DRAFT

Chapter from a book, *The Material Theory of Induction,* now in preparation.

Why Not Bayes

John D. Norton

Department of History and Philosophy of Science

University of Pittsburgh

http://www.pitt.edu/~jdnorton

0. Prelude

A central proposition of this book is that there are no universal rules for inductive inference. The chapters so far have sought to argue for this proposition and to illustrate it by showing how several popular accounts of inductive inference fail to provide universally applicable rules. Many in an influential segment of the philosophy of science community will judge these efforts to be mistaken and futile. In their view, the problem has been solved, finally and irrevocably.

This segment of the community are "Bayesians" who work in what has come to be called "Bayesian epistemology." Its central idea is that issues of belief and inductive inference are to be treated solely by means of the probability calculus. The central structure is a conditional probability measure $P(A|B)$, the probability of proposition A against the background proposition B. The term "Bayesian" derives from an easily proven theorem in the probability calculus, Bayes' theorem:

$$P(H \mid E \& B) = \frac{P(E \mid H \& B)}{P(E \mid B)} P(H \mid B)$$

It provides the central engine for inference in Bayesian epistemology. The inference starts with some prior belief or inductive strength of support for an hypothesis H on the background B, $P(H|B)$. Learning evidence E leads the prior probability to be updated to the posterior probability $P(H|E\&B)$, which now incorporates the full import of the evidence E. This posterior probability

1

is computed via Bayes' theorem using the auxiliary quantities, the "likelihood" $P(E|H\&B)$ and the "expectedness" $P(E|B) = P(E|H\&B) \, P(H|B) + P(E|\text{-}H\&B) \, P(\text{-}H|B)$.

This is the barest sketch of the core notions of the Bayesian approach. They are now so widely know as not to require further elaboration. There is much more to the general Bayesian approach and there are many variant forms. Recalcitrant cases that do not easily fit with the core notion in its bare form are treated by "imprecise probabilities." The imprecision derives from replacing a single probability measure by a set of measures; or by replacing an additive measure by a superadditive measure. These are conceived as providing a generalized probabilistic analysis. Recently, Bayesian epistemology has been subsumed under the new heading of "formal epistemology." Its leading idea is that epistemic problems are to be addressed by formal and mathematical methods. There is little real change, however, as far as belief and inductive inference is concerned. Probability measures remain the principal instrument used to treat them.

For my purposes, the core commitment of this tradition resides in a single idea:

It's all probabilities.

Here just what it is to be probabilistic can be construed differently according to one's interpretive inclinations. However that general conception is taken to solve the essential problems addressed in this book. There are universal rules, this tradition holds. They are axioms of the probability calculus. Once that is recognized, all that remains are the finer details of determining just how they are to be applied to each problem. The big problem is solved.

The purpose of this and the following chapters is to explain why I am dissatisfied with this Bayesian solution.

1. Introduction

The case against the universality of probabilities will be made in this chapter in two parts. The first part will apply the general argument developed in Chapter 2 against the idea that any calculus, probabilistic or otherwise, can be universally applicable. Its core resides in the following:

> *Any logic of induction must restrict what happens in ways that go beyond logical*
> *consistency. Hence a logic of induction is applicable in some domain if the facts of*
> *that domain match the factual restrictions of the logic of induction. Since there is no*
> *universally applicable factual restriction, in general, different domains require*
> *different inductive logics.*

This argument will be developed more fully in Sections 3 and 4. It concludes that any calculus of induction must eventually reach a boundary to its domain of applicability beyond which it fails.

I will argue in Section 5 that efforts to develop theories of imprecise probabilities are misplaced attempts to disguise these boundaries. They use an additive measure merely as adjuncts to simulate the non-additive inductive logic of a new domain. In foundational terms, they mislead by fostering the impression that "it's all probabilities" even when the logic simulated is inherently non-additive.

As a foil for further analysis, Section 6, 7 and 8 will present an extreme but simple example of such a non-additive inductive logic. It is the relation of "completely neutral support," which is derived from the principle of indifference and illustrated by von Mises' example of different mixtures of wine and water. Section 9 reviews the extent to which theories of imprecise probability can accommodate completely neutral support. In so far as they do not accommodate it, they are inadequate; in so far as they do, they are superfluous. Any success in this one case merely postpones the inevitable failures, I urge, that must arise when they seek to accommodate more exotic logics.

