

June 6, June 22, August 29, 2006

New abstract, Nov. 1, 2006

Minor rev. Dec 5 2006

Ignorance and Indifference

John D. Norton

Center for Philosophy of Science and

Department of History and Philosophy of Science

University of Pittsburgh

Pittsburgh PA 15260

www.pitt.edu/~jdnorton

The epistemic state of complete ignorance is not a probability distribution. In it, we assign the same, unique ignorance degree of belief to any contingent outcome and each of its contingent, disjunctive parts. That this is the appropriate way to represent complete ignorance is established by two instruments, each individually strong enough to identify this state. They are the principle of indifference (“PI”) and the notion that ignorance is invariant under certain redescriptions of the outcome space, here developed into the “principle of invariance of ignorance” (“PII”). Both instruments are so innocuous as almost to be platitudes. Yet the literature in probabilistic epistemology has misdiagnosed them as paradoxical or defective since they generate inconsistencies when conjoined with the assumption that an epistemic state must be a probability distribution. To underscore the need to drop this assumption, I express PII in its most defensible form as relating symmetric descriptions and show that paradoxes still arise if we assume the ignorance state to be a probability distribution. By separating out the different properties that characterize a probability measure, I show that the ignorance state is incompatible with each of the additivity and the dynamics of Bayesian conditionalization of the probability calculus.

1. Introduction

In one ideal, a logic of induction would provide us with a belief state representing total ignorance that would evolve towards different belief states as new evidence is learned. That the Bayesian system cannot be such a logic follows from well-known, elementary considerations. In familiar paradoxes to be discussed here, the notion that indifference over outcomes requires equality of probability rapidly leads to contradictions. If our initial ignorance is sufficiently great, there are so many ways to be indifferent that that the resulting equalities contradict the additivity of the probability calculus. We can properly assign equal probabilities in a prior probability distribution only if our ignorance is not complete and we know enough to be able to identify which is the right partition of the outcome space over which to exercise indifference. Interpreting zero probability as ignorance also fails multiply. Additivity precludes ignorance on all outcomes, since the sum of probabilities over a partition must be unity; and the dynamics of Bayesian conditionalization makes it impossible to recover from ignorance. Once an outcome is assigned a zero prior, its posterior is also always zero. Thus it is hard to see that any prior can properly be called an “ignorance prior,” to use the term favored by Jaynes (2003, Ch. 12), but is at best a “partial ignorance prior.” For these reasons the growing use of terms like “noninformative priors,” “reference priors” or, most clearly “priors constructed by some formal rule” (Kass and Wasserman, 1996) is a welcome development.

What of the hope that we may identify an ignorance belief state worthy of the name? Must we forgo it and limit our inductive logics to states of greater or lesser ignorance only? The central idea of this paper is that, *if we forgo the idea that belief states must be probability distributions*, then there is a unique, well-defined ignorance state and the project of this paper is to identify it.

The instruments more than sufficient to specify this state already exist in the literature and are described in Section 2. They are the familiar principle of indifference (“PI”) and also the notion that ignorance states can be specified by invariance conditions. I will argue, however, that common uses of invariance conditions do not employ them in their most secure form. The most defensible invariance requirements use perfect symmetries and these are governed by what I shall call the “principle of the invariance of ignorance” (PII). In Section 3, I will review the familiar paradoxes associated with PI, identifying the strongest form of the paradoxes as those associated with competing but otherwise perfectly symmetric descriptions. I will also argue that

invariance conditions are beset by paradoxes analogous to those troubling PI. They arise even in the most secure confines of PII since simple problems can exhibit multiple competing symmetries, each generating a different invariance condition.

In Section 4, I will argue that we have misdiagnosed these paradoxes as some kind of deficiency of the principle of indifference or an inapplicability of invariance conditions. Rather, they are such innocuous principles of evidence as to be near platitudes. They both derive from the notion that beliefs must be grounded in reasons and, in the absence of distinguishing reasons, there should be no difference of belief. How could we ever doubt the notion that, if we have no grounds at all to pick between two outcomes, then we should hold the same belief for each? The aura of paradox that surrounds the principles is an illusion created by our imposing the additional and incompatible assumption that an ignorance state must be a probability distribution. In the remaining sections, it will be shown that these instruments identify a unique, epistemic state of ignorance that is not a probability distribution.

Section 5 describes the weaker theoretical context in which this ignorance state can be defined. It is based on a notion of non-numerical degrees of confirmation that may be compared qualitatively; and it may be selectively enriched to bring it closer to the full probability calculus. In Section 6, we shall see that implicit in the paradoxes of indifference is the notion that the state of ignorance is unchanged under disjunctive coarsening or refinement of the outcome space; and that this same state is invariant under a transformation that exchanges propositions with their negations. These two conditions each pick out the same ignorance state, in which a unique ignorance degree is assigned to all contingent propositions. In particular, in that state, we assign the same ignorance state of belief to all contingent propositions and each of their contingent, disjunctive parts. We shall see in Section 7 that this state is incompatible with both the additivity of the degrees of belief and also with the property that allows for Bayesian conditionalization, when these properties are separated out by algebraic means, so that any logic of induction that employs this state can use neither property. Section 8 contains some concluding remarks.

2. Instruments for Defining the State of Ignorance

The present literature in probabilistic epistemology has identified two principles that can govern the distribution of belief. They are both based on the simple notion that beliefs must be grounded in reasons, so that when there are no differences in reasons there should be no

differences in belief. Applying this notion to different outcomes gives us the Principle of Indifference (Section 2.1); and applying it to two perfectly symmetric descriptions of the same outcome space gives us what I call the Principle of Invariance of Ignorance (Section 2.2).

2.1 Principle of Indifference

The “principle of indifference” was named by Keynes (1921, Ch.IV) to codify a notion long established in writings on probability theory. I will express it in a form independent of probability measures.¹

(PI) *Principle of Indifference*. If we are indifferent among several outcomes, that is, if we have no grounds for preferring one over any other, then we assign equal belief to each.

Applications of the principle are familiar. In cases of finitely many outcomes, such as the throwing of a die, we assign equal probabilities of 1/6 to each of the 6 outcomes. If the outcomes form a continuum, such as the selection of a real magnitude between 1 and 2, we assign a uniform probability distribution.

2.2 Principle of Invariance of Ignorance

A second, powerful notion has been developed and exploited by Jeffeys (1961, Ch.III) and Jaynes (2003, Ch. 12). The leading idea is that a state of ignorance can remain unchanged when we redescribe the outcomes; that is, there can be an invariance of ignorance under redescription. That invariance may powerfully constrain and even fix the belief distribution. Jaynes (2003, pp. 39-40) uses this idea to derive the principle of indifference as applied to probability measures over an outcome space with finitely many mutually exclusive and exhaustive outcomes A_1, A_2, \dots, A_n . If we are really ignorant over which outcome obtains, our

¹ For completeness, I mention that this principle is purely *epistemic*. It is to be contrasted with an *ontic* symmetry principle, according to which outcomes A, B, C, ... are assigned equal weights if, for every fact that favors A, there are corresponding facts favoring B, C, ...; and similarly for B, C, In the familiar cases of die throws and dart tosses, it is this physical symmetry that more reliably governs the assigning of probabilities.

distribution of belief would be unchanged if we were to permute the labels A_1, A_2, \dots, A_n in any arbitrary way:

$$A_{\pi(i)}' = A_i \quad (1)$$

where $(\pi(1), \pi(2), \dots, \pi(n))$ is a permutation of $(1, 2, \dots, n)$. A probability measure P that remains unchanged under all these permutations must satisfy² $P(A_1) = P(A_2) = \dots = P(A_n)$. If the outcomes A_i are mutually exclusive and exhaust the outcome space, then the measure is unique: $P(A_i) = 1/n$, for $i = 1, \dots, n$. This is the equality of belief called for by PI.

This example illustrates a principle that I shall call:

(PII) *Principle of the Invariance of Ignorance*. An epistemic state of ignorance is invariant under a transformation that relates symmetric descriptions.

The new and essential restriction is the limitation to “symmetric descriptions,” which, loosely speaking are ones that cannot be distinguished other than through notational conventions. More precisely, *symmetric descriptions* are defined here as pairs of descriptions meeting two conditions:

(S1) The two describe exactly the same physical possibilities; and each description can be generated from the other by a relabeling of terms, such as the additional or removal of primes, or the switching of words.

An example is the permutation of labels of (1) above and a second is found below in (2a), (2b).

(S2) The transformation that relates the two descriptions is “self-inverting.” That is, the same transformation takes us from the first description to the second, as from the second to the first.

An example is the permutation that merely exchanges two labels; a second exchange of the same pair takes us back from the second description to the first.

This principle is the most secure way of using invariance to fix belief distributions. What makes it so secure is the insistence on the perfect symmetry of the descriptions. That defeats any

² The simplest way to arrive at this result is to consider a transformation that merely exchanges two labels, A_i and A_k , say, for i and k unequal. If the probability measure is to remain unchanged under all such exchanges, then we must have $P(A_i) = P(A_k)$ for each pair i, k , which entails the equality stated.

attempt to find reasons upon which to base a difference in the distribution of belief in the two cases; for any feature of one description will, under the symmetry, assuredly be found in the second. So any difference in the two epistemic states cannot be grounded in reasons, but must reflect an arbitrary stipulation. We shall see, however, that common invocations of invariance conditions in the literature do not adhere strictly to this symmetry in the transformations and are thus less secure.

