition of Ω, where the partition reflects "relevant alternatives" to the conditioning event. Which alternatives are relevant depends upon the context of inquiry. Thus, Easwaran relativizes conditional probability to a contextually determined parameter. He does not say why conditioning on a null event should be relativized to this contextually determined parameter. Rather than *explain* why conditional probabilities are relativized, Easwaran's account merely stipulates that they are.

My account explains the relativity of conditional probability by grounding it in more general aspects of mental representation. I postulate no relativity beyond the relativity one expects from all rational mental activity: relativity to representational perspective. In particular, I do not postulate relativity to a contextually determined parameter. The proposition through which one represents the conditioning event $\{X = x_0\}$ already determines an appropriate conditioning sub-σ-field $\sigma(X)$, quite independently of context. Probability theory establishes that,

[11] When I use the phrase "a constant density over longitude" here and in subsequent passages, I mean the conditional pdf $p(\gamma \mid \Psi = \psi) = \frac{1}{\pi}$ described in note 7.

for certain theoretical purposes, one can ignore relativity to representational perspective when studying conditionalization on non-null events. As the Borel-Kolmogorov paradox demonstrates, the relativity becomes inescapable when one studies conditionalization on null events.

§7.2 A model of doxastic state

Let us examine more closely how Kolmogorov uses σ-fields to model doxastic states. The basic idea is that an agent learns whether outcome ω belongs to G, for each G in the conditioning sub-σ-field $\mathcal{G}$. In the case that concerns us, $\mathcal{G} = \sigma(X)$ for some random variable X. One can easily show that $\sigma(X)$ consists of the sets

$$X^{-1}(B) \qquad B \subseteq \mathbb{R} \textit{ is any Borel set}.$$

Thus, our agent learns the answer to each question

$$\text{Is } \omega \in X^{-1}(B)? \qquad B \subseteq \mathbb{R} \textit{ is any Borel set}.$$

If X has uncountably many values (as in the sphere and diagonal line examples), then there are uncountably many such sets $X^{-1}(B)$. No finite agent can simultaneously represent membership facts about uncountably many sets. Taken literally, then, Kolmogorov's framework postulates infinitary mental capacities that vastly outstrip actual human capacities.

We can reinterpret Kolmogorov's framework so as to reduce its reliance upon infinitary mental capacities. $\sigma(X)$ is *countably generated*: there exists a countable sequence $S_1, S_2, \ldots S_n, \ldots$ such that $\sigma(X)$ is the minimal σ-field containing $S_1, S_2, \ldots S_n, \ldots$ For example, we may take $S_1, S_2, \ldots S_n, \ldots$ to be sets of the form

$$X^{-1}(a, b) \qquad a, b \in \mathbb{Q}.$$

Many other choices for $S_1, S_2, \ldots S_n, \ldots$ are possible. The key point is that $\sigma(X)$ results when we start with $S_1, S_2, \ldots S_n, \ldots$ and close under complementation and countable union. Hence:

> If knowledge of membership facts is closed under complementation and countable union, then knowing whether outcome ω belongs to each S_1, S_2, … S_n, … entails knowing whether ω belongs to any member of $\sigma(X)$.

When we use $\sigma(X)$ to model an agent's knowledge, we need not say that she explicitly represents all elements of $\sigma(X)$. Nor need we say that her knowledge is closed under complementation and countable union. We need only say that, if her knowledge *were* closed under complementation and countable union, then she could extrapolate from knowledge about S_1, S_2, … S_n, … to full membership knowledge for members of $\sigma(X)$. In that sense, she has *implicit knowledge* of membership facts for members of $\sigma(X)$. Call this *the implicit knowledge maneuver*.

If we adopt the implicit knowledge maneuver, then we need not postulate an agent who simultaneously represents membership facts about $\sigma(X)$'s uncountably many elements. We need merely postulate an agent who learns the answer to each question

$$\text{Is } \omega \in S_n?$$

For example, suppose our agent learns the answer to every question

$$\text{Is } \omega \in X^{-1}(a, b)? \qquad a, b \in \mathbb{Q}.$$

By learning answers to these questions, she gains implicit knowledge. Complete membership information for $\sigma(X)$ models that implicit knowledge.

We may therefore use $\sigma(X)$ to model implicit knowledge gained by an ideal agent who learns whether X's value falls inside each rational open interval (a, b). To illustrate, take the sphere example. Suppose our agent learns the answer to each question

$$\text{Is } \Phi(\omega) \in (a, b)? \qquad a, b \in \mathbb{Q}.$$

In other words, she learns whether ω's longitude falls within any given rational open interval. She can thereby readily discern the answer to each question

(1) Is $\omega\in\Phi^{-1}(a, b)$? $a, b\in\mathbb{Q}$.

$\Phi^{-1}(a, b)$ is a bundle of meridians (as in Figure 3). We use $\sigma(\Phi)$ to model implicit knowledge gained by the agent. By contrast, suppose the agent learns the answer to each question

(2) Is $\omega\in\Psi^{-1}(a, b)$? $a, b\in\mathbb{Q}$.

$\Psi^{-1}(a, b)$ is a bundle of half-parallels (as in Figure 4). We use $\sigma(\Psi)$ to model implicit knowledge gained by the agent. Answers to all questions (1) determine a unique null event $\{\Phi = \varphi\}$, while answers to all questions (2) determine a unique null event $\{\Psi = \psi\}$. We may suppose that $\{\Phi = \varphi\} = \{\Psi = \psi\}$. Nevertheless, knowing the answers to all questions (1) is a different doxastic state than knowing the answers to all questions (2). The two doxastic states involve membership knowledge about different regions of the sphere. To model the disparate impact that these distinct doxastic states exert upon rational credal allocation, Kolmogorov invokes distinct rcds $P_{\sigma(\Phi)}$ and $P_{\sigma(\Psi)}$.

§7.3 Infinitary mental capacities

The implicit knowledge maneuver presupposes that our agent can learn answers to each question

Is $\omega\in S_n$?

for some sequence $S_1, S_2, \ldots S_n, \ldots$ There are continuum many sequences of answers "yes" or "no" to these questions. Representing an arbitrary sequence drawn from continuum many options is an infinitary mental capacity. It is a relatively modest infinitary capacity --- roughly comparable to a capacity for representing arbitrary real numbers. Still, the postulated capacity far outstrips our own finite mental capacities. Thus, even though the implicit knowledge maneuver *reduces* reliance on infinitary mental capacities, Kolmogorov's modeling framework ultimately posits infinitary capacities unavailable to finite minds.

Readers may question this reliance upon infinitary capacities. How much can epistemology gain by studying agents whose cognitive resources so vastly outstrip our own?

I offer a two-pronged response. First, Kolmogorov's model of conditional probability captures a core aspect of conditioning on null events: representational perspective rationally influences credal assignment. Kolmogorov tracks how representational perspective informs conditionalization *for an ideal agent whose infinitary mental capacities exceed our own actual finite mental capacities*. Rational credal allocation depends upon the way one represents conditioning null events *even for the ideal agent*. Thus, Kolmogorov's model highlights crucial links between probability and intensionality, links that persist even for ideal agents with infinitary mental capacities.[12] If rational credal assignments of an infinitary mind vary with representational perspective, then all the more so should rational credal assignments of a finite mind. As we pursue theories that honor the finite nature of human cognition, we should expect that these theories will assign continued importance to the relation between representational perspective and conditional probability. The tight links between probability and intensionality seem to reflect basic mathematical constraints on conditionalization. Those constraints encompass both finite minds *and* minds whose infinitary capacities vastly outstrip our own.