The second part of the case against the universality of probabilities is a general rebuttal of the many proofs offered in the literature as demonstrating the necessity of probabilities. These proofs come in different guises. One of the oldest and best known is Ramsey and de Finetti's Dutch book argument. It is used to infer than non-probabilistic distributions of belief are "incoherent," which is a form of irrationality. All such proofs must fail and they must fail in the same way, for this reason:

> *A proof of necessity of probabilities is a deductive argument whose premises must be at least as strong logically as the conclusion. Therefore the assumptions of the proof must already presuppose the necessity of probabilities or something logically stronger. Hence by dominance we are better off simply presupposing the necessity of probabilities at the outset and forgoing the proof.*

This general argument is developed in Sections 10 and 11. It is used to predict that a careful analysis of a proof of the necessity of probabilities triggers a regress of reasons. For the assumptions used in the proof will always be found to be improperly grounded. Attempts to provide proper grounding will require new assumptions that will then also prove to be improperly grounded.

The principal illustration of this circularity and the ensuing regress of reasons will be the recent efforts to vindicate probabilities by means of notions of accuracy and scoring rules. Since the analysis is quite extensive, it is postponed to the next chapter. To show that other attempts at vindication fail in the same way, two more examples are given briefer treatment in this chapter. Section 12 recalls the Dutch book argument. It identifies which assumptions already have the axioms of the probability calculus built into them and recounts the failure of attempts to remove the circularity. Section 13 repeats this analysis for a different approach developed by Cox (1961)

and Jaynes (2003). In it, necessary conditions are identified for strengths of inductive support and from them the computational rules of the probability calculus are recovered by functional analysis. We shall see that the necessity of the conditions requires further grounding, triggering the now familiar regress of reasons. Conclusions in Section 14 suggest further directions of exploration.

As a preliminary, in Section 2, I will distinguish objective from subjective approaches and apologize to the reader for not always distinguishing them clearly as the chapter unfolds.

Finally before proceeding I would like to give Bayesian epistemology its due. My view is far from a complete dismissal of Bayesian epistemology. I view it in the same way as I view all other candidate logics of induction. Whether it applies in some domain is determined by the background facts of the domain. These background facts will also determine the variety of Bayesianism applicable. Stronger facts will authorize strict Bayesianism in which inductive support or, subjectively speaking, beliefs are measured by a single probability measure. Weaker facts will authorize a relaxed Bayesianism in which these supports are represented by sets of probability measures or upper and lower bounds. There are many domains in which varieties of Bayesian analysis are authorized and can be applied. In them, it provides a wonderful instrument.

It is formally precise where other accounts flounder. Arguments from analogy struggle to separate the strong from the weak analogies. Accounts that reward simplicity cannot provide a clear and unobjectionable notion of simplicity whose measure translates mechanically into inductive strength. In contrast, once the probability space is well-defined, the Bayesian analysis has no such trouble. Determining all its relations is reduced to well-defined computations in the probability space. When the system under investigation becomes very complicated, other approaches provide little guidance on how apparently conflicting evidence is to be combined. The fossil record is best explained by an old earth. The earth's cool temperature is best explained through Newton's law of cooling by a young earth. Using inference to the best explanation, to which do we infer? If they can pass the formidable hurdle of providing a well-defined probability space, Bayesians can answer the corresponding questions by mechanical computation. For all the information needed to trade off competing items of evidence lies within the conditional probabilities. Indeed, if any general question about belief or inductive support can be translated into a precise query in probability theory, it can be decided by a theorem that affirms or denies it.

With virtues as strong as these, it is all too appealing to hope that Bayesian analysis can be applied universally. When the inevitable problems arise, it is easy to dismiss them as the routine teething troubles of any infant who will outlive them and grow to boisterous maturity. Once that was a defensible attitude. As time passes and the problems remain unsolved, we can no longer afford to indulge the universal aspirations. If we are to understand what inductive inference is fundamentally, we need a different approach.

2. Objective and Subjective Bayesianism

My concern in this work are the objective relations of inductive support. Bayesians are also interested in these relations in so far as they expect them to be embraced by their analyses in one form or another. What complicates responding to Bayesian ambitions of universality is that Bayesianism is not a univocal doctrine. Rather it comes in many varieties.