If the outcomes form a continuum, the application of PII is identical in spirit to the deduction of the principle of indifference, though slightly more complicated. A clear illustration that gives the template for computing other cases is provided by applying PII to von Mises' (1951, pp. 77-78) celebrated case of wine and water. We are given a glass with some unknown mixture of water and wine and know only that the mixtures lies somewhere between 1:2 parts water to wine and 2:1 parts water to wine. That is,

$$\text{the ratio of water to wine } x \text{ lies in the interval } 1/2 \text{ to } 2; \quad (2a)$$

$$\text{and the ratio of wine to water } x' = 1/x \text{ also lies in the interval } 1/2 \text{ to } 2. \quad (2b)$$

If we represent our uncertainty over x with the probability density $p(x)$ and our uncertainty over x' with the probability density $p'(x')$, the idea that our ignorance is unchanged by redescription turns out to fully specify both densities. The calculation that shows this has two parts. First we note that the transformation from x to x' merely redescribes the same outcome, so the two should agree in assigning the same probabilities to the same outcomes. The outcome of x being in the small interval x to $x+dx$ is the same outcome as x' lying in x' to $x'+dx'$, where $x' = 1/x$. Since the two outcomes must agree in probabilities we have:³

$$p(x')dx' = -p(x)dx$$

That is, more precisely,

$$\text{A. Agreement in probability} \quad p'(x') = -p(x) dx/dx' \quad (3a)$$

In the second part, we note that there is a perfect symmetry between the two descriptions (2a) and (2b). Loosely speaking, that means that whatever our ignorance may be of the ratio of water to wine, it is just the same as our ignorance of the ratio of wine to water. Indeed had we mistakenly switched the labels in (2a) and (2b) it would make no difference to the problem posed. Formally that is expressed in the two descriptions (2a) and (2b) meeting the conditions

³ The negative sign arises since the increments dx and dx' increase in opposite directions.

(S1) and (S2) above. The first condition (S1) is met in that (2a) becomes (2b) if we switch the words “water” and “wine” and replace the variable x by x' ; and (2a) and (2b) still describe exactly the same outcome space.⁴ Condition (S2) is met since x relates to x' in exactly the same way as x' relates to x . That is, the function that transforms x to x' is exactly the same as the function that transforms x' to x ; they are both the taking of the arithmetic inverse

$$x' = 1/x \quad x = 1/x' \quad (4)$$

In other words, they are self-inverting, since composing the transformation with itself yields the identity map.

We have complete symmetry of descriptions. So PII requires that the two probability distributions are the same:

$$\text{B. Symmetry} \quad p'(\cdot) = p(\cdot) \quad (3b)$$

Since $dx/dx' = -x^2$, the system of equations (3a), (3b) and (4) entail that any $p(x)$ must satisfy the functional equation

$$p(1/x) (1/x) = p(x) x \quad (5)$$

Notably, solutions of (5) do *not* include $p(x) = \text{constant}$. The most familiar solution is⁵

$$p(x) = K/x \quad (5a)$$

where the requirement that $p(x)$ normalize to unity fixes $K = 1/\ln 4$.

The example of wine and water cleanly embodies the symmetry of descriptions needed to trigger the requirements of PII. Other familiar cases of symmetry appear to be a little less symmetric in so far as the transformations between the descriptions are not self-inverting. Take for example the redescription of all the reals by unit translation:

$$x' = x - 1 \quad x = x' + 1 \quad (6)$$

The transformation is not self-inverting so the perfect symmetry of descriptions fails in that we proceed from the x description to x' by adding unity; and from the x' to x by subtracting unity.

⁴ This symmetry can easily fail as it did in Von Mises' original presentation. He took the ratio to lie in 1:1 to 2:1, so that permuting “wine” and “water” and replacing x with x' does not lead to a description of the same outcome space.

⁵ Briefly, arbitrarily many solutions can be constructed by stipulating $p(x)$ for $1 \leq x < 2$ and using (5) to define $p(x)$ for $1/2 < x \leq 1$, where the resulting function may need to be multiplied by a constant to ensure normalization to unity.

However, in this example, self-inverting transformations are readily found. The transformation (6) can be created by composing two transformations that are individually self-inverting:

$$x'' = -x \quad x = -x'' \quad (6a)$$

$$x' = -1 - x'' \quad x'' = -1 - x' \quad (6b)$$

So the invariance of epistemic state required by PII obtains for each transformation (6a) and (6b) individually and thus also for (6) as a whole. (This is analogous to the decomposition of an arbitrary permutation (1) into a sequence of many pair-wise exchanges, each of which is self-inverting.)

Not all deductions of prior probabilities in the objective probability literature conform to the strict and most defensible conditions of PII, a perfect symmetry of descriptions. The most familiar use of invariance in the objective probability literature that lacks this symmetry is the deduction of the Jeffreys prior for a real-valued physical magnitude x that can take any value greater than zero. Jaynes' method requires that our prior probability distribution $p(x)$ be unaffected by changes in the units used. For example, Jaynes (2003, p. 382) requires that the prior probability distribution $p(t)$ for a time constant t must be unchanged by a change in the units used to measure time. That is, if we measure the constant in different units $t' = qt$, for q a constant of unit conversion, we recover a new prior distribution $p'(t')$ that must be unchanged. This gives us two conditions analogous to those of the wine and water problem

$$\text{A. Agreement in probability} \quad p'(t') = p(t) dt/dt' \quad (7a)$$

$$\text{B. Symmetry} \quad p'(\cdot) = p(\cdot) \quad (7b)$$

Given that q can be any real, these two equations admit a unique solution⁶

$$p(t) = \text{constant}/t \quad (8)$$

which is the Jeffreys prior.

The weakness of this deduction is that the condition B. Symmetry is not deduced from a perfect symmetry of the two descriptions. Rather it arises from something a little weaker and more nebulous. Jaynes writes of the two hypothetical experimenters using the different systems of units: "But Mr. X and Mr.X' are both completely ignorant and they are in the same state of

⁶ Since $dt'/dt=q$, the two equations entail $p(qt) q = p(t)$. Holding t fixed and differentiating with respect to q we have $p(qt) + qt dp(qt)/d(qt) = 0$; that is, $dp(t')/dt' = -p(t')/t'$. whose unique solution is the Jeffrey's prior (8).

knowledge, and so [p] and [p'] must be the same function..." What Jaynes says here is wrong. Mr. X and Mr. X' may know very little. But they do know how their units relate. If t is measured in minutes and t' in seconds, so that $q=60$, then there is the following asymmetry, knowable to both: Mr. X's measurement will be 1/60th that of Mr. X'. Switching t and t' does not leave everything unchanged, as it did when we switched x and x' in (2a) and (2b). So they are *not* "in the same state of knowledge." Jaynes' plausible presumption is that this is not enough asymmetry to overturn (7b) and perhaps he is right. However that reliance on plausibility falls short of the condition of perfect symmetry of descriptions needed by PII and routinely used in physics, Jaynes' model science, to underwrite such invariances.

These two principles express platitudes of evidence whose acceptance seems irresistible. They follow directly from the simple idea that we must have reasons for our beliefs. So if no reasons distinguish among outcomes, we must assign equal belief to them; or if two descriptions of the outcomes are exactly the same in every non-cosmetic aspect, then we must distribute beliefs alike in each.⁷ Yet, as I will now review, both principles lead to paradoxes in the ordinary, probabilistic context.

3. Their Failure if Belief States are Probability Distributions

3.1 Paradoxes of Indifference

Since at least time of Keynes' baptism (1921, Ch.IV) of the principle of indifference, it has been traditional to besiege the principle in paradoxes. Indeed they have become a fixture in

⁷ One may think that subjective Bayesians can escape the platitudes since they allow prior probability distributions to be set arbitrarily in so far as the evidence fails to fix them. Since that is the only constraint on the priors, the resulting freedom in the *range* of admissible priors already conforms to both PI and PII. If no evidence picks outcomes A_1 and A_2 apart, then for any admissible prior that favors A_1 in some way, there must be an admissible counterpart that favors A_2 in the same way. Thus, if one's beliefs are expressed by the range of admissible priors, they treat indifferent outcomes alike, as PI requires. Similarly this range must respect the perfect symmetry of descriptions of PII. In the wine-water example, for any admissible prior over the parameter x, there must be a corresponding admissible prior over the alternative parameter x'.

the routine, now nearly ritualized dismissals of the classical interpretations of probability.⁸ Laplace (1825, p.4) famously defined probability as the ratio of favorable to all cases among “equally possible cases,” which he defined as “cases whose existence we are equally uncertain of.” While Laplace seems to have taken the notion of an equally possible case as a primitive term not in need of further definition within his system, the later literature insists that on pain of circularity, he must provide an independent specification of which are the equally possible cases.⁹ The principle of indifference is pressed into service and the outcome is shown to be paradoxical.¹⁰

The paradoxes reveal that the principle of indifference only partially captures the notion of indifference presumed in the literature. All the paradoxes have the same structure. We are given some outcomes over which we are indifferent and thus assign equal probability. The outcomes are redescribed; indifference once again dictates that the outcomes are of equal probability; and it turns out the two assignments of probability contradict. Thus, it is also routinely assumed that the indifference described by the principle can persist when we redescribe the outcome space in certain ways, to be specified below.