Second, even though Kolmogorov's framework is tailored to agents with infinitary mental capacities, we can fruitfully use it to model finite agents. Admittedly, a finite agent cannot represent any arbitrary sequence of answers to the questions:

---

[12] My discussion involves *normative* claims about how ideal agents should update their credences. It thereby differs significantly from the idealizations employed in natural science, which simplify complex reality without making normative claims (Wimsatt, 2007, pp. 15-25). For discussion of ideal agents as an epistemological tool, see (Shaffer, 2007). For discussion of idealization in scientific modeling, see (Shaffer, 2012) and (Wimsatt, 2007, pp. 3-4, pp. 26-36, pp. 94-132, pp. 152-154).

Is $\omega \in S_n$?

Thus, a finite agent cannot instantiate every possible doxastic state modeled through σ(X). Nevertheless, there are *countably many* doxastic states modeled through σ(X) that she can instantiate. She can understand countably many expressions "*a*" that represent real numbers (e.g. "½", ".5684", "π", or "the positive square root of 2"). If expression "*a*" is suitably informative, then knowledge that $X = a$ allows her to answer each question

$$\text{Is } \omega \in X^{-1}(a, b)? \qquad a, b \in \mathbb{Q}.$$

For example, suppose our agent learns that X = ½. Then she can readily discern that

$$\omega \in X^{-1}(a, b) \text{ iff } a < ½ \text{ and } b > ½ \qquad \textit{for all } a, b \in \mathbb{Q}.$$

Or suppose she learns that X = .5684. Then she can readily discern that

$$\omega \in X^{-1}(a, b) \text{ iff } a < \frac{5684}{10^4} \text{ and } b > \frac{5684}{10^4} \qquad \textit{for all } a, b \in \mathbb{Q}.$$

In such cases, complete membership information for σ(X) models the implicit knowledge gained by learning that $X = a$.

Kolmogorov uses an rcd $P_{\sigma(X)}$ to model rational credal allocation upon learning X's value. Assuming that X has uncountably many values, Kolmogorov's framework models rational credal allocation for all uncountably many values. A finite agent can only represent countably many values of X. So Kolmogorov's framework models uncountably scenarios that our finite agent cannot realize. This is no reason to eschew the framework when addressing those countably many scenarios that she *can* realize. For suitably informative expressions "*a*", we may fruitfully use $P_{\sigma(X)}(\, . \mid \omega \,)$ to model the agent's rational credal reallocation upon learning that $X(\omega) = a$.

Thus, even though Kolmogorov's framework is tailored to agents with infinitary mental capacities, we can use it to model conditioning by a finite agent. In doing so, we will not exploit the framework's full resources. We will ignore how $P_{\sigma(X)}$ handles the uncountably many doxastic states that the agent cannot instantiate. But we will use $P_{\sigma(X)}$ to model rational credal allocation for the countably many doxastic states that the agent *can* instantiate.

§7.4 Beyond random variables

I have discussed situations where an agent conditions upon the value of a random variable X: $\Omega \rightarrow \mathbb{R}$. There are many situations where it is unnatural or impossible to encapsulate an agent's evidence through a real number. Can we extend my analysis to handle the general case of a probability space (Ω, $\mathscr{F}$, P) and a sub-σ-field $\mathscr{G} \subseteq \mathscr{F}$? Do conditioning sub-σ-fields yield good models of doxastic state in the general case? Does the relativistic aspect of Kolmogorov's theory likewise seem compelling?

These questions merit extensive further discussion. But my goal here is not to offer a complete treatment of conditional probability, nor to offer a blanket defense of Kolmogorov's theory. My goal is to defend an oft-criticized feature of Kolmogorov's approach: relativization to a conditioning sub-σ-field. I have argued that this feature is not problematic *for the special case of conditioning on a random variable*. It follows that relativization *per se* is no problem for Kolmogorov. Despite relativization, Kolmogorov provides a compelling treatment of conditioning on a random variable. Whether he can handle other cases equally well is a question best addressed on an individual basis. It may emerge that Kolmogorov's theory yields plausible results in certain cases but not others.

§8. Residual indeterminacy

To address indeterminacy worries raised by the Borel-Kolmogorov paradox, I have argued that conditional probabilities are relative to representational perspective on the conditioning event. Unfortunately, my approach does not eliminate all indeterminacy. I now want to discuss some residual indeterminacy phenomena.

§8.1 Indeterminacy in Kolmogorov's theory

Kolmogorov's theory leaves conditional probabilities indeterminate *even after we specify a conditioning sub-σ-field*. Rcds are unique only up to sets of measure 0. Consider again the sphere example. We have seen that density $\frac{\cos\theta}{2}$ yields an rcd for P given $\sigma(\Phi)$. But there are infinitely many other candidate rcds. For any meridian C, one can find an rcd $P_{\sigma(\Phi)}: \mathscr{B}(E) \times E \rightarrow \mathbb{R}$ that induces whatever probability measure $P_{\sigma(\Phi)}(\,.\mid\omega)$ one likes on points $\omega \in C$. Indeed, one can assign conditional probabilities however one pleases on any countable collection of meridians, since the union of countably many meridians is a null event.

The difference between two candidate rcds does not matter for normal mathematical applications. In particular, the difference is irrelevant if we use rcds only inside integrals, since any two rcds integrate to the same result. Still, the indeterminacy is troubling. Kolmogorov's theory leaves conditional probabilities far less constrained than one would expect. Shouldn't we demand a unique solution, at least in certain cases? We have a strong intuition that $\frac{\cos\theta}{2}$ is rationally superior to other candidates for $P_{\sigma(\Phi)}$. We also have a strong intuition that a constant density over longitude is rationally superior to other candidates for $P_{\sigma(\Psi)}$. Kolmogorov provides an infinite array of candidate solutions when we want a single determinate solution.

Ad hoc arguments sometimes favor a particular rcd. For instance, rotational symmetry in the sphere example favors $\frac{\cos\theta}{2}$ over other candidates for $P_{\sigma(\Phi)}$ (Easwaran, 2008). However, such *ad hoc* devices are not available in the general case. Anyway, we would like a systematic theory that handles as many examples as uniformly possible.

For present purposes, the key point I want to emphasize is that residual indeterminacy does not undermine Kolmogorov's theory. Kolmogorov places constraints on admissible conditional probabilities. The constraints leave conditional probabilities dismayingly indeterminate, which suggests that we need additional constraints. We would like to augment Kolmogorov's theory so as to select a unique or semi-unique solution. Residual indeterminacy does *not* impugn Kolmogorov's constraints. We should take residual indeterminacy as an impetus to *supplement* Kolmogorov's theory, not a reason to *reject* Kolmogorov's theory.

Several authors have explored how one might supplement Kolmogorov's theory to eliminate or at least reduce residual indeterminacy (Pfanzagl, 1979), (Rao, 2005), (Tjur, 1980). So far, no supplementation has gained widespread acceptance. Developing a suitably supplemented theory is an important task for future research. Since I have not indicated how a satisfying supplementation might proceed, I do not claim to have provided a complete account even for the special case of conditioning on a random variable.

§8.2 Combining representational perspectives

In Sects. 6-7, I discussed cases where an agent conditions on a null event *as represented a specific way*. In particular, I compared conditioning on the proposition that $\Phi = \varphi_0$ with conditioning on the proposition that $\Psi = \psi_0$. What happens when an agent simultaneously represents the same null event in two distinct ways? Combination of distinct representational perspectives on the same entity is a familiar phenomenon from the literature on Frege cases. An agent may discover that an entity *as represented one way* is identical to that same entity *as represented a different way*. For example, she may discover *that Hesperus is Phosphorus*.