A major division is between the objective Bayesians, such as Edwin Jaynes, and the subjective Bayesians, such as Bruno de Finetti. The objective Bayesians are distinguished by the claim that, in any epistemic situation, there is one correct probability distribution applicable. There is, in particular, one correct prior probability. In this regard, the project of objective Bayesians is closest to mine. I am comfortable regarding the objective Bayesians' conditional probability $P(H|E\&B)$ as an attempted expression of the objective strength of inductive support provided by evidence E in background B for hypothesis H. As I will remark briefly in the concluding section, the primary obstacle facing the objective Bayesians specifically is that rules needed to define this one correct prior are arbitrary.

Subjective Bayesians permit many probability distributions, constrained only by conformity with the axioms of the probability calculus. They characterize the freedom in our choice among the probability distributions as a free exercise of opinion. Thus, antecedent to the consideration of evidence, we are free to choose any prior probability distribution we like. Thus, at best, for a subjective Bayesian the conditional probability $P(H|E\&B)$ cannot simply express *the* strength of inductive support accrued to H since it has no unique value. Rather it is, at best, one of many possible mixes of evidential support and opinion. The hope is that eventually, in some longer term limit, the balance will move decisively toward objective support. There are also many attempts to derive measures from confirmatory support from the subjective probabilities.

My principal concern with subjective approach is that it demotes strengths of inductive support to a derived quantity. The primary quantity, the conditional probability, is a measure of belief or credence. Strengths of inductive support are to be recovered from them. That a notion of belief should be more primitive than a notion of inductive support has it the wrong way round. We wish to assess the strength of inductive support evolutionary theory derives from the fossil record or big bang cosmology derives from the cosmic background radiation. To make this assessment, we should not first have to determine our beliefs or credences. These strengths should not be dependent on our beliefs, else the objectivity of science is at risk. Worse, the project of assessing these strengths of inductive support from the beliefs has proven to be so troublesome that no univocal assessment is recoverable from the present literature in subjective Bayesianism. It was an approach that was risky to try and has not yet succeeded.

Since there is so much in common in the objective and subjective approaches to Bayesianism, it is impractical in what follows for me to keep them fully separated. There is often no need since argumentation concerning subjective probabilities can often be adapted to apply to objective probabilities; and conversely. Thus, in this chapter and the related chapters that follow, I will move freely between treating probabilities as objective relations of support and subjective credences. Please accept my apologies in advance for any confusions this may cause.

3. The Main Failing of Bayesianism

To repeat the disclaimer of the introduction, I have no quarrel with the application of Bayesian analysis in specific cases. There are many of successful and interesting cases that arise when the background facts provide the warrant needed. For example, a physical theory may supply the probabilities as the physical chances; and the particular conditions of the system may provide unambiguous prior probabilities.

My concern is the claim that Bayesian analysis is *universally* applicable to all systems, where it can supply the one, true logic of inductive inference. On the contrary, I will argue, there is a boundary beyond which Bayesian analysis fails. The existence of this limit is a corollary of the more general claim that there are no universal rules of inductive inference. For, without the boundary, Bayesian analysis would be providing a universal rule of inductive inference. The argument for this general claim was developed in Chapter 2. It is recapitulated here for the special case in which a mathematical calculus is used in one manner or another as a logic of inductive inference. The argument applies to all calculi used this way, not just to the probability calculus.

The key premise in the argument is that a calculus of inductive inference must place limitations on what can happen that is more restrictive than logical consistency.[1] Otherwise the calculus is not part of a system that implements inductive inferences. If the limitation goes beyond logical necessity, then it is by definition a contingent restriction, that is, one that may without logical contradiction be true or false. It follows immediately that there will be some

---

[1] That it must do this might not be immediately clear. The standard manipulations within the probability calculus, for example, are all deductive. Take ten independent tosses of a fair coin. If the probability of heads is 1/2 in each toss, then the probability that at least one heads appears is $1 - 1/2^{10} = 0.999$. So far all the reasoning has been deductive. The inductive component only enters when we employ an interpretive rule that tells us that outcomes of near unit probability are to be expected. Without some sort of interpretive rule like this, the probability is simply a mathematical quantity with no import for real things in the world.

conceivable domains that conform with the restriction and some that do not. Whether the calculus in question can be applied in some domain as a logic of inductive inference will depend on whether the requisite facts obtain. When they do, these facts warrant the use of this calculus in this domain.