The paradoxes divide into two classes. The first employs a finite outcome space and often a very small one of mutually exclusive outcomes. In one of Keynes’ (1921, Ch. IV) examples, the unknown country of a man is one of

France, Ireland, Great Britain

so, by the principle of indifference, we assign a probability $1/3$ to each. We can coarsen the same space by forming a disjunctive outcome “British Isles,” to arrive at

France, British Isles (=Great Britain v Ireland)

⁸ For surveys of these paradoxes both in the context of the classical interpretation and otherwise, see Galavotti (2005, § 3.2), Gillies (2000, pp. 37-49), Howson and Urbach (1996, pp. 59-62) and Van Fraassen (1989, Ch. 12)

⁹ This insistence has seems unfair to me. On pain of circularity or infinite regress, we cannot demand explicit definitions for all our terms. Some may be introduced by implicit definitions, that is by assertions in which they figure in some essential way.

¹⁰ It has also been suggested that the logical interpretation of probability is troubled by these paradoxes. (Gillies, 2000, p. 37)

By the principle of indifference we now assign the probability of 1/2 to each. We have now assigned both probability 1/3 and 1/2 to France. In another example, we do not know whether the color of the cover of a book is

red, blue, black

and assign probability 1/3 to each. However we also have the coarsened space

red, -red = (blue or black)

to whose outcomes we assign the incompatible probability of 1/2. The additivity of the probability calculus clearly drives the paradoxes. Since it requires that probabilities sum to unity, it precludes leaving the probabilities of the two outcomes red and -red of the coarsened space at 1/3, as the first space requires. Alternatively, additivity forces the probability of -red to be the sum of the probabilities of blue and black, so that it cannot equal the probability assigned to red in the first space.

This analysis can be continued with Keynes' other examples, as well as those offered elsewhere. Cramer (1966, pp. 15-16) describes a game in which two coins are tossed and player A wins if there is at least one head.¹¹ On one analysis, we are indifferent over the outcomes

HH, HT, TH, TT,

¹¹ This example is equivalent to Keynes' example of two cards, each drawn from different packs. One card proves to be black so by an application of the principle of indifference we can infer that the probability that the other is black is 1/2. The inference requires that we count the disjunctive outcome of (black-red or red-black) as a single outcome and assign it a probability equal to the only remaining outcome of black-black. A similar treatment of disjunctive outcomes lies at the heart of the mathematically much more complicated urn sampling problem discussed by Keynes (1921, pp. 49-50) arising in discussions of Laplace's rule of succession, with Howson and Urbach (1996, pp. 55-59) laying out the detailed computations. An urn contains n balls that may each be black or white. We may be indifferent over each of the possible numbers of white balls in the urn, in which case each possible number has probability $1/(n+1)$. Or we may be indifferent over each constitution, that is each of the 2^n assignments of black and white to each ball in succession. That there are exactly $n/2$ white balls (for n even), is a single outcome in the first case. In the second, it is a disjunction of $n!/((n/2)!(n/2)!)$ constitutions, each of which is a distinct outcome.

where “HT” represent the outcome of a head and then a tail. In another analysis, the first two outcomes are replaced by a single H = (HH v HT) since if a head is tossed on the first throw, player A has won and the second coin will not be tossed. So now we are to be indifferent over the three outcomes of the coarsened space

$$H = (HH \vee HT), \quad TH, \quad TT,$$

which, by analogous reasoning, yields incompatible probability assignments.

This second class of paradox employs a continuous outcome space, indexed by a continuous parameter, and the redescription arises through a manipulation of this parameter. We shall see, however, that all the paradoxes still employ the same notion that indifference can persist through a coarsening of the outcome space. This second class of paradox is often associated with so-called “geometrical probabilities” (Borel, 1950, Ch.7) since these cases commonly arise in geometry. The *locus classicus* is Bertrand’s¹² (1907, Ch. 1) discussion of a series of examples. In the simplest, we are to pick a real number at random between 0 and 100. We are indifferent to the number being in either half of this interval, 0 to 50 or 50 to 100, so we assign probability 1/2 to each. Our outcome space is

$$0 \text{ to } 50, \quad 50 \text{ to } 100$$

and we assign probability 1/2 to each. If however we consider the squares of these numbers, then we can divide the corresponding interval of squares of the chosen number, 0 to 10,000 into four equal parts, to create a refined outcome space (the inverse of coarsening) in which the original outcome of 50 to 100 is divided into three parts

$$0 \text{ to } 2500 (= 0^2 \text{ to } 50^2), \quad 2500 \text{ to } 5000, \quad 5000 \text{ to } 7500, \quad 7500 \text{ to } 10,000.$$

By indifference, we assign a probability 1/4 to each. The contradiction is immediate. The interval 0 to 50 was first assigned a probability 1/2. That same interval corresponds to the interval of squares 0 to 2500, to which we now assign the incompatible probability 1/4.

The basic idea of this construction can be replicated in many ways. Van Fraassen (1989, pp. 302-303) describes a conversion of it into a perfect cube factory that produces iron cubes

¹² Bertrand seemed to make no connection with something like the not yet named principle of indifference. Rather he took his constructions merely to illustrate the danger of analyses that involve the infinities associated with continuous magnitudes. The existence of incompatible outcomes was for him simply evidence that the original problem was badly posed.

with sides between 0 and 2 cm. If we are indifferent to the length of the side, then we assign probability $1/2$ to the outcome that one cube has a side with a length between 0 and 1 cm. For the 1cm sided cube lies midway in the interval of side lengths of 0 to 2 cm. If we are indifferent to the surface area of the cubes, we end up assigning a probability $1/4$ to this outcome. For side of area 1 cm^2 lies $1/4$ th into the interval of side areas of 0 to $2^2=4 \text{ cm}^2$. If we are indifferent to the volume of the cube, we end up assigning a probability $1/8$ to this outcome. For the cube of volume 1 cm^3 lies $1/8$ th into the interval of volumes of 0 to $2^3=8 \text{ cm}^3$.

Bertrand himself gave a series of more complicated examples. The most discussed was the problem of deciding whether a randomly chosen chord of a circle is greater in length than the side of an inscribed equilateral triangle. Bertrand gave three constructions that yielded probabilities of $1/3$, $1/2$ and $1/4$. In another example, he asked after the distance between two points chosen at random on a sphere. The probabilities of various distances differ according to whether the choosing is uniformly distributed on a great circle or uniformly over the area of the sphere.

These many multiplications of the one construction share a defect. They all depend on the assumption that it is appropriate to distribute beliefs indifferently in both the original description and the new coarsened or refined description. Perhaps that is just a mistake. Perhaps, as Gillies (2000, p. 46) notes, it is appropriate to exercise the principle of indifference in just one but not the other description, because of some difference intrinsic to the first description. The side, the area and the volume of a cube are all mathematically different in their properties and perhaps something about those differences dictate that we should be indifferent only over the length of the side. Sentiments such that these presumably lay behind Borel's (1950, pp. 81-83) response to Bertrand's problem of selecting two points on a sphere. Borel essentially insisted that selecting a point at random on a sphere must mean that the selection is uniformly distributed over the sphere's surface area.

It seems far-fetched to me to imagine that we may find some property of one description that warrants us exercising indifference only over it. And if there is some reason that privileges one description over another, do the paradoxes not return if we are so ignorant that we do not know it? Nonetheless, the threat remains: we may imagine that we can find some distinctive property of one description that licenses the use of an accordingly embellished principle of indifference, where that distinctive property would not be present in the other description.

This threat to the cogency of the paradox can be defeated fully by means of ideas introduced in the discussion of PII. Von Mises' wine-water problem (above) provides a traditional illustration of the paradoxes of indifference. When we are indifferent to x , the ratio of water to wine, we distribute probability uniformly over x . So we are indifferent to the three outcomes with x in the ranges:

$$x = 1/2 \text{ to } 1, \quad x = 1 \text{ to } 3/2, \quad x = 3/2 \text{ to } 2$$

Under the redescription in terms of x' , the ratio of wine to water, we are indifferent to the outcomes

$$x' = 2 \text{ to } 3/2, \quad x' = 3/2 \text{ to } 1, \quad x' = 1 \text{ to } 1/2$$

The paradox follows. We assign probability $1/3$ to outcome $x'=1$ to $1/2$, although it corresponds to the disjunctive outcome $(x=1 \text{ to } 3/2) \vee (x=3/2 \text{ to } 2)$, each of whose disjuncts is also assigned probability $1/3$.

What is distinctive in this formulation of the wine-water problem is that, as described above, there is a perfect symmetry between the two descriptions that includes the transformations used to relate them, which are self-inverting. So there can be no intrinsic difference between the two for an embellished version of the principle of indifference to exploit. Any distinctive feature that the principle may call up from one will, by the symmetry, assuredly be found in the other.

Indeed the example is even more troublesome for the principle of indifference. Exactly because of the perfect symmetry of the two descriptions, we should assign the same distributions of belief in each. The class of probability densities that respect the symmetry was specified by (5). It turns out not to contain the uniform distribution licensed by the principle of indifference. That is, the question is not to decide to which of the two descriptions the principle of indifference may be applied. Rather it turns out that the principle is applicable to neither. This last example is the strongest form of the paradoxes of indifference.

What we learn from this little survey of the paradoxes of indifference is a fact about routine expectations in the literature: the indifference expressed by the principle of indifference persists through coarsenings and refinements of the outcome space. While this expectation of

persistence may be challenged in many of the familiar paradoxes as a way of defeating the paradoxes,¹³ we can find symmetric cases in which it cannot be challenged.

3.2 Paradoxes of Invariance

While it is not generally recognized, it turns out that the use of invariance conditions as a way of specifying probability distributions is beset by paradoxes akin to those that trouble the principle of indifference. The paradoxes of indifference arose because greater ignorance gave us more partitions of the outcome space over which to exercise indifference. Each new partition corresponds to a new mathematical constraint on our probability measure. They quickly combine to produce a contradiction. The same thing happens with invariance conditions. Each ignorance is associated with an invariance. Thus the greater our ignorance, the more invariances we expect and the greater threat that these compounding invariances impose contradictory requirements on our probability distribution.