Many expositions of the Borel-Kolmogorov paradox elicit a realization along these lines. One motivates a conditional probability distribution for arc C (defined by the equation $\Phi = \varphi_0$) and a distinct conditional probability distribution for arc C^* (defined by the equation $\Psi = \psi_0$). One then notes that $C = C^*$ --- i.e. that $(\forall \omega)(\Phi(\omega) = \varphi_0 \text{ iff } \Psi(\omega) = \psi_0)$ --- engendering an apparent contradiction. I have argued that the apparent contradiction is spurious. Yet a related problem remains. Once an agent realizes that $C = C^*$, she can no longer rationally maintain distinct conditional distributions for C and C^*. Although it is perfectly rational for her to maintain distinct conditional probabilities $P(D \mid C)$ and $P(D \mid C^*)$ *before* she realizes that $C = C^*$, it is not rational to do so *after* she realizes that $C = C^*$. How, then, should an agent who knows that $C = C^*$ assign conditional probabilities? There is no intuitively clear answer to this question. Certainly, my analysis from Sects. 6-7 does not recommend any determinate answer. I have addressed rational credal allocation if one learns *that* ω *falls on* C or if one learns *that* ω *falls on* C^*, but I have not addressed rational credal allocation if one learns *that* ω *falls on* C *&* ω *falls on* C^* *&* $C = C^*$.

Nor is there any evident way to model the relevant doxastic state within Kolmogorov's framework. Kolmogorov uses conditioning sub-σ-fields to model doxastic states. In any given case, one must choose a *single* conditioning sub-σ-field. Thus, there is no natural way for Kolmogorov to model situations where an agent coordinates distinct representational perspectives with one another. There is no natural way to model knowledge *that* $C = C^*$. This lacuna is an undeniable limitation of Kolmogorov's framework.

Kolmogorov can model a related situation: the agent knows *that* ω *falls on C &* ω *falls on C**. (Such knowledge may or may not be accompanied by knowledge *that* $C = C^*$.) Kolmogorov can cite the conditioning sub-σ-field σ(Φ, Ψ), i.e. the smallest sub-σ-field relative to which Φ and Ψ are both measurable. He can use σ(Φ, Ψ) to model an ideal agent's implicit knowledge upon learning the values of Φ and Ψ.

Let $P_{\sigma(\Phi,\Psi)}$ be an rcd for P given σ(Φ, Ψ). Since C has measure 0, the integral formula is satisfied no matter what values $P_{\sigma(\Phi,\Psi)}(A \mid .)$ assumes on points $\omega \in C$. One can choose $P_{\sigma(\Phi,\Psi)}$ so that $P_{\sigma(\Phi,\Psi)}(\,.\mid \omega)$ is any probability measure on points $\omega \in C$. Different choices yield different instructions for conditioning upon outcomes $\omega \in C$ as filtered through σ(Φ, Ψ). This is a special case of the indeterminacy noted in Sect. 8.1: Kolmogorov only dictates rcds up to set of measure 0.

The indeterminacy seems particularly recalcitrant for σ(Φ, Ψ). As noted in Sect. 8.1, we have a strong intuition that density $\frac{\cos\theta}{2}$ gives the rationally privileged result for conditioning on σ(Φ) and that a constant density over longitude gives the rationally privileged result for conditioning on σ(Ψ). A complete theory will likely privilege these respective rcds over rival candidates. We have no analogous intuitions for σ(Φ, Ψ). There is no clear reason to privilege one determinate choice among the infinitely many candidate rcds $P_{\sigma(\Phi,\Psi)}$. Specifically, either $\frac{\cos\theta}{2}$ or a constant density over longitude seems to yield a reasonable choice for $P_{\sigma(\Phi,\Psi)}(\,.\mid \omega)$ when $\omega \in C$.

In effect, we have landed back in the quandary from Sect. 1. The arc length solution and the surface area solution each engenders considerable conviction *when taken on its own*. Someone who hears just one solution typically finds it compelling. When one hears *both* solutions, confusion and indecision ensue. Each solution has considerable merit, yet neither seems sufficiently superior to command firm assent. We have found ourselves unable to resolve this quandary, despite deploying the vast resources of modern probability theory.

Does a determinate choice emerge if we stipulate more details about the agent modeled through probability space $(E, \mathscr{B}(E), P)$? Consider the following scenario:

(1) Fix coordinates (θ, φ) for the unit sphere, where arc C consists of the points with coordinate φ_0. An agent knows that numbers θ and φ were selected through a stochastic process, where repeated applications of that process would yield outcomes (θ, φ) uniformly distributed over the sphere's surface. She awaits news regarding which outcome was selected.

Since our agent knows that the θ-φ coordinate system plays a privileged role in selecting ω, one might argue that news formulated using θ-φ coordinates deserves privileged status over news formulated using other coordinates. In particular, one might argue that the rational policy is to prioritize news *that* $\Phi = \varphi_0$ over news *that* $\Psi = \psi_0$. So one might argue that the surface area solution specifies the rationally privileged response to news *that* $\Phi = \varphi_0$ & $\Psi = \psi_0$.

For the sake of argument, let us grant that the foregoing analysis is correct. By parity of reasoning, we should also grant that the arc length solution is rationally privileged if an agent learns *that* $\Phi = \varphi_0$ & $\Psi = \psi_0$ in the following scenario:

(2) Fix coordinates (ψ, γ) for the unit sphere, where arc C consists of the points with coordinate ψ_0. An agent knows that numbers ψ and γ were selected through a stochastic process, where repeated applications of that process would yield outcomes (ψ, γ) uniformly distributed over the sphere's surface.

Moreover, there are scenarios where no single coordinate system is known to play a privileged role in selecting ω:

(3) An agent awaits news regarding which location ω on the unit sphere's surface is my grandmother's favorite. The agent deems the likelihood of ω falling inside some region to be proportional to that region's surface area, but otherwise he has no relevant beliefs about my grandmother's preferences. Specifically, the agent knows nothing about the coordinate system (if any) through which my grandmother represents locations.

In scenario (3), neither the proposition that $\Phi = \varphi_0$ nor the proposition that $\Psi = \psi_0$ appears privileged. So neither the surface area solution nor the arc length solution seems like a rationally favored response to news *that* $\Phi = \varphi_0 \;\&\; \Psi = \psi_0$.

Scenarios (1)-(3) involve the same initial credences over $\mathcal{B}(E)$. Even if we grant that scenario (1) yields a determinate solution, scenario (2) then appears to yield a distinct determinate solution. And scenario (3) yields no evident determinate solution. Apparently, $(E, \mathcal{B}(E), P)$ in itself does not support a single determinate response to news *that* $\Phi = \varphi_0 \;\&\; \Psi = \psi_0$. By similar reasoning, $(E, \mathcal{B}(E), P)$ in itself does not support a single determinate response to news *that* $\Phi = \varphi_0 \;\&\; \Psi = \psi_0 \;\&(\forall\omega)(\Phi(\omega) = \varphi_0 \text{ iff } \Psi(\omega) = \psi_0)$.

§8.3 Living with indeterminacy

When you first encounter the Borel-Kolmogorov paradox, you feel pressure to seek the "right" answer. You want to ask: what are the true probabilities? From a subjectivist viewpoint, the question is how one should reallocate credence in light of various doxastic states. As I have argued, distinct doxastic states can represent the same null event. Epistemic norms may dictate different responses to distinct doxastic states that represent the same null event. There is nothing paradoxical or even surprising about this non-uniform treatment: a null event *as represented in different ways* can yield different rational credal assignments. In some cases (e.g. learning *that* $\Phi = \varphi_0$, or learning *that* $\Psi = \psi_0$), rationality seems to dictate determinate new credences. In other cases (e.g. learning *that* $\Phi = \varphi_0 \,\&\, \Psi = \psi_0$), rationality does not seem to dictate even semi-determinate new credences. The resulting indeterminacy is disconcerting but hardly paradoxical. Why should we expect that epistemic norms dictate even a semi-determinate credal assignment in response to all possible evidence? Rationality may sometimes allow wide discretion in credal reallocation. Current credences need not determine how to update credences in light of any possible news.