This consideration applies directly to objective Bayesian approaches, for according to them, the probability calculus is the "logic of science." Or so proclaims Jaynes (2003) in the title of his treatise, *Probablity Theory: The Logic of Science.* The consideration also applies to subjective Bayesian approaches. While a probability measure for them mixes opinion and evidential warrant, the probability calculus does constrain someone to conform their beliefs with the evidence. The expectation is that conditionalization on a sufficiently rich body of evidence for some theory will lead a subjective Bayesian to mass the probability almost entirely on the true theory. Since such a body of evidence is finite but the theory's scope is infinite, this is a form of inductive inference. In this sense, a subjective Bayesian is implementing a scheme of inductive inference, although not as directly as an objective Bayesian.

We have now concluded that contingent facts warrant the applicability of some particular calculus of inductive inference in some domain. We might still hope that a single calculus is warranted universally. That would happen if it turns out that there is a single, contingent warranting fact that obtains in all domains in which we might conceivably practice inductive inference. Such a fact was pursued, in effect, in the nineteenth century under the guise of a search for a principle of the uniformity of nature. The search for such a universally applicable principle failed. As described in greater detail in Chapter 2, all candidates either proved empirically false or so hedged as to be vacuous.

The facts prevailing in some domain warrant the inductive logic applicable there. There is no single warranting fact common to them all. It follows that different inductive logics are warranted in different domains. This locality does not preclude one domain being very large. The success of probabilistic methods suggests that the domain or domains in which they are warranted as a logic of inductive inference are large. However every such domain is bounded and there are others beyond it in which a different logic is warranted. The logic warranted in these other domains may be governed by a calculus. But there may well be domains so irregular in their facts that no well-developed calculus can systematize whichever inductive inferences are warranted in them.

4. Probabilities without Warrants

The need for some sort of warrant for the use of probabilities becomes quite apparent if we consider cases in which there is no warrant. What results are striking inductive fallacies.

A simple example is provided by Van Inwagen's (1996, p. 95) question "Why is there anything at all?" He notes that there is only one possible world with nothing at all. There are infinitely many possible worlds with something, however, each differing in the configuration of something. Since we are assuming antecedently to have no basis for knowing what there is, if anything at all, we distribute our probabilities roughly uniformly over all the possible worlds. The result is that all the probability mass is attracted to the set of worlds in which there is something. It follows that the probability of there being nothing is zero. It is, Van Inwagen concludes (p. 99), "as improbable as anything can be."

This conclusion is derived fallaciously. Not even the prodigious powers of the probability calculus can legitimately extract such a strong conclusion from premises so bereft of content. The fallacy derives directly from employing a probabilistic analysis in a context in which no background facts warrant the probabilities. This particular fallacy is unfortunately widespread. It is a version of what I have called elsewhere (Norton, 2010, §4) the "inductive disjunctive fallacy." More examples are described in (Norton, 2010, §4).

The "doomsday argument"[2] provides another illustration of a related fallacy. It uses only the evidence that our world has survived for t years. It asks after the probability that our world will end in $T > t$ years—"doom." Since our t can equally be any of the total T years of our world, the probability that our present world has survived for t years is $P(t|T) = 1/T$. The quantity we seek is the posterior probability $P(T|t)$, the probability that the world meets its doom after T years, given that it has survived for t years. An application of the ratio form of Bayes' theorem tells us that[3]

$$P(T_1 \mid t) / P(T_2 \mid t) = P(t \mid T_1) / P(t \mid T_2) = T_2 / T_1$$

Substitute $T_1 = t$ and $T_2 = 10t$, and we recover $P(T_1|t) / P(T_2|t) = 10$. This extraordinary conclusion tells us that doom is ten times more likely right now at t than is survival for another ten ages to $10t$.

Once again, the analysis delivers too much. The evidence is just that our world has survived t years. That is too thin an evidential basis for the strong conclusions drawn. They are not a reflection of what the evidence authorizes. They are merely artifacts of the use of an unwarranted inductive logic.[4] See Norton (2010, §6) for further analysis and for a proposal for a reduced inductive logic more appropriate to the problem.

---

[2] See Bostrom (2002a, chaps. 6–7) for an entry into the earlier literature on this argument.

[3] Assume equal prior probabilities $P(T_1) = P(T_2)$.

[4] Bostrom (2002, p. 57) seeks to warrant probabilistic analysis with his "self-sampling assumption": "One should reason as if one were a random sample from the set of all observers in

5. Mapping the Boundaries: the Fate of Imprecise Probability

It is an interesting exercise to map out the boundaries for a probabilistic inductive logic. A simple axiom system, such as Kolmogorov's (1950) celebrated system, guides us to the boundaries.[5] The axioms have us posit a set of propositions[6] in which:

- non-negativity: we assign a non-negative, real valued probability $P(A)$ to proposition A;

- normalization: we assign a probability of unity $P(\Omega) = 1$ to the universal proposition (tautology) Ω; and

- additivity: $P(A \lor B) = P(A) + P(B)$ when proposition A and B contradict.