It takes only a little exploration to realize this threat. It is easily seen in a simple example that sufficient ignorance forces contradictory invariance requirements. Consider some magnitude x whose values lies in the open interval $(0,1)$; and that is all we know about it. What is our prior probability density $p(x)$ for x ? The problem remains unchanged if we reparameterize the magnitude, retaining the range of parameter values in $(0,1)$. In order to have the most defensible restrictions, let us consider only self-inverting transformations as reparameterizations, so that we have a perfect symmetry between the two descriptions and PII applies. The simplest of these is just

$$x' = 1 - x \tag{9a}$$

One may imagine that such self-inverting transformations are rare. They are not and very many can be found. Loosely speaking, as will be seen from the construction below, they are about as

¹³ Keynes' (1921, pp. 52-64) efforts to eliminate the paradoxes of indifference depend upon blocking the application of the principle to systems connected by disjunctive coarsening or refinement. Similarly Kass and Wasserman (1996, §3.1) propose that the paradoxes be avoided "practically" in that we use "scientific judgment to choose a particular level of refinement that is meaningful for the problem at hand."

common in the world of functions on reals as are symmetric functions.¹⁴ The self-inverting (9a) is the $\lambda=\infty$ member of the λ -indexed family of self-inverting transformations¹⁵

$$x' = f_\lambda(x) = \lambda - \sqrt{\lambda^2 + (\lambda - 1)^2 - (\lambda - x)^2} \quad (9)$$

displayed graphically in Figure 1. Another simple member is $\lambda=1$, which is

$$x' = f_1(x) = 1 - \sqrt{2x - x^2} \quad (9b)$$

¹⁴ The device used to generate the paradoxes of indifference, the rescaling of variables, is not sufficient to generate competing invariances, such as are needed to generate the paradoxes of invariance. That is, let $p(x)$ and $p'(x')$ satisfy the conditions (10a) and (10b) and rescale the variables to $X=f(x)$ and $X'=f(x')$, where both rescalings are effected by the same monotonic function $f(\cdot)$. Two new probability distributions are induced by

$$P(X) = p(x) dx/dX \quad P'(X') = p'(x') dx'/dX'$$

It follows immediately that $P(\cdot)$ and $P'(\cdot)$ are the same functions since $p(\cdot)$ and $p'(\cdot)$ are the same functions; and dx/dX is the same function of X as dx'/dX' is of X' . That is, the induced distributions P and P' will satisfy B. Symmetry, and, by their construction, A. Agreement in probability as well.

¹⁵ One can affirm that (9) is self-inverting by directly expanding $f_\lambda(f_\lambda(x))$ to recover x or by noting that $x' = f_\lambda(x)$ is equivalent to

$$(\lambda - x')^2 + (\lambda - x)^2 = \lambda^2 + (\lambda - 1)^2 \quad (a)$$

The expression for $x' = f_\lambda(x)$ is found by solving (a) for x' in terms of x ; and the expression for $x = f_\lambda^{-1}(x')$ by solving (a) for x in terms of x' . The two functions recovered must be the same since x and x' enter symmetrically into (a). The properties of (9) are more easily comprehended geometrically. As equation (a) indicates, each curve in the graph is simply the arc of a circle with a center at $(x', x) = (\lambda, \lambda)$ and radius R satisfying $R^2 = \lambda^2 + (\lambda - 1)^2$.

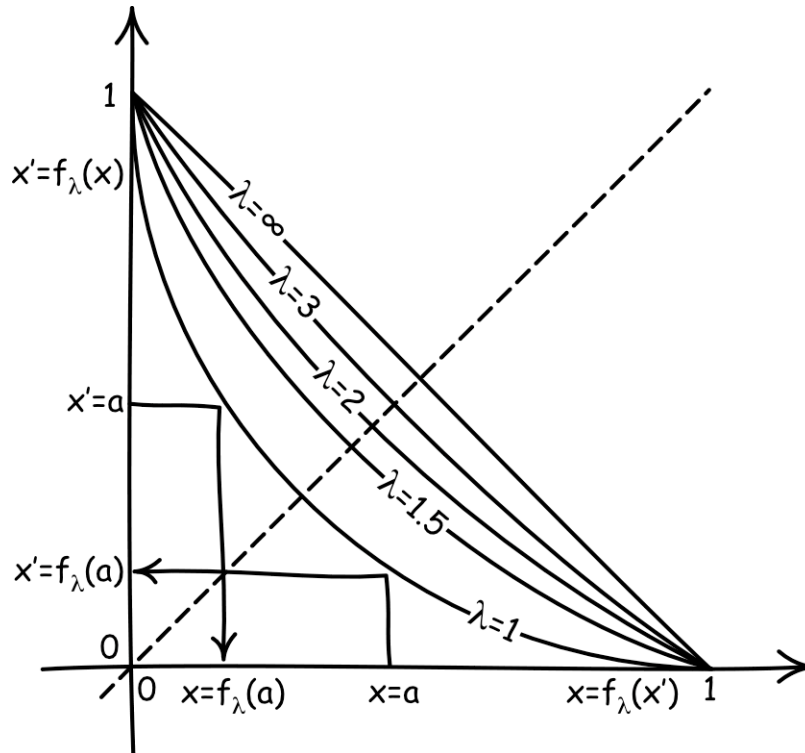


Figure 1. A Family of Self-Inverting Transformations

That these functions are self-inverting manifests as a symmetry of the graph about the $x'=x$ diagonal, represented as a dashed line in Figure 1. Indeed any function with this symmetry property will be self-inverting. It ensures that if $x=a$ is transformed to $x'=f_\lambda(a)$, then $x'=a$ will also be transformed to the same $x=f_\lambda(a)$ by the inverse transformation, as shown in the figure.¹⁶

¹⁶ It is more cumbersome but possible to generate self-inverting functions by algebraic manipulation. To generate such a function with the domain $[L,H]$, where $L<H$ and L may be $-\infty$ and H may be ∞ , select some K such that $L<K<H$. Now pick any strictly decreasing function $x'=g_h(x)$, defined for $K\leq x\leq H$, the “h” high part of the domain, such that $g_h(K)=K$ and $g_h(H)=L$. Since g_h is strictly decreasing, it is invertible. Define $g_l(x)$ for $L\leq x\leq K$, the “l” low portion of the domain, by stipulating for x in this domain that $g_l(x) = g_h^{-1}(x)$. The self-inverting function $f(x)$ is defined by $f(x)=g_l(x)$, for $L\leq x\leq K$, and $f(x)=g_h(x)$, for $K\leq x\leq H$. Even if g_h and g_l are arbitrarily differentiable, the combined f may fail to be differentiable at the join $x=K$. Arbitrary differentiability may be assured for f , such as in (9), by suitable selection of the derivatives of g_h at $x=K$ and $x=H$.

The probability distribution $p(x)$ must remain invariant under each transformation f_λ , according to PII, for each yields a perfectly symmetric redescription of the problem. We can see, however, that no $p(x)$ can exhibit this degree of invariance. To see this, we do not need all the properties of the family of self-inverting transformations f_λ . We need only assume that it has two members, h_1 and h_2 , say, such that $h_1(x) < h_2(x)$ for all $0 < x < 1$. That is trivially achieved using any pair of members of f_λ , since, for fixed x in $(0,1)$, $f_\lambda(x)$ is strictly increasing in λ . As before, PII requires that $p(x)$ and its transform $p'(x')$ under a self-inverting transformation $x' = f_\lambda(x)$ must satisfy

$$\text{A. Agreement in probability} \quad p'(x') = -p(x) dx/dx' \quad (10a)$$

$$\text{B. Symmetry} \quad p'(\cdot) = p(\cdot) \quad (10b)$$

Now select any X , such that $0 < X < 1$. For $x_1'(x) = h_1(x)$, we have from (10a) and (10b) that

$$\int_0^X p(x) dx = - \int_{x=0}^X p(x_1') \frac{dx_1'}{dx} dx = \int_{x_1'=x_1'(X)}^1 p(x_1') dx_1'$$

Similarly for $x_2'(x) = h_2(x)$, we have

$$\int_0^X p(x) dx = - \int_{x=0}^X p(x_2') \frac{dx_2'}{dx} dx = \int_{x_2'=x_2'(X)}^1 p(x_2') dx_2'$$

Subtracting and relabeling the variable of integration, we have

$$0 = \int_{x_1'(X)}^{x_2'(X)} p(y) dy \quad (11a)$$

Since $p(x) \geq 0$ for all x , it follows that¹⁷ $p(x) = 0$ for all $x_1'(X) < x < x_2'(X)$. Since X is chosen arbitrarily, we may select X so that any nominated value of x lies in the interval $(x_1'(X), x_2'(X))$.¹⁸ Hence it follows that $p(x) = 0$ for all $0 < x < 1$, which contradicts the requirement that a probability distribution normalize to unity.

¹⁷ Or, if $p(x)$ is discontinuous, it may differ from zero at most at isolated points, so that these non-zero values do not contribute to the integral. Therefore they are discounted, since they cannot contribute to a non-zero probability of an outcome.