To motivate this viewpoint, let us consider a simple example that minimizes technical details.[13] Imagine a six-sided die colored as follows:

1 = *yellow* 2 = *black* 3 = *green* 4 = *red* 5 = *purple* 6 = *blue*

Imagine an agent who knows the color of each even side. She also knows that odd sides are colored yellow, green, and purple, but she does not know which side has which color. Her credence that a roll yields a certain outcome is given follows:

$P(1) = P(3) = P(5) = 0$ $P(2) = P(4) = P(6) = 1/3$

Our agent cannot use the ratio formula to define probabilities conditional on the event *the outcome is odd*. However, we may suppose that she has the following conditional probabilities:

$P(1 \mid odd) = 2/3$

$P(3 \mid odd) = P(5 \mid odd) = 1/6$

$P(2 \mid odd) = P(4 \mid odd) = P(6 \mid odd) = 0,$

[13] Thanks to an anonymous referee for suggesting that I discuss this example.

reflecting a policy for updating credences if she learns *that the outcome is odd*. Since she does not know the precise correlation between colors and odd numbers, those conditional probabilities seem consistent with the further conditional probabilities:

$P(green \mid odd) = 2/3$

$P(yellow \mid odd) = P(purple \mid odd) = 1/6$.

She represents the same side in two different ways (*the side labeled 1* vs. *the side colored yellow*). These distinct representational perspectives induce different conditional probabilities (2/3 vs. 1/6) regarding the same worldly event.

Now suppose our agent learns that the side labeled 1 *is* the side colored yellow. In this case, it is no longer rational to maintain both conditional probabilities

$P(1 \mid odd) = 2/3$

$P(yellow \mid odd) = 1/6$.

She must revise her conditional probabilities so that

$P(1 \mid odd) = P(yellow \mid odd)$.

But how? Should she choose 2/3, or 1/6, or some other value?

I think that the case *as described above* does not support a determinate answer. One must say more about the basis for the agent's initial conditional probabilities. Where do these conditional probabilities come from? Are they based on evidence? Depending on the details, there may or may not be a single determinate rational way for the agent to revise $P(1 \mid odd)$ and $P(yellow \mid odd)$. In particular, suppose the agent's initial conditional probabilities are mere ungrounded hunches lacking any rational basis. Then rationality does not seem to dictate any determinate way of revising $P(1 \mid odd)$ and $P(yellow \mid odd)$ once she learns that the side labeled 1 *is* the side colored yellow. Either 2/3 or 1/6 seems rationally permissible.

Readers may feel disappointed that my account yields no determinate solution in the die example. However, we should not require a theory to yield a determinate solution when a case is genuinely indeterminate. Do you really think that the stipulated facts determine a single rational value for *P*(1 | *odd*)? If so, which value? Why is that value privileged over alternative values?

Before the agent learns the precise correlation between odd numbers and colors, her unconditional probabilities appear to leave her conditional probabilities

P(1 | *odd*)	*P*(*yellow* | *odd*)
P(3 | *odd*)	*P*(*green* | *odd*)
P(5 | *odd*)	*P*(*purple* | *odd*)

unconstrained. Any numbers seem admissible, as long as the three numbers in each column sum to 1. This extreme permissiveness differentiates the die example from the sphere example, where the integral formula severely restricts $P_{\sigma(\Phi)}$, $P_{\sigma(\Psi)}$, and $P_{\sigma(\Phi,\ \Psi).}$ Unconditional probabilities constrain conditional probabilities far more in the sphere example than in the die example. Nevertheless, I have argued that unconditional probabilities in the sphere example leave *certain* conditional probabilities indeterminate. Rationality significantly underdetermines probabilities conditional upon knowledge *that* $\Phi = \varphi_0 \,\&\, \Psi = \psi_0$, or upon knowledge *that* $\Phi = \varphi_0 \,\&\, \Psi = \psi_0 \,\&(\forall\omega)(\Phi(\omega) = \varphi_0 \text{ iff } \Psi(\omega) = \psi_0)$.

In both the die example and the sphere example, current credences do not determine how one should reallocate credence in light of certain news. The indeterminacy arises for different reasons in the two cases:

- In the die example, we simply stipulate the conditional probabilities *P*(1 | *odd*) = 2/3 and *P*(*yellow* | *odd*) = 1/6. When the agent learns *that the outcome is odd & the side* *labeled 1 is the side colored yellow*, we are hard pressed to find a compelling argument favoring either 2/3 or 1/6 as rationally privileged.

- In the sphere example, a compelling argument depicts the arc length solution as the rationally privileged response to news *that* $\Phi = \varphi_0$. A second compelling argument depicts the arc length solution as the rationally privileged response to news *that* $\Psi = \psi_0$. The probability space (E, $\mathfrak{B}(E)$, P) supplies no principled basis for choosing between these two conflicting solutions in response to news *that* $\Phi = \varphi_0$ & $\Psi = \psi_0$, or in response to news *that* $\Phi = \varphi_0$ & $\Psi = \psi_0$ &$(\forall\omega)(\Phi(\omega) = \varphi_0$ *iff* $\Psi(\omega) = \psi_0)$.

Indeterminacy in the die example arises because neither solution recommends itself as particularly compelling, whereas indeterminacy in the sphere example arises because two distinct compelling solutions recommend themselves.

Admittedly, I have not conclusively established the alleged indeterminacy in either example. However, the past century of debate provides little encouragement that a determinate solution will emerge, especially for the sphere example. Do you really think that (E, $\mathfrak{B}(E)$, P) determines a single rationally correct rcd $P_{\sigma(\Phi, \Psi)}$? If so, which? Why is it privileged over alternative rcds? Those who hold that there is a single determinate solution must defend their position through persuasive argumentation. If there is no determinate solution, then my failure to offer a determinate solution is no defect.

In many familiar Frege cases, sufficient investigation can resolve conflicts that arise from contrasting representational perspectives on the same entity. If I believe *that Hesperus has craters* and *that Phosphorus does not have craters*, and if I subsequently discover *that Hesperus is Phosphorus*, then one hopes that sufficient astronomical investigation will suggest a determinate verdict regarding the presence of craters. In the sphere example, sufficient investigation may sometimes yield a determinate verdict between the arc length solution and surface area solution (e.g. if I discover that some fixed coordinate system played a privileged role in selecting ω). Yet there is no reason to expect the same determinate solution in all scenarios. There is also no reason to expect that every scenario will yield a determinate solution, no matter how thoroughly I investigate. This indeterminacy does not rise to the level of paradox, but it seems puzzling. We would like a general theory that explains why conditional probabilities are sometimes determinate and sometimes indeterminate. I have not provided such a theory. Thus, I do not say that we can dispel all the mysteries surrounding the Borel-Kolmogorov paradox simply by treating it as a Frege case. I do say that treating it as a Frege case is an essential first step.

§9. Conclusion

The Borel-Kolmogorov paradox reveals important links between probability and intensionality --- links that mathematical and philosophical expositions neglect all too often. The links surface within Kolmogorov's framework as relativity to conditioning sub-σ-fields. Once we place the links in proper philosophical context, any vestige of contradiction evaporates. Nevertheless, numerous philosophical and mathematical difficulties persist. I highlighted some especially important difficulties in Sect. 8. Resolving these difficulties will require extensive further research, much of it straddling philosophy and mathematics. Eighty years later, Kolmogorov's theory of conditional probability remains an outstanding framework for future inquiry.

Acknowledgments. I am indebted to Tim Butzer, Kenny Easwaran, Greg Grandenberger, Colin Howson, Agustín Rayo, Frederick Paik Schoenberg, and two anonymous referees for this journal for helpful comments on this paper. I also thank Melanie Schoenberg for assistance in preparing Figures 1-7 and RJ Duran for creating the final Mathematica versions of the figures.