Inductive problems in which each of these fail are well-known. This chapter and chapters that follow will recount some of them. The "completely neutral support" relation below violates additivity, as do the indeterministic systems of a later chapter. Normalization and additivity are violated by the infinite lottery. That the structure is a real valued function is violated by the quantum inductive logic of a later chapter.

That some boundary has been reached is not controversial. What is controversial, at least in my mind, is how we should respond to it. I believe the correct response is to recognize that we have found the boundaries of probabilistic logic; that we should recognize that different logics prevail in the domains beyond it; and that we should begin the task of identifying them.

The standard response in the literature is different. It is to weaken the probability calculus until the generalized calculus has been weakened enough to encompass whatever troublesome counterexample has arisen. For example, as we shall see shortly, the field of imprecise probability encompasses systems that violate additivity by replacing a single probability measure

---

one's reference class." Here "random sampling" implies equal probability of each sample drawn. Since it is an assumption without factual basis, it provides no warrant. Rather it enables us to identify and name the arbitrary posit that is the origin of the inductive fallacy.

[5] Kolmogorov's axioms are simpler since they specify an unconditional probability $P(A)$. Conditional probabilities are introduced through the definition $P(A|B) = P(A\&B)/P(A)$. This approach is more restrictive than providing more complicated axioms directly for conditional probabilities. But the simpler approach suffices for the present analysis since the more complicated approaches only move the boundaries slightly.

[6] More precisely, we posit a Boolean algebra of propositions, which is a set of propositions closed under finite or countable disjunction $\lor$, conjunction $\&$ and negation $-$.

with a set or them; or it may employ superadditive measures. This is the project of variety of approaches grouped under the heading of "imprecise probability."[7]

This approach is, in the end, an ill-fated attempt to preserve the core idea that "It's all probabilities." The domains covered are ones in which an inherently non-probabilistic inductive logic is warranted. What imprecise probability does is to employ an additive calculus to simulate a non-additive logic. It thus preserves the illusion that an additive measure is somehow still the core of the logic. The better approach would simply have been to recognize that a qualitatively distinct logic is required and to map out its behavior as a distinct logic.

In any case, this general stratagem of extending the calculus offers only a temporary respite. For as long as the generalized calculus supports inductive inference, it must place restrictions on the system that go beyond mere logical consistency. These are contingent constraints that, recalling the argument of Section 3 above, will not obtain in all domains. Further investigation will reveal new boundaries and new domains beyond them, as we shall see shortly in Section 9 below.

This standard response is driving towards an unhappy ending. Each time a calculus is generalized to embrace new examples, it is weakened in the sense that it becomes less restrictive. As long as the generalized logic places some restrictions on systems beyond logical consistency, the domains in which it applies are limited. The need to generalize to embrace unanticipated counterexamples will continue. The process of generalization can only assuredly terminate when the logic places no restrictions beyond logical consistency on its domain. But then it has ceased to be an inductive logic.

6. The Principle of Indifference

To make these concerns more concrete, it will be helpful to develop the simplest case in which the boundaries of probabilistic analysis are breached. It is "completely neutral support" and arises when we have inductive support that is, in objective terms, maximally uninformative. In subjective terms, it corresponds to the case of complete ignorance. This case has been explored extensively in Norton (2008, 2010) as part of an investigation of the import of the principle of indifference. We will first recall the principle and its application; and then develop the notion of completely neutral support in the next section.

The form of the principle that I prefer is:

---

[7] For an introduction to this literature, see Bradley (2016) and the resources on the website of the *Society for Imprecise Probability*, http://www.sipta.org.

Principle of Indifference. If one has no grounds for distinguishing several outcomes, then we should assign equal inductive support to them.

The principle in this form is a truism of evidence. It just reflects the requirement that discriminations in inductive support cannot be made arbitrarily.