¹⁸ Proof: For a given $0 < x < 1$, by we have from the initial supposition on h_1 and h_2 that $x_1'(x) < x_2'(x)$. A suitable X is any value $x_1'(x) < X < x_2'(x)$. For, from $x_1'(x) < X$, we have $x = x_1'(x_1'(x)) > x_1'(X)$, where we use the fact that x_1' is strictly decreasing in x . Similarly from $X <$

4. What Should We Learn from the Paradoxes?

The moral usually drawn from the paradoxes of indifference is a correct but short-sighted one: indifference cannot be used as a means of specifying probabilities in cases of extensive ignorance. That there are analogous paradoxes for invariance conditions is less widely recognized. They are indicated obliquely in Jaynes' work. He described (1973, §7) how he turned to the method of transformational invariance as a response to the paradoxes of indifference, exemplified in Bertrand's paradoxes. They allowed him to single out just one partition over which to invoke indifference, so that (to use the language of Bertrand's original writing) the problem becomes "well-posed." The core notion of the method was (§7):

Every circumstance left unspecified in the statement of a problem defines an invariance which the solution must have if there is to be any definite solution at all.

We saw above that this notion leads directly to new paradoxes if our ignorance is sufficiently great to yield excessive invariance. Jaynes reported (§8) that this problem arises in the case of von Mises' wine-water problem:

On the usual viewpoint, the problem is underdetermined; nothing tells us which quantity should be regarded as uniformly distributed. However, from the standpoint of the invariance group, it may be more useful to regard such problems as *overdetermined*; so many things are left unspecified that the invariance group is too large, and no solution can conform to it.

It thus appears that the "higher-level" problem of how to formulate statistical problems in such a way that they are neither underdetermined nor overdetermined may itself be capable of mathematical analysis. In the writer's opinion it is one of the major weaknesses of present statistical practice that we do not seem to know how to formulate statistical problems in this way, or even how to judge whether a given problem is well posed.

When the essential content of this 1973 paper was incorporated into Chapter 12 of Jaynes' (2003) final and definitive work, this frank admission of the difficulty no longer appeared, even

$x_2'(x)$, we have $x_2'(X) > x$. The function x_1' is strictly decreasing since it is invertible, $x_1'(0)=1$ and $x_1'(1)=0$; and similarly for x_2' .

though no solution had been found. Instead Jaynes (2003, pp. 381-82) sought to dismiss cases of great ignorance as too vague for analysis on the manifestly circular grounds that his methods were unable to provide a cogent analysis:

If we merely specify ‘complete initial ignorance’, we cannot hope to obtain any definite prior distribution, because such a statement is too vague to define any mathematically well-posed problem. We are defining this state of knowledge far more precisely if we can specify a set of operations which we recognize as transforming the problem into an equivalent one. Having found such set of operations, the basic desideratum of consistency then places nontrivial restrictions on the form of the prior.

My diagnosis, to be developed in the sections below, is that Jaynes was essentially correct in noting that invariance conditions may *overdetermine* an ignorance belief state. Indeed the principle of indifference also overdetermines such a state. In this regard, we shall see that both instruments are very effective at distinguishing a unique state of ignorance. The catch is that this state is not a probability distribution. Paradoxes only arise if we assume in addition that it must be. We thereby fail to see that PI and PII actually work exactly as they should.

5. A Weaker Structure

In order to establish that PI and PII do pick out a unique state of ignorance, we need a structure hospitable to non-probabilistic belief states. Elsewhere, drawing on an extensive literature in axiom systems for the probability calculus, I have described such a structure (Norton, forthcoming). For a precise synopsis of its content, see Appendix. Informally, its basic entity is $[A|B]$ is introduced through the properties of F . *Framework*. It represents the degree to which proposition B inductively supports proposition A, where these propositions are drawn from a (usually) finite set of propositions closed under the Boolean operations of \sim (negation), \vee (disjunction) and $\&$ (conjunction). The degrees are *not* assumed to be real valued. Rather it is only assumed that they form a partial order so that we can write $[A|B] \leq [C|D]$ and $[A|B] < [C|D]$. These comparison relations will be restricted by one further notion. Whatever else may happen, we do not expect that some proposition B can have less support than one of its disjunctive parts A on the same evidence. That is, we require *monotonicity*: if $A \Rightarrow B \Rightarrow C$, then $[A|C] \leq [B|C]$. This much of the structure will provide the background for the analysis to follow.

The structure has further notions that will prove incompatible with the ignorance state to be defined. The first, introduced through the property *A. Addition*, is an addition operator \oplus , which allows the combining of degrees of support of mutually contradictory propositions to yield the degree of support of their disjunction. It is the surrogate of the additivity of the probability calculus. The second, introduced through the property *B. Bayes Property*, is a multiplication operator \otimes , which allows for the distinctive dynamics of updating associated with Bayes' theorem. Both of these properties are compatible with the probability calculus and express essential elements of it.

The existence of these two operators is logically independent of one another; we can have systems that have either one without the other. We shall see, however, that the unique state of ignorance will be incompatible with each of them individually.

This structure is formulated in terms of “degrees of support.” On the supposition that we believe what we are warranted to believe, I will presume that our degrees of beliefs agree with these degrees of support.

6. Characterizing Ignorance

Once we dispense with the idea that a state of ignorance must be represented by a probability distribution, we can return to the ideas developed in the context of PI, PII and their paradoxes and deploy them without arriving at contradictions. We can discern two properties of a state of ignorance: invariance under disjunctive coarsenings and refinements; and invariance under negation. As we shall see below, each is sufficient to specify the state of ignorance fully and it turns out to be the same state: a single ignorance degree of belief “I” assigned to all contingent propositions in the outcome space.

6.1 Invariance under Disjunctive Coarsenings and Refinements

We saw in Section 3.1 above that the paradoxes of indifferences all depended upon a single idea: if our ignorance is sufficient, we may assign equal beliefs to all members of some partition of the outcome space and that equality persists through disjunctive coarsenings and refinements. This idea is explored here largely because of its wide acceptance in that literature. I have already indicated above in Section 3.1 that the idea is less defensible in cases in which there is not a complete symmetry between the two descriptions. We shall see in Section 6.2 below that

the same results about ignorance as derived here in Section 6.1 can be derived from PII using descriptions that are fully symmetric, related by self-inverting transformations.

Let us develop the idea of invariance of ignorance under disjunctive coarsenings and refinements. If we have an outcome space Ω partitioned into mutually contradictory propositions $\Omega = A_1 \vee A_2 \vee \dots \vee A_n$, an example of a disjunctive coarsening is the formation of the new partition of mutually contradictory propositions $\Omega = B_1 \vee B_2 \vee \dots \vee B_{n-1}$, where

$$B_1 = A_1, B_2 = A_2, \dots, B_{n-1} = A_{n-1} \vee A_n \quad (12)$$

All disjunctive coarsenings of A_1, A_2, \dots, A_n are produced by finitely many applications of this coarsening operation along with arbitrary permutations of the propositions, as defined by (1) above. A partition and its coarsening are each non-trivial if none of their propositions is Ω or \emptyset . The inverse of a coarsening is a refinement.

Assume that we have no grounds for preferring any of the members of the non-trivial partition of $\Omega = A_1 \vee A_2 \vee \dots \vee A_n$, then by PI we assign equal belief to each, and, by supposition, this ignorance degree of belief $[\emptyset|\Omega] < I < [\Omega|\Omega]$ is neither certainty nor complete disbelief:

$$[A_1|\Omega] = [A_2|\Omega] = \dots = [A_n|\Omega] = I \quad (13a)$$

Now consider the coarsening (12) and assume that we have no grounds for preferring any of the members. Once again there exists a possibly distinct ignorance degree of belief I' , neither certainty nor complete disbelief, such that

$$[B_1|\Omega] = [B_2|\Omega] = \dots = [B_{n-1}|\Omega] = I' \quad (13b)$$

Since from (12) $B_1 = A_1$, we have $[A_1|\Omega] = [B_1|\Omega]$ so that

$$I' = I$$

Since all coarsenings are produced by successive applications of (12) along with permutations, it follows that the one ignorance degree of belief I is unique. Finally, if C is a non-trivial disjunction of some proper subset of the $\{A_1, A_2, \dots, A_n\}$ —written $C = A_a \vee \dots \vee A_b$ —some sequence of coarsening and permutations allows us to infer that its degree of confirmation is I . That is, we infer the distinctive property of ignorance presumed in the paradoxes of indifference

$$[A_a|\Omega] = \dots = [A_b|\Omega] = [A_a \vee \dots \vee A_b|\Omega] = I \quad (14)$$

6.2 Invariance of Ignorance Under Negation

We saw in Section 2.2 above that one application of PII was Jaynes' deduction of the principle of indifference by requiring that a state of ignorance over a finite outcome space remains invariant under a permutation of the propositions. This same idea can also be applied to a transformation that switches propositions with their negations. If we are really ignorant over some outcome A , then our degree of belief in A would be unchanged if A had been somehow switched with its negation $\sim A$. A short parable may help clarify the transformation.