Works Cited

Arnold, B. C., and Roberston, C. A. 2002. "The Conditional Distribution of X given X = Y Can Be Almost Anything!", In *Advances on Theoretical and Methodological Aspects of Probability and Statistics,* ed. N. Balakrishnan. New York: Taylor and Francis.

Arntzenius, F., Elga, A., and Hawthorne, J. 2004. "Bayesianism, Infinite Decisions, and Binding." *Mind* 113:251–283.

Bangu, S. 2010. "On Bertrand's Paradox." *Analysis* 70: 30-35.

Bertrand, J. 1889. *Calculs des Probabilités*. Paris: Gauthier-Villars.

Billingsley, P. 1995. *Probability and Measure*. 3rd ed. New York: Wiley.

Borel, E. 1909/1956. *Elements of the Theory of Probability*. Trans. J. Freund. Englewood Cliffs: Prentice-Hall.

Burge, T. 2009. "Five Theses on *De Re* States and Attitudes." In *The Philosophy of David Kaplan*, eds. J. Almog and P. Leonardi. Oxford: Oxford University Press.

Chalmers, D. 2011. "Frege's Puzzle and the Objects of Credence." *Mind* 120: 587-365.

Davidson, D. 2004. *Problems of Rationality*. Oxford: Oxford University Press.

de Finetti, B. 1972. *Probability, Induction, and Statistics*. New York: Wiley.

Dubins, L. 1975. "Finitely Additive Conditional Probabilities, Conglomerability and Disintegrations." *Annals of Probability* 3: 89-99.

Easwaran, K. 2008. *The Foundations of Conditional Probability*. PhD dissertation, University of California, Berkeley. Ann Arbor: ProQuest/UMI. (Publication No. 3331592.)

---. 2011. "Varieties of Conditional Probability." In *Philosophy of Statistics,* eds. P. Bandyopadhyay and M. Forster. Burlington: Elsevier.

Field, H. 2001. *Truth and the Absence of Fact*. Oxford: Oxford University Press.

Gorroochurn, P. 2012. *Classic Problems of Probability*. Hoboken: Wiley.

Hájek, A. 2003. "What Conditional Probability Could Not Be." *Synthese* 137: 273-323.

---. 2011. "Conditional Probability." In *Philosophy of Statistics*, eds. P. Bandyopadhyay and M. Forster. Burlington: Elsevier.

Hill, B. 1980. "On Some Statistical Paradoxes and Non-conglomerability." *Trabajos de Estadistica de Investigacion Operativa* 31: 39-66.

Howson, C. 2014. "Finite Additivity, Another Lottery Paradox, and Conditionalization." *Synthese* 191: pp. 989-1012.

Jaynes, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Kadane, J., Schervish, M., and Seidenfeld, T. 1986. "Statistical Implications of Finitely Additive

Probability." In *Bayesian Inference and Decision Techniques,* eds. P. Goel and A. Zellner. Amsterdam: North-Holland.
---. 2001. "Improper Regular Conditional Distributions." *Annals of Probability* 29: 1612-1624.
Kolmogorov, A. N. 1933/1956. *Foundations of the Theory of Probability*. 2nd English ed. Trans. N. Morrison. New York: Chelsea.
Lindley, D. 1982. "The Bayesian Approach to Statistics." *Some Recent Advances in Statistics,* eds. J. T. de Oliveria and B. Epstein. New York: Academic Press.
---. 1997. "Some Comments on Bayes Factors." *Journal of Statistical Planning and Inference* 61: 181-189.
Myrvold, W. Forthcoming. "You Can't Always Get What You Want: Some Considerations Regarding Conditional Probabilities." *Erkenntnis*.
Pfanzagl, J. 1979. "Conditional Distributions as Derivatives." *Annals of Probability* 7: 1046-1050.
Popper, K. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
Proschan, M., and Presnell, B. 1998. "Expect the Unexpected from Conditional Expectation." *The American Statistician* 52: 248-252.
Rao, M. M. 2005. *Conditional Measures and Applications*. 2nd ed. Boca Raton: CRC Press.
Rényi, A. 1955. "On a New Axiomatic Theory of Probability." *Acta Mathematica Academiae Scientiarum Hungarica* 6: 285-335.
Shafer, G. 1985. "Conditional Probability." *International Statistical Review* 53: pp. 261-277.
Shaffer, M. 2007. "Bealer on the Autonomy of Philosophical and Scientific Knowledge." *Metaphilosophy* 38: pp.44-54.
---. 2012. *Counterfactuals and Scientific Realism*. New York: Palgrave-MacMillan.
Tjur, T. 1980. *Probability Based on Radon Measures*. New York: Wiley.
van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Clarendon Press.
von Mises, R. 1957. *Probability, Statistics, and Truth*. 2nd English ed. London: Allen and Unwin.
Wimsatt, W. 2007. *Re-Engineering Philosophy for Limited Beings*. Cambridge: Harvard University Press.

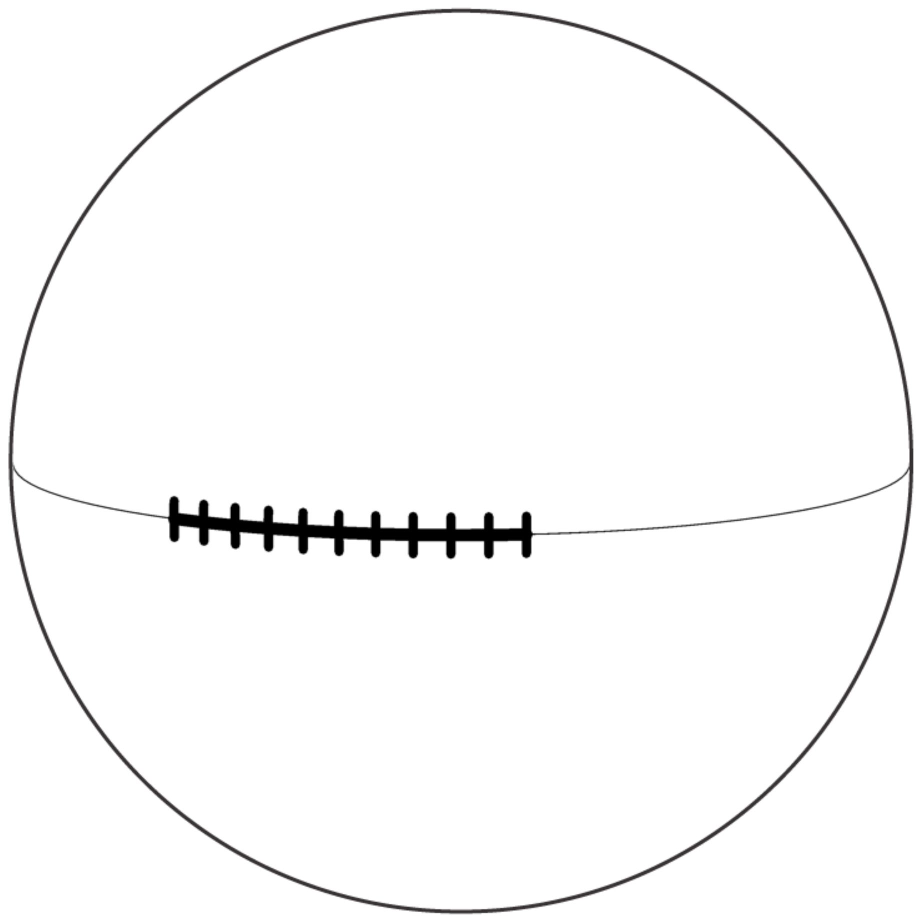

Figure 1. Arc *C* is the visible half of the equator. Arc *D* (indicated by hash marks) spans ¼ of *C*. The solution $P(D \mid C)$ = ¼ seems equally compelling no matter where *D* falls on *C*.