The principle applies when we have indistinguishable outcomes. They are realized most securely through invariances, which are transformations that leave the relations of inductive support unchanged. Their use is familiar and unproblematic, initially. When we have a fair coin toss, our formal analysis is unchanged if we switch the labels on the sides of the coin. Whatever ground we have for favoring heads would remain unchanged if reassign the label "heads" to the other side of the coin; and reassigned the label "tails" similarly. That is we have no grounds for distinguishing the outcome of either side. The principle of indifference requires us to assign equal support to each. If the relations of support are probabilistic, then each side is assigned equal probability.

In case of a coin toss, the invariance under this permutation of the labels is derived from background physical facts: the mechanical conditions of tossing are such that they favor both sides equally. One can also have a more epistemic version. The coin need not even be tossed. We might just imagine it as sitting somewhere, untouched, in a drawer. But since our information about the coin is so limited, we have no grounds for distinguishing whether it is heads up or tails up.

Invariances like these seem benign until we start to combine them. Then they yield the well-known paradoxes of indifference. An early and familiar example is due to Keynes (1921, Ch.4), who also named the principle of indifference. We ask after the country of a man who may be from

$$\text{France, Ireland, Great Britain} \qquad (1)$$

Since we have no grounds for discriminating among them, so we assign equal probability of 1/3 to each. However the disjunction of two outcomes (Ireland or Great Britain) is equivalent to just British Isles at the time Keynes first wrote. So we might equally ask if the man is from

$$\text{France, British Isles} \qquad (2)$$

Since again we have no grounds for discriminating, we assign equal probability of 1/2 to each. We have now arrived at contradictory assignments. For we have assigned both probability 1/3 and probability 1/2 to France as the man's country.

Examples like these are usually used to impugn the principle of indifference. That is a misdiagnosis. The principle is a truism of evidence and not readily discarded. It just says that the support for outcomes should not differ without a reason. What is overlooked in these efforts to impugn the principle is that the real cause of the trouble lies elsewhere. It is the presumption that

relations of inductive support must always be probabilistic. These paradoxes of indifference are an early indication that they need not always be so.

Recently there have been several alterative interpretations of the import of the principle of indifference on representing the neutrality of support. Benétreau-Dupin (2015) draws on the existing ideas in imprecise probability and explores representing completely neutral support through sets of probability measures. Eva (manuscript) proposes a novel accommodation in which not all degrees of support are comparable.

7. Completely Neutral Support

7.1 Invariance under Redescription

The transformation from (1) to (2) is a "disjunctive coarsening" of the outcome space. Two outcomes are replaced by a single outcome, their disjunction. The reverse transformation is a "disjunctive refinement." If we conceive these operations as redescriptions of the outcomes, Keynes' example depends on a particular invariance:

Invariance under redescription. In cases of completely neutral support, equality of inductive support over outcomes remains under disjunctive coarsening and refinement.

We arrive at the formal representation of completely neutral support by applying this invariance to an outcome space that has finitely many, mutually exclusive atoms $A_1, A_2, A_3, \ldots, A_n$, where an atom is the logically strongest proposition in the outcome space.[8] If our circumstance is maximally uninformative concerning these outcomes, then the principle of indifference enjoins us to assign equal support to each. Writing:

$$[A|B] = \text{inductive support accrued to proposition } A \text{ from } B.$$

we then infer:

$$[A_1|\Omega] = [A_2|\Omega] = [A_3|\Omega] = \ldots = [A_n|\Omega] = I \qquad (3)$$

where I represents the common inductive strength of support.

We can disjunctively coarsen the outcome space by replacing the first two propositions A_1 and A_2 by their disjunction, which we will write as $A_{1v2} = A_1 \vee A_2$. The new outcome space

---

[8] As before, the outcomes space is a Boolean algebra of propositions whose universal proposition or tautology Ω is the disjunction of all the atoms $\Omega = A_1 \vee A_2 \vee \ldots \vee A_n$. Proposition A_i is an atom just if any proposition A that entails it is either A_i itself or the contradiction.

is $A_{1v2}, A_3, \ldots, A_n$ and has only n-1 propositions. Proceeding as did Keynes, we remain maximally uniformed about these propositions, so we must assign equal support to each. That is,

$$[A_{1v2}|\Omega] = [A_3|\Omega] = \ldots = [A_n|\Omega] = I \tag{4}$$

The same strength of support I must be used in both (3) and (4) since they have many common terms. For example, both include $[A_3|\Omega]$, so we can infer

$$[A_1|\Omega] = [A_2|\Omega] = [A_3|\Omega] = [A_{1v2}|\Omega] = [A_1 \text{ v } A_2|\Omega] = I$$