Let us imagine that we are to receive a message pertaining to the two compatible outcomes "land" and "sea" by means of a secret code that assigns numbers to each outcome. The code was devised by a very dedicated logician, so there are numbers for all possible 16 logical combinations of the outcomes. For greater security, we have two code books to choose from. Their values are (shh—don't tell!):

Code in Book A	Secret Message	Code in Book B
1	$\emptyset = \sim\Omega$	16
2	$\text{land} \ \& \ \text{sea} = \sim(\sim\text{land} \vee \sim\text{sea})$	15
3	$\sim\text{land} \ \& \ \text{sea} = \sim(\text{land} \vee \sim\text{sea})$	14
4	$\text{land} \ \& \ \sim\text{sea} = \sim(\sim\text{land} \vee \text{sea})$	13
5	$\sim\text{land} \ \& \ \sim\text{sea} = \sim(\text{land} \vee \text{sea})$	12
6	$\text{sea} = \sim(\sim\text{sea})$	11
7	$\text{land} = \sim(\sim\text{land})$	10
8	$(\text{land} \ \& \ \text{sea}) \vee (\sim\text{land} \ \& \ \sim\text{sea}) =$ $\sim((\text{land} \ \& \ \sim\text{sea}) \vee (\sim\text{land} \ \& \ \text{sea}))$	9
9	$(\text{land} \ \& \ \sim\text{sea}) \vee (\sim\text{land} \ \& \ \text{sea}) =$ $\sim((\text{land} \ \& \ \text{sea}) \vee (\sim\text{land} \ \& \ \sim\text{sea}))$	8
10	$\sim\text{land} = \sim(\text{land})$	7
11	$\sim\text{sea} = \sim(\text{sea})$	6
12	$\text{land} \vee \text{sea} = \sim(\sim\text{land} \ \& \ \sim\text{sea})$	5
13	$\sim\text{land} \vee \text{sea} = \sim(\sim\text{land} \ \& \ \text{sea})$	4
14	$\text{land} \vee \sim\text{sea} = \sim(\sim\text{land} \ \& \ \text{sea})$	3
15	$\sim\text{land} \vee \sim\text{sea} = \sim(\text{land} \ \& \ \text{sea})$	2
16	$\Omega = \sim\emptyset$	1

Table 1. Code Books Illustrating the Negation Map

Prior to its receipt we are in complete ignorance over which message may come and begin to contemplate how credible the content of each message may be. On the presumption that Book A is in use, we assign beliefs *not* over which message may come, but over the truth of the 16 possible messages. We must assign maximum belief to the content of code 16, since we know that Ω is necessarily true; we must assign minimum belief to the contents of code 1, since the contradiction \emptyset is always false; and we assign intermediate, ignorance degrees of belief to everything in between. In convenient symbols

$$[1|\Omega] = [\emptyset|\Omega] \quad [2|\Omega] = I_2 \quad [3|\Omega] = I_3 \quad \dots \quad [14|\Omega] = I_{14} \quad [15|\Omega] = I_{15} \quad [16|\Omega] = [\Omega|\Omega]$$

We now find that it is not Book A, but Book B that will be used. How will that affect our distribution of belief? We must exchange our degrees of belief for the message with codes 1 and 16, since code 1 now designate the necessarily true Ω and code 16 the necessarily false \emptyset . What of the remaining messages? We had assigned degree of belief I_6 to what we thought was the message “sea”. It now turns out to be the message “ \sim sea”. If our ignorance is sufficient, that will have no effect on our degree of belief. “sea”? “ \sim sea”? We just do not know! That is, we must assign equal degree of belief to each, so that $I_6 = I_{11}$. This analysis can be repeated for all the remaining outcomes 2 to 15. In each case, the switching of the codebooks has simply switched one message with its negation, as the table reveals. For example, under the Book A, code 14 designates “land \vee \sim sea.” Under Book B, code 14 designates its negation “ \sim land & sea” and the original “land \vee \sim sea” is designated by code 3. So, by analogous reasoning, $I_3 = I_{14}$. We can also infer that all four values are the same by using the property of *monotonicity* mentioned in Section 5 above.¹⁹ Continuing in this way, we can conclude that all intermediate degrees of ignorance have the same value $I = I_2 = \dots = I_{15}$. Rather than displaying the argument in all detail, it is sufficient to proceed to the general case, whose proof covers this special case.

An outcome space Ω consists of propositions C_1, C_2, \dots, C_m generated by closing a set of atomic propositions A_1, A_2, \dots, A_n , under the usual Boolean operators \sim, \vee and $\&$, taking note of the usual logical equivalences. The remapping of codebooks corresponds to the negation map N between set of proposition labels C_1, C_2, \dots, C_m and a duplicate set of proposition labels, C_1', C_2', \dots, C_m' , in which

$$\sim C_i' = N(C_i) \quad (15)$$

What is important about this negation map (15) is that it is self-inverting—the negation of a negation returns the original proposition (or, in this case, its label clone). Thus the sets C_i and C_i' are symmetric descriptions²⁰ and, if we are in ignorance over the outcomes, PII may be applied.

¹⁹ Since \sim land & sea \Rightarrow sea, we have $I_3 \leq I_6$; and since \sim sea \Rightarrow land \vee \sim sea, we have $I_{11} \leq I_{14}$.

Recalling $I_6 = I_{11}$ and $I_3 = I_{14}$, we must have $I_6 = I_{11} = I_3 = I_{14}$.

²⁰ This symmetry may not be evident immediately since the negation map (15) can take an atomic propositions (such as A_1) and map it to a disjunctive propositions (here $A_2 \vee \dots \vee A_n$) and

The analysis proceeds with the two-component calculation already shown in Section 2.2. Since the map simply relabels the same outcomes, we have

$$\text{A. Agreement in degrees of belief} \quad [\sim C_i' | \Omega'] = [C_i | \Omega]$$

But since we have a perfect symmetry in the two descriptions of the outcomes, we also have for the contingent propositions (i.e. those that are not always true or always false)²¹

$$\text{B. Symmetry} \quad [C_i' | \Omega'] = [C_i | \Omega]$$

It follows immediately from these two conditions A. and B. that $[\sim C_i' | \Omega'] = [C_i' | \Omega']$; or, re-expressed in the original proposition labels

$$[\sim C_i | \Omega] = [C_i | \Omega] = I_i \tag{16}$$

where this condition holds for every contingent proposition in the outcome space Ω . So far, we cannot preclude that the ignorance degree I_i defined is unique for each distinct pair of contingent propositions $C_i, \sim C_i$. As before, *monotonicity* allows us to infer that all I_i are equal to a common value I . To see this, take any two contingent propositions C and D . There are two cases:

(I) $C \Rightarrow D$ or $D \Rightarrow C$ or $\sim C \Rightarrow D$ or $C \Rightarrow \sim D$. Since they are the same under relabeling, assume $C \Rightarrow D$, so that $\sim D \Rightarrow \sim C$. We have from *monotonicity* that $[C | \Omega] \leq [D | \Omega]$ and $[\sim D | \Omega] \leq [\sim C | \Omega]$. But we have from (16) that $[\sim C | \Omega] = [C | \Omega]$ and $[\sim D | \Omega] = [D | \Omega]$.

Combining, it now follows that $[\sim C | \Omega] = [C | \Omega] = [\sim D | \Omega] = [D | \Omega]$.

(II) Neither $C \Rightarrow D$ nor $D \Rightarrow C$ nor $\sim C \Rightarrow D$ nor $C \Rightarrow \sim D$. This can only happen when $C \& D$ is not \emptyset . Since $C \& D \Rightarrow D$, we can repeat the analysis of (I) to infer that $[\sim(C \& D) | \Omega] =$

conversely; whereas our earlier examples of self-inverting maps (such as an exchange of two labels A_i and A_k) mapped atomic propositions to atomic propositions. This greater complexity does not compromise the facts that C_i and C'_i label the same set of propositions and that the map between them is self-inverting, which is all that is needed for the symmetry.

²¹ Restricting B. Symmetry to contingent propositions only is really a stipulation on the type of ignorance being characterized. We are assuming that the ignorance does not extend to logical truths, such as Ω , and logical falsities, such as \emptyset . Without the restriction, we would recover a more extensive ignorance state in which we would be uncertain even over logical truths and falsities. There is no contradiction in such a state, but it is of lesser interest, since knowledge of logical truths can at least in principle be had without calling upon external evidence.

$[C \& D | \Omega] = [\sim D | \Omega] = [D | \Omega]$. Similarly, we have $C \& D \Rightarrow C$, so that $[\sim(C \& D) | \Omega] = [C \& D | \Omega] = [\sim C | \Omega] = [C | \Omega]$. Combining we have the result sought: $[\sim C | \Omega] = [C | \Omega] = [\sim D | \Omega] = [D | \Omega]$.

We have now used PII to infer that, in cases of ignorance, every contingent proposition in the outcome space Ω must be assigned the same ignorance value I . This is the same result as arrived at in Section 6.1 above by means of the idea that the ignorance state must be invariant under disjunctive coarsenings and refinements. Thus we affirm that both approaches lead us to a unique state of ignorance, in which all contingent propositions are assigned the same ignorance value I .

There is a more formal way of understanding the generation of this unique ignorance state from the condition of invariance under negation. Elsewhere (manuscript), I have investigated how the familiar duality of truth and falsity in a Boolean algebra may be extended to real-valued measures defined on the algebra. To each additive measure m , there is a dual additive measure M , defined by the dual map $M(A) = m(\sim A)$, for each proposition A in the algebra. The dual of an additive measure is itself additive but not in the familiar way. The familiar rule that allows us to add the measures of propositions (whose conjunction is \emptyset) when they are disjoined is replaced by an unfamiliar rule that allows us to add the dual measures of propositions (whose disjunction is Ω) when they are conjoined. Because additive measures and their duals obey different calculi, additive measures are not self-dual.

We arrive at the ignorance state for the case of real valued measures by the simple condition that the measure be self-dual in its contingent propositions. That is, it is a monotonic, real-valued measure²² m that remains invariant under the dual map $M(A) = m(\sim A)$, for contingent propositions A , so that it satisfies the defining condition $m_I(A) = m_I(\sim A)$. If we set by convention that the extreme values of this ignorance map are $m_I(\emptyset) = 0$ and $m_I(\Omega) = 1$, then it follows²³ that all contingent propositions must be assigned the same real value $m_I(A) = I$, where I is some fixed real, $0 < I < 1$. Clearly m_I is not additive.