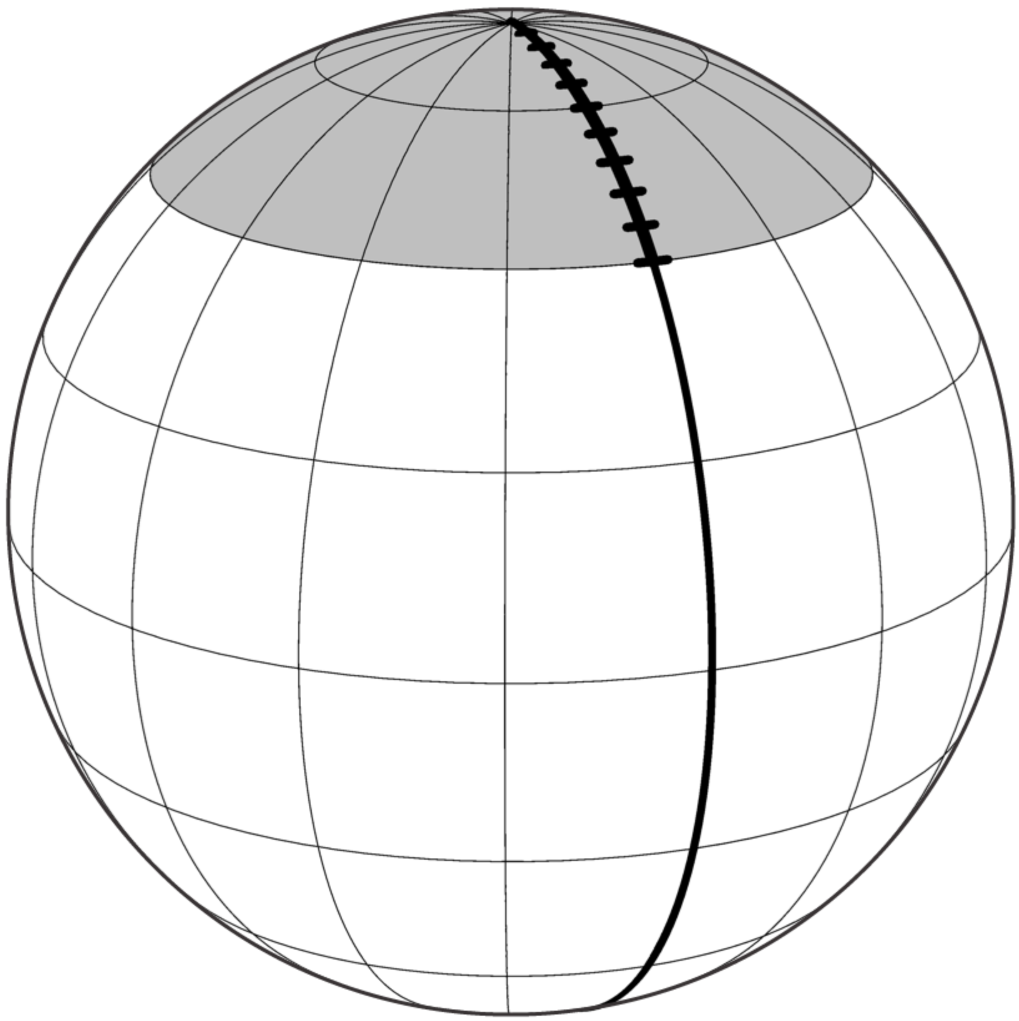

Figure 2. Arc *C* is the darkened meridian. Arc *D* (indicated by hash marks) extends halfway from the North Pole to the equator. In other words, it stretches from latitude 90° to latitude 45°. The grey polar cap contains all points with latitude ≥ 45°. The polar cap's surface area is less than ¼ the sphere's surface area. Thus, $P(\ \textit{latitude} \geq 45°\ ) < ¼$.

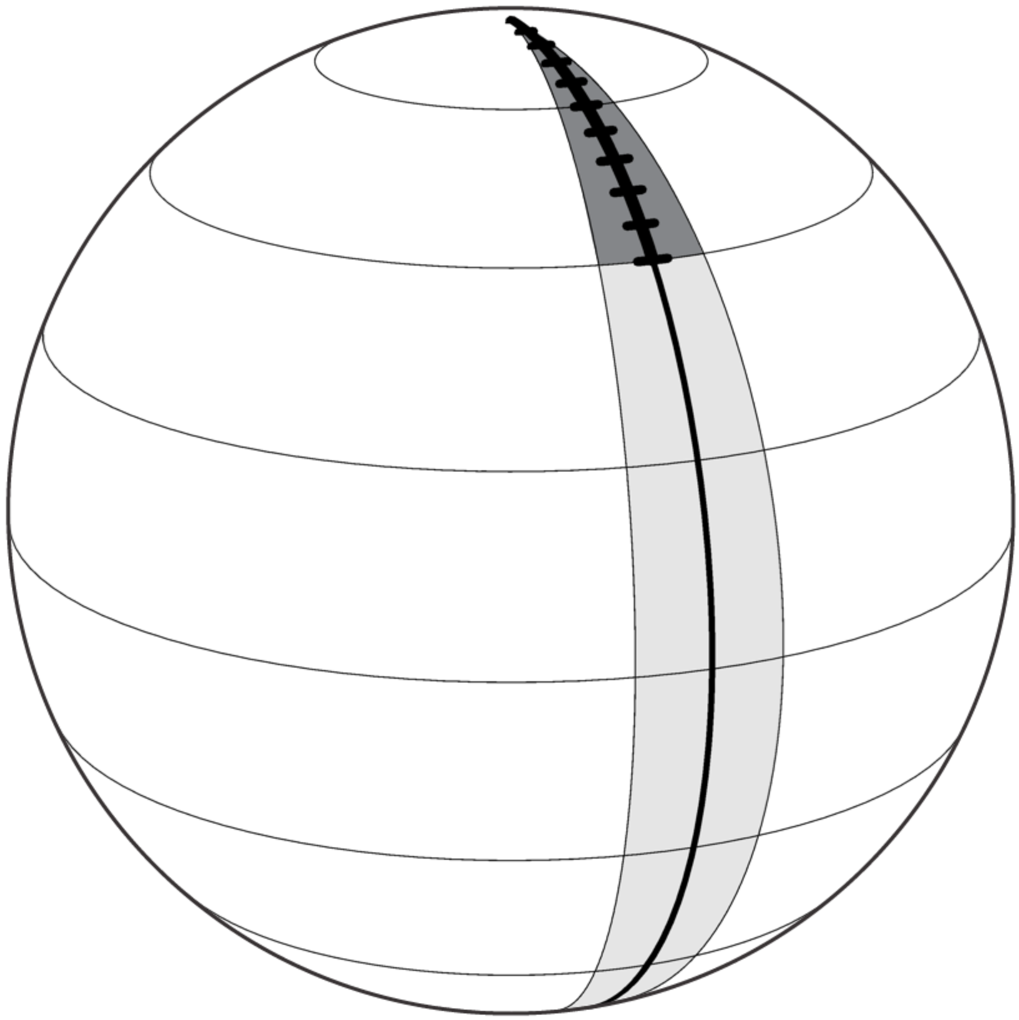

Figure 3. A bundle of meridians centered on meridian *C*. The dark grey region contains points with latitude ≥ 45°. Dividing the surface area of the bundle into the surface area of the dark grey region yields a value less than ¼. Thus, *P*(ω *has latitude* ≥ 45° | ω *falls in the bundle of meridians*) < ¼. In the general case, the bundle need not be centered on *C*.