Continuing by forming more disjunctive coarsenings of the original outcome space, it is easy to see that the support offered to any contingent disjunctions of atoms[9] is the same strength of support I:

$$[\text{any contingent disjunction of atoms} \mid \Omega] = I$$

However every contingent proposition in outcome space is equivalent to some disjunction of atoms. Thus we arrive at[10]

Completely neutral support (5)

$$[\text{tautology } \Omega \mid \Omega] = 1$$

$$[\text{any contingent proposition} \mid \Omega] = I$$

$$[\text{contradiction} \mid \Omega] = 0$$

The strengths "1" and "0" have been chosen for continuity with the familiar probabilistic case. Their arithmetic properties are not invoked.

   This subsection argues for a unique characterization (5) of completely neutral support. It is tempting to block the argument at equation (4) by asserting that the principle of indifference should only be applied to the most refined outcome space, which is (3) in this case. I set aside the question of whether this knowledge is adequate to block the argument. For whether it is or not, the difficulty is easily escaped. It is presumed that we know that (3) is the case of maximum refinement. So let us consider the case in which we do not know that (3) is the maximum refinement; or that we know positively that there is no maximum refinement. Then the argument for completely neutral support goes through.

---

[9] This requirement of contingency excludes the tautology Ω.

[10] Keeping distinct values for the tautology and contradiction presumes sufficient logical knowledge that we can discern them from the contingent propositions. One could also define a still more extreme case in which we are maximally ignorant of deductive relations amont the propositions, so that all the strengths are I.

How might there be no maximum refinement? Such a case arises if the propositions represent ranges of some real-valued parameter x. E.g. A_m might correspond to $m \leq x < m+1$. Then we can disjunctively refine by replacing A_m by a disjunction BvC, where B corresponds to $m \leq x < m+1/3$ and C corresponds to $m+1/3 \leq x < m+1$. Since these intervals can be divided indefinitely, there is no most refined outcome space.[11]

7.2 Invariance under Negation

The invariance under redescription of the last section is already well represented in the literature. There is a second, less familiar invariance that leads to the same result. To see it consider some proposition A about which you know nothing at all. How well supported is it? Now consider its negation, not-A. Is the negation any more or any less well supported? If there is any doubt about how to answer, imagine that the first question asked concerned the proposition not-A, which we initially labeled B. Is not-B any more or less supported than B?

Invariance under negation asserts that the two strengths of support are the same, simply because, by supposition, we have no basis for discriminating between them. Switching their labels makes no difference to the strengths of support.

Invariance under negation. In cases of completely neutral support, the inductive support for a contingent[12] proposition and its negation are the same.

Let us implement this invariance in the outcome space with atoms $A_1, A_2, A_3, \ldots , A_n$. Any contingent proposition consists of a disjunction of some number of atoms from one to n-1. For example, the negation of A_1 is A_2 v A_3 v … v A_n; and the negation of A_2 v A_3 v … v A_n is A_1; and so on for all other possible combinations. If we have a case of completely neutral support, we infer from invariance under negation that

$[A_1 | \Omega] = [A_2 \text{ v } A_3 \text{ v } \ldots \text{ v } A_n | \Omega] = I_1$

$[A_1 \text{ v } A_2 | \Omega] = [A_3 \text{ v } A_4 \text{ v } \ldots \text{ v } A_n | \Omega] = I_{1,2}$

etc.

---

[11] It is tempting to argue that the refinement must divide parameter values into uniform intervals. That requirement fails since what may be a uniform division for one scaling of the parameter will not be so for another. The wine and water example below illustrates the problem.

[12] A stronger version arises if we cannot discern the tautology and contradiction from the contingent propositions. Then the restriction to contingent propositions can be dropped.

The strengths of support must be distinguished as $I_1, I_{1,2}, \ldots$ at this stage, since negation invariance, by itself, is not strong enough to force all the strengths to the same value. Their equality, however, can be recovered if we add the condition:

Monotonicity. The strength of support of a proposition is no greater than[13] that of its consequences. If A entails B, then $[A|\Omega] \leq [B|\Omega]$.

Since A_1 entails A_1 v A_2 and A_2 v A_3 v … v A_n is entailed by A_3 v A_4 v … v A_n, we have

$$I_1 = [A_1 \,|\Omega] \leq [A_1 \text{ v } A_2 \,|\Omega] = I_{1,2}$$

$$I_1 = [A_2 \text{ v } A_3 \text{ v } \ldots \text{ v } A_n \,|\Omega] \geq [A_3 \text{ v } A_4 \text{ v } \ldots \text{ v } A_n \,|\Omega] = I_{1,2}$$

These two inequalities can only obtain if the two strengths are equal:

$$I_1 = I_{1,2} = I$$

Proceeding analogously for all the other cases, we recover the equality of the strengths of support with the single value I for all contingent propositions. The details of the recovery are straightforward but a little tedious and are given in Norton (2008, §6.3).

7.3 Invariance from Ignorance or Positive Warrant

The representation of completely neutral support has been developed assuming that the invariances prescribed can come about in some circumstance. It is tempting to invert the argumentation. Since the totality of these invariances is incompatible with a probabilistic treatment of inductive support, might we then infer that it is impossible for us ever to be in a position in which these invariances are realized? I have argued that sufficient ignorance will realize these invariances. However my real concern in this work is not ignorance but strengths of support warranted by background facts. We might well wonder what sort of facts could realize these invariances. What sort of coin tosses or die throws or other similar machines could yield probabilities such that $P(A_1) = P(A_2) = \ldots = P(A_n) = P(A_1 \text{ v } A_2) = P(A_1 \text{ v } A_2 \text{ v } A_3) = \ldots$ as required by (5)? Since these probability assignments violate the additivity axiom of the probability calculus, no probabilistic randomizer can realize them, precisely because its mechanism is probabilistic.

This inversion of the arguments succeeds only in so far as we are restricted to warrants for support arising from probabilistic randomizers. The chapter on "Indeterministic Physical Systems" describes many physical systems whose indeterminism is not probabilistic and which

---

[13] That is, we assume that there is a partial order "<" defined over the strengths. It is transitive and antisymmetric. Monotonicity is widely assumed but is not necessary and one can conceive logics in which it fails. An example is the specific conditioning logic of Norton (2010a, §11.2).

realizes the two invariances employed above. The inductive logic warranted for these systems conforms with completely neutral support (5). However, where here I have used the three values 0, I and 1, in the later chapter I relabel these values as *imp* ("impossible"), *poss* ("possible") and *nec* ("necessary") to reflect better they physical underpinnings of these new cases.

8. Von Mises' Wine and Water.

The argument for completely neutral support derives from either of the two invariances just stated. It is tempting to try to defeat them by calling upon asymmetries in propositions that are not respected by the invariances. For example, while there may be no finest disjunctive refinement, there are some that are more refined and some that are less so. All negations are not the same in their atom counts. The negation of a single atom proposition A_1 has a different atom count from the negation of a disjunction of n-1 atoms, A_2v…v A_n. These are all asymmetries among the cases that are not reflected in the invariances. We may well ask how the invariances can be maintained with these asymmetries.

The uniform response to all concerns of this type is merely to reduce our knowledge still further, until the symmetries are restored. Then the invariances apply and completely neutral support is recoverable. The asymmetries depend on choosing a particular outcomes space to describe some system's behavior. A proposition may consist of one atom in one outcome space, but a disjunction of n-1 atoms in another, where both spaces describe the same system. These differences of atom counts are then immaterial to the invariance if we have no way to discern which of the outcome spaces is the "right one."

A version of von Mises' wine and water example illustrates this effect. A goblet contains a mixture of wine and water. All we know is that the ratio x of wine to water lies in the interval $0.5 < x < 2$. It follows that the ratio $y = 1/x$ of water to wine lies in the interval $2 > y > 0.5$. The variables x and y each allow the definition of outcome spaces that describe the same physical goblet. The first has atoms:[14]

$$A_1: 0.5 < x < 1 \quad A_2: 1 < x < 1.5 \quad A_3: 1.5 < x < 2$$

and the second has atoms

$$B_1: 0.5 < y < 1 \quad B_2: 1 < y < 1.5 \quad B_3: 1.5 < y < 2$$

The principle of indifference requires us to assign uniform support across the atoms:

$$[A_1|\Omega] = [A_2|\Omega] = [A_3|\Omega]$$

---

[14] To avoid the need to juggle too many "<" and "≤," I employ the expedient assumption that x and y can never adopt exactly the values 0.5, 2/3, 1, 1.5 and 2.

$$[B_1|\Omega] = [B_2|\Omega] = [B_3|\Omega]$$

One might try to adapt probabilities to these equalities, by setting

$$P(A_1|\Omega) = P(A_2|\Omega