²² A monotonic measure is defined analogously to the property of monotonicity (Appendix). If $A \Rightarrow B \Rightarrow C$, then $m(A|C) \leq m(B|C)$.

²³ Proof: set $m_I(C_i) = [C_i | \Omega] = I_i$. Then the uniqueness of I_i follows from the demonstration of the uniqueness of I_i of (16) and requires the monotonicity of the measure.

6.3 Comparing Ignorance Across Different Outcome Spaces

The arguments of Sections 6.2 and 6.3 establish that, for each outcome space, there exists a unique ignorance degree of belief for all contingent propositions in it. That leaves the possibility that this ignorance degree of belief is different for each distinct outcome space. It is a natural expectation that the same ignorance degree of belief can be found in all outcome spaces. Naturalness, however, is no substitute for demonstration. The difficulty in mounting a demonstration is that, if *B. Bayes Property* is not assumed, the framework set up in Section 5 is too impoverished to enable comparison of degrees of belief between different outcome spaces. Some further assumption is needed to enable the comparison. In this section, I will show that introducing a very weak notion of independence and assuming that it is occasionally instantiated is sufficient to allow us to infer that the same ignorance degree of belief arises in all finite outcome spaces.

Consider an outcome space “A” defined by the logical closure under Boolean operations of m mutually exclusive and exhaustive, contingent propositions A_1, A_2, \dots, A_m , so that $\Omega = A_1 \vee A_2 \vee \dots \vee A_m$. We shall assume that we are in a state of complete ignorance over A so that

$$I_A = [A_i | \Omega]_A$$

for $i = 1, \dots, m$. The subscripts A allow the possibility that the ignorance degree of belief and other degrees of belief are peculiar to the outcome space A . A second outcome space “B” is defined analogously with n propositions B_1, B_2, \dots, B_n , for which

$$I_B = [B_k | \Omega]_B$$

where $k = 1, \dots, n$. Finally we define a product outcome space “AB” as generated in the same way by the mn mutually exclusive and exhaustive propositions, $(A_1 \& B_1), (A_1 \& B_2), \dots, (A_m \& B_n)$. Degrees of belief relative to this product outcome space are designated by $[\cdot | \Omega]_{AB}$.

A very weak notion of independence of the two spaces is:

Weak independence of outcome spaces A and B. The degree of belief in some proposition of A is unaffected by the mere knowledge that outcomes in B are possible; and conversely. That is expressed by the condition

$$[A_i | \Omega]_A = [A_i | \Omega]_{AB} \quad \text{and} \quad [B_k | \Omega]_B = [B_k | \Omega]_{AB} \quad (17)$$

for all admissible i, k .

This condition is much weaker than the usual condition of probabilistic independence. In the latter, the probability assigned to an outcome of one space is unaffected when the outcome is conditionalized on the supposition that some outcome of the other space obtains. In (17) we conditionalize only on the knowledge that the other outcome space exists, not that one of its outcomes obtains.

An example illustrates the differing strengths. Nothing in the discussion above precludes the propositions B_1, B_2, \dots, B_n of the second outcome space being merely a permutation of the propositions A_1, A_2, \dots, A_m of the first outcome space. In that case, ordinary probabilistic independence between the two spaces would fail. However the weaker sense of (17) would still hold. To see that this weaker sense is not vacuous, imagine a second example in which the propositions of the second outcome space are $B_1=A_1, B_2=A_2, \dots, B_n=A_{m-1}$; that is, the second space allows everything in first but denies A_m . (In effect this is the space generated from the A outcome space by conditioning on $A_1 \vee \dots \vee A_{m-1}$.) Since the B but not A outcome space presumes A_m is false, we would not expect (17) to hold; for learning the range of possibility admitted by B supplies new information that can alter judgments of degrees of belief. Indeed relation (17) must fail in case $i=m$, for $[A_m|\Omega]_A=I_A$ but²⁴ $[A_m|\Omega]_{AB}=[\emptyset|\Omega]_{AB}$.

If outcome spaces A and B are independent in the sense of (17), then we can show that their ignorance degrees of belief are the same. First note that for this case, we must also have an ignorance distribution in the product space AB, with an ignorance degree I_{AB}

$$[A_i|\Omega]_{AB} = [B_k|\Omega]_{AB} = I_{AB}$$

for admissible i, k . From the weak notion of independence (17), we also have

$$I_A = [A_i|\Omega]_A = [A_i|\Omega]_{AB} \quad \text{and} \quad I_B = [B_k|\Omega]_B = [B_k|\Omega]_{AB}$$

Combining we have

$$I_{AB} = I_A = I_B \tag{18}$$

so that the ignorance degree of belief in two independent outcome spaces is the same.

²⁴ In the outcome space AB, A_m is represented by the disjunction $A_m = (A_m \& B_1) \vee \dots \vee (A_m \& B_n) = (A_m \& A_1) \vee \dots \vee (A_m \& A_{m-1}) = \emptyset \vee \dots \vee \emptyset = \emptyset$.

This last conclusion is enough to enable us to conclude the uniqueness of the ignorance degree of belief for all outcome spaces, even ones that are not independent in the sense of (17). To see this, imagine that the outcome spaces A and B are not independent and that their ignorance degrees of belief are I_A and I_B . We need only assume that there exists a third outcome space, C, that is independent from each of A and B in the sense of (17), with ignorance degree of belief I_C . It now follows from (18) that $I_C = I_A$ and $I_C = I_B$. Combining them, we have $I_A = I_B$, which establishes the equality of the degrees of ignorance for any two outcome spaces A and B.

7. Incompatibility with Degrees of Belief as Probability Measures

It is evident that the assigning of the same ignorance degree of belief to all contingent propositions in the outcome space is incompatible with the distribution of belief being a probability measure. What we shall see in this section is that this incompatibility runs quite deep. It contradicts two, important, independent aspects of the probability calculus.

7.1 Incompatibility with A. Addition

The surrogate here for the additivity of the probability calculus is the existence of an addition operator \oplus with properties A. *Addition* defined in the Appendix. It is incompatible with the ignorance state defined in Section 6. To see this, assume that we have such a state on an outcome space Ω that includes two mutually contradictory propositions A and B. For them we have

$$[A \vee B|\Omega] = [A|\Omega] = [B|\Omega] = I$$

where I is distinct from $[\Omega|\Omega]$ and $[\emptyset|\Omega]$. If, for purposes of contradiction, we presume that we also have an addition operator defined on the space, that operator relates these degrees as

$$[A \vee B|\Omega] = [A|\Omega] \oplus [B|\Omega]$$

where $[A|\Omega] \oplus [B|\Omega]$ is strictly increasing, and thus invertible, in both arguments. Combining, we have

$$[A|\Omega] \oplus [B|\Omega] = [A|\Omega] = [B|\Omega] = I$$

That is, the ignorance degree I acts as a zero of the operator in the restricted sense that composing it with itself leaves it unchanged

$$[A|\Omega] \oplus I = I \oplus [A|\Omega] = [A|\Omega]$$

However, since $A \vee \emptyset = \emptyset \vee A = A$, the minimal value $[\emptyset|\Omega]$ also is a zero of \oplus in the more general sense that composing it with anything leaves the latter unchanged:

$$[A|\Omega] \oplus [\emptyset|\Omega] = [\emptyset|\Omega] \oplus [A|\Omega] = [A \vee \emptyset|\Omega] = [A|\Omega]$$

Comparing the last two expressions and recalling that \oplus is invertible in both arguments, we conclude that

$$I = [\emptyset|\Omega]$$

That is an unacceptable outcome, for $[\emptyset|\Omega]$ is the minimum degree of belief assigned to the contradiction \emptyset that we are certain is false. The incompatibility is thereby established.

That the ignorance state is incompatible with A . *Addition* was to be expected on quite general grounds. The presence of this operator amounts to a particular interpretation for the degrees in the system: as they range from high to low, they span justification of complete belief to complete disbelief. As is described in greater detail in Norton (forthcoming, Section 4.1), this follows from a reciprocal relationship between belief and disbelief: high belief in A entails high disbelief in $\sim A$, and conversely. (More or less ignorance concerning A is not reciprocally related to less or more ignorance in $\sim A$.) If one translates this reciprocal relationship into functional terms, the existence of the \oplus operator follows.

7.2. Incompatibility with B. Bayes Property

B. Bayes Property, with its multiplication operator \otimes , gives the system the characteristic dynamics of conditionalization associated with Bayes' theorem. From it, for hypothesis H and evidence E , we infer a version of Bayes' theorem

$$[H\&E|\Omega] = [H|E] \otimes [E|\Omega] = [E|H] \otimes [H|\Omega]$$

Since the operator \otimes is almost²⁵ always strictly increasing and invertible in both arguments, the usual dependencies follow. The posterior $[H|E]$ increases when, other terms fixed, the expectedness $[E|\Omega]$ decreases; or the likelihood $[E|H]$ increases; or the prior $[H|\Omega]$ increases.