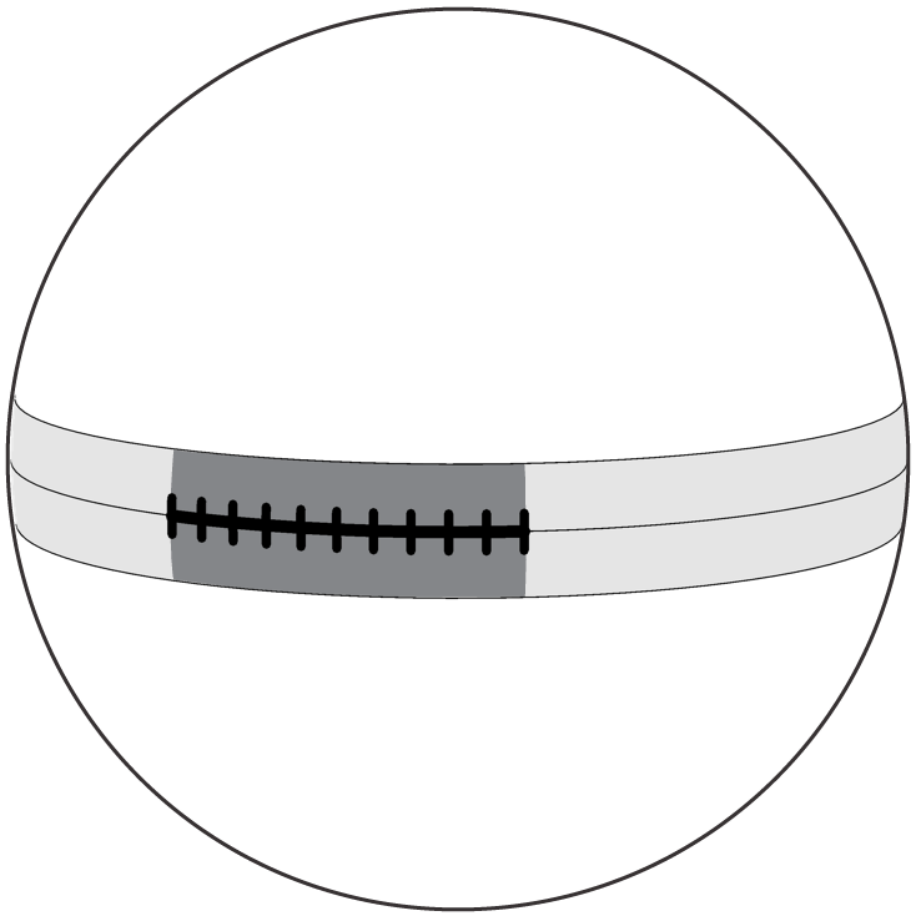

Figure 4. A bundle of half-parallels centered on the equator. The dark grey region is delimited by the two meridians that bound arc *D*. Dividing the surface area of the bundle into the surface area of the dark grey region yields ¼. Thus, *P*(*ω falls between the two meridians* | *ω falls in the bundle of half-parallels*) = ¼. We obtain the same answer no matter where *D* falls along the equator. In the general case, the bundle need not be centered on the equator.

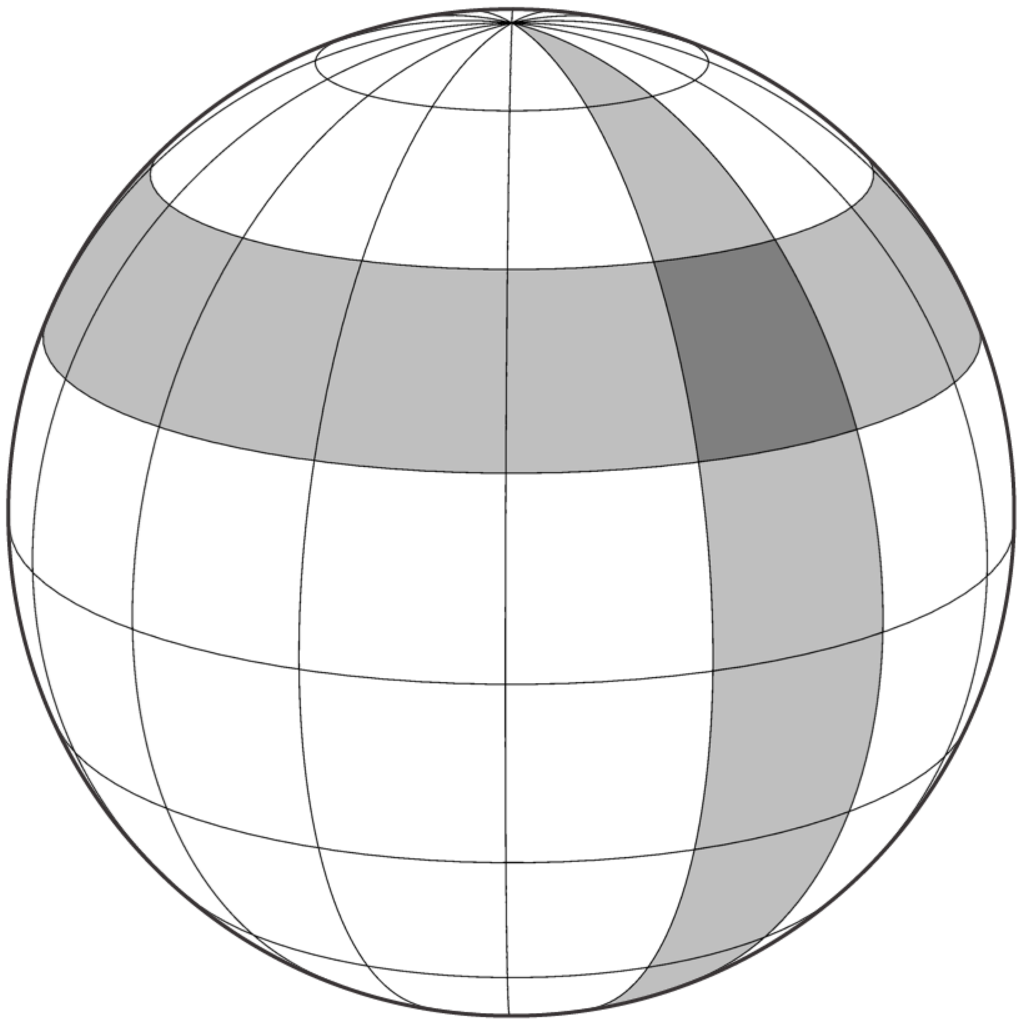

Figure 5. The dark grey tile is a (magnified!) surface area element. Tiles grow smaller as one moves away from the equator, but their size remains constant as one rotates around the polar axis.

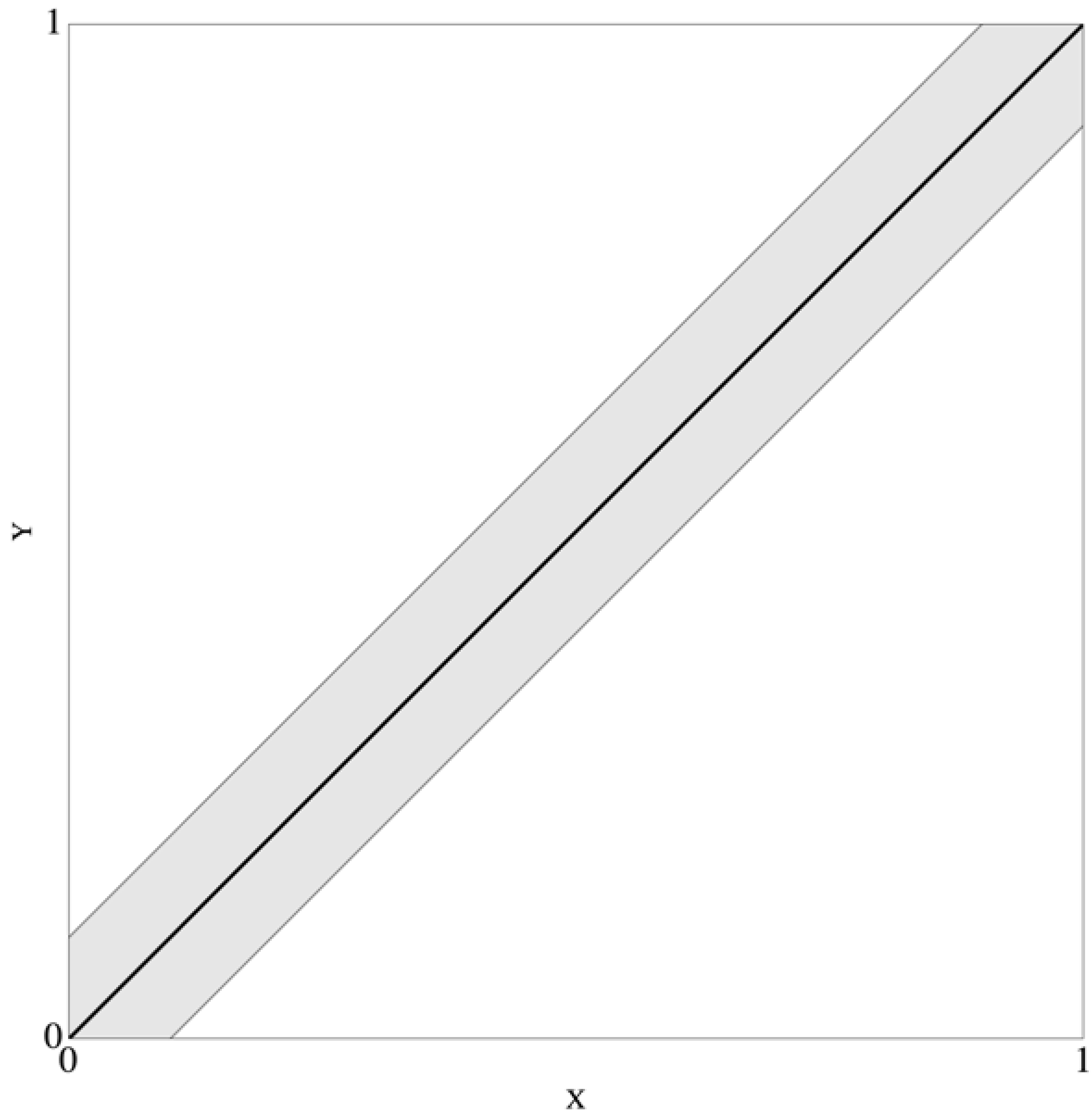

Figure 6. The diagonal line contains points such that $X = Y$, i.e. $V = 0$. The grey region contains points such that $|V| \leq \varepsilon$, for small ε. Points with $X > ½$ occupy half the area of the light grey region. Thus, $P(X > ½ \mid |V| \leq \varepsilon) = ½$. Taking the limit as $\varepsilon \rightarrow 0$ yields $|V| \rightarrow 0$, which suggests that $P(X > ½ \mid X = Y) = P(X > ½ \mid V = 0) = ½$.

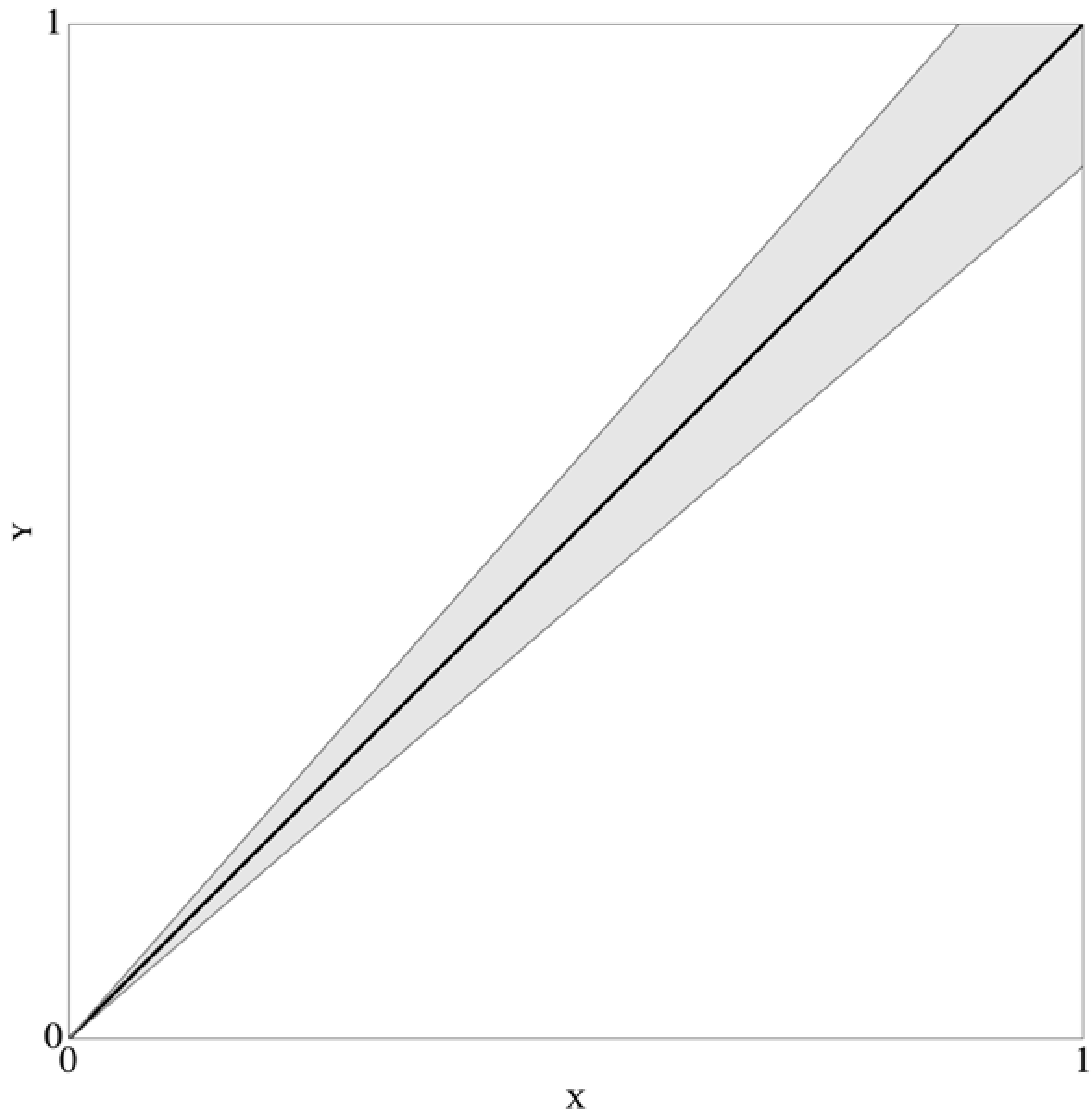

Figure 7. The diagonal line contains points such that $X = Y$, i.e. $W = 1$. The grey region contains points such that $|W–1| \leq \varepsilon$, for small ε. Points with $X > ½$ occupy much more than half the area of the grey region. Thus, $P(X > ½ \mid |W–1| \leq \varepsilon)$ is substantially greater than ½. Taking the limit as $\varepsilon \to 0$ yields $W \to 1$, which suggests that $P(X > ½ \mid X = Y) = P(X > ½ \mid W = 1) > ½$.