An ignorance state on some outcome space Ω is incompatible with *B. Bayes Property*, as long as we also require that it be possible for there to remain some uncertainty after we

²⁵ For example, when $[E|H] = [\emptyset|H]$, $[E|H] \otimes [H|\Omega]$ is not strictly increasing in $[H|\Omega]$. $[H|E] \otimes [E|\Omega]$ is undefined when $[E|\Omega] = [\emptyset|\Omega]$; and similarly for $[E|H] \otimes [H|\Omega]$.

conditionalize in a state of ignorance. To see this, assume that we have two contingent propositions A and B in Ω where $A \Rightarrow B$ and *B. Bayes Property* holds. We have

$$[A|\Omega] = [A|B] \otimes [B|\Omega]$$

Since $[A|\Omega] = [B|\Omega] = I$, we have

$$I = [A|B] \otimes I \tag{19}$$

Indeed, if we set $A=B$, this becomes

$$I = [A|A] \otimes I \tag{19a}$$

Since \otimes always invertible in the first argument, it follows from (19) that $[A|B]$ will be the same for all contingent A and B, where $A \Rightarrow B$. Worse, from (19a) we see that this unique value is certainty

$$[A|B] = [A|A] = [\Omega|\Omega]$$

That is an extraordinarily unintuitive result. It tells us that if we are completely ignorant over 1000 mutually exclusive outcomes A_1, \dots, A_{1000} but condition A_1 on the first 999, we become certain of A_1 since $[A_1|A_1 \vee \dots \vee A_{999}] = [\Omega|\Omega]$.

We may seek to avoid these outcomes by portraying the ignorance degree I as a zero of the operator \otimes so that (19) would not be invertible. Indeed that would have the very desirable outcome of enabling non-trivial conditionalization. Since (19) would now no longer be invertible, we could maintain ignorance degrees for both $[A|\Omega]$ and $[B|\Omega]$ and then be free to let other considerations fix the value of $[A|B]$.²⁶ Attractive as it is, this possibility contradicts *B. Bayes Property*, which does not admit a zero in the second position of the operator \otimes . The natural zero in $[A|B] \otimes [B|C]$ would be $[B|C] = [\emptyset|C]$. It is precluded since *B. Bayes Property* requires B to be “admissible,” that is, not of minimum measure. The analogous prohibition in probability theory is that if $A \Rightarrow B$ and $P(A)=P(B)=0$, then $P(A|B) = P(A)/P(B) = “0/0”$ is not defined.

²⁶ This possibility is realized in the “theory of random propositions,” described in Norton (1994, §7.3).

8. Conclusion

How should an epistemic state of ignorance be represented? My contention here is that we have long had the instruments that uniquely characterize it in the principle of indifference and the principle of invariance of ignorance. However our added assumption that epistemic states must also be probability distributions has led to contradictions that we have misdiagnosed as arising from some deficiency in the two principles.

There are other proposals for representing states of ignorance. In the Shafer Dempster theory of belief functions (Shafer, 1976, pp. 23-34), ignorance is represented by a belief function that assigns zero belief both to a proposition A and its negation, $\text{Bel}(A) = \text{Bel}(\sim A) = 0$, but unit belief to their certain disjunction, $\text{Bel}(A \vee \sim A) = 1$. A weakness of this proposal is that it is what I shall call “contextual.” That is, our ignorance concerning some outcome A is not expressed simply by the value assigned directly to A . $\text{Bel}(A)=0$ can mean ignorance if $\text{Bel}(\sim A) = 0$, or it can mean disbelief if $\text{Bel}(\sim A) = 1$. Its meaning varies with the context. One may also represent ignorance through convex sets of probability measures; complete ignorance consists of the set of all possible measures on some outcome space. I have elsewhere (Norton, forthcoming, §4.2) explained my dissatisfaction with this last proposal. Briefly my concern is the indirectness of using probability measures, which do have properties *A. Addition* and *B. Bayes Property*, to simulate the behavior of distributions of belief that do not. One difficulty suggests that the simulation is not complete. We expect an epistemic state of ignorance to be invariant under the negation map (15). Convex sets of probability measures are not invariant under that map, for, under that map, an additive probability measures is transformed to a dual additive measure, which obeys a distinct calculus.

Finally, one may well wonder about the utility of the epistemic state of ignorance defined here. It invokes a single degree of belief that is neither complete belief nor disbelief, assigned equally to all contingent propositions in the outcome space, and is resistant to both addition and Bayesian updating. Might such a state really arise in some non-trivial problem? My contention elsewhere (forthcoming, §8.3) is that it already has. There is an inductive logic naturally adapted to inferences over the behavior of indeterministic physical systems. Its basic belief state coincides with the ignorance state described here.

Appendix

The following is a synopsis of the system described in Norton (forthcoming).

F. Framework

A (usually) finite set of propositions (sometimes assumed mutually exclusive and exhaustive) A_1, A_2, \dots is closed under the familiar Boolean operations \sim (negation), \vee (disjunction) and $\&$ (conjunction) and, occasionally, countable disjunction. The formula $A \Rightarrow B$ (“A implies B”) means that the propositions are so related that $\sim A \vee B$ must always be true. The universal proposition, Ω , is implied by every proposition in the algebra and is always true. The proposition, \emptyset , implies every proposition and is always false.

The symbol $[A|B]$ represents the degree to which proposition B confirms proposition A. It is undefined when B is of minimum degree, which means that $B=\emptyset$, or there is a C such that $B \Rightarrow C$ and $[B|C] = [\emptyset|C]$. The sentences $[A|B] \leq [C|D]$ and $[C|D] \geq [A|B]$ means ‘D confirms C at least as strongly as B confirms A.’ The relation \leq is a partial order; that is, it is reflexive, antisymmetry and transitive. The sentences $[A|B] < [C|D]$ and $[C|D] > [A|B]$ hold just in case $[A|B] \leq [C|D]$ but not $[A|B] = [C|D]$. For all admissible²⁷ propositions A, B, C and D:

$$[\emptyset|\Omega] \leq [A|B] \leq [\Omega|\Omega]$$

$$[\emptyset|\Omega] < [\Omega|\Omega]$$

$$[A|A] = [\Omega|\Omega] \text{ and } [\emptyset|A] = [\emptyset|\Omega]$$

$$[A|B] \leq [C|D] \text{ or } [A|B] \geq [C|D] \text{ (universal comparability)}$$

$$\text{if } A \Rightarrow B \Rightarrow C, \text{ then } [A|C] \leq [B|C] \text{ (monotonicity)}$$

A. Addition.

For any admissible proposition Z and mutually contradictory propositions X and Y, there exists an addition operator \oplus such that $[X \vee Y|Z] = [X|Z] \oplus [Y|Z]$ where \oplus is strictly increasing in both $[X|Z]$ and $[Y|Z]$.

B. Bayes Property is the conjunction of N. and M.:

N. Narrowness.

$$\text{For any proposition A and any admissible B, } [A|B] = [A \& B|B]$$

²⁷ Here and elsewhere, “admissible” precludes formation of the undefined $[.|B]$, where B is of minimum degree.

M. Multiplication.

For any proposition A and admissible propositions B and C such that $A \Rightarrow B \Rightarrow C$, there exists a multiplication operator \otimes such that $[A|C] = [A|B] \otimes [B|C]$ where \otimes is strictly increasing and thus invertible in both arguments (excepting $[B|C]$, when $[A|B]=[\emptyset |B]$).

R. Real Values.

For any admissible propositions A, A', B and B', the set of values possible for degrees of confirmation $[A|B]$ can be mapped one-one onto a closed set of reals such that the mapped real values $f([A|B]) > f([A'|B'])$ just in case $[A|B] > [A'|B']$.

All these properties combined are sufficient to entail the existence of real valued degrees of support that can be rescaled to yield a conditional probability measure.

References

- Bertrand, Joseph (1907) *Calcul des Probabilités*. 2nd Ed. Gauthier-Villars: Paris; repr. New York: Chelsea.
- Borel, Émile (1950) *Elements of the Theory of Probability*. Trans. John E. Freund. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1965.
- Cramer, Harald (1966) *The Elements of Probability Theory and Some of its Applications*. Huntington, New York: Robert E. Kreiger Publishing Co., 2nd Ed., 1966; reprinted 1973.
- Galavotti, Maria Carla (2005) *Philosophical Introduction to Probability*. Stanford: CSLI Publications.
- Gillies, Donald (2000) *Philosophical Theories of Probability*. London: Routledge.
- Howson, Colin and Urbach, Peter (1996) *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Chicago and La Salle, IL: Open Court.
- Jaynes, E. T. (1973) "The Well-Posed Problem," *Foundations of Physics*, **3**, pp. 477-493.
- Jaynes, E. T. (2003) *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
- Jeffreys, Harold (1961) *Theory of Probability*. 3rd Edition. Oxford: Oxford University Press.
- Kass, Robert E. and Wasserman, Larry (1996) "The Selection of Prior Distributions by Formal Rules," *Journal of the American Statistical Association*, **91**, pp. 1343-70.
- Keynes, John Maynard (1921) *A Treatise of Probability*. London: MacMillan; reprinted New York: AMS Press, 1979.

- Laplace, Pierre-Simon (1825) *Philosophical Essay on Probabilities*/ 5th ed. Trans. Andrew I. Dale. New York: Springer-Verlag, 1995.
- Norton, John D. (1994) "The Theory of Random Propositions," *Erkenntnis*, **41**, pp. 325-352.
- Norton, John D. (forthcoming) "Probability Disassembled," *British Journal for the Philosophy of Science*.
- Norton, John D. (manuscript) "Disbelief as the Dual of Belief."
- Shafer, Glen (1976) *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.
- Van Fraassen, Bas (1989) *Laws and Symmetries*. Oxford: Clarendon.
- Von Mises, Richard (1951) *Probability, Truth and Statistics*. 3rd German ed.; 2nd revised English translation. London: George Allen and Unwin, Ltd., 1957; repr. New York: Dover, 1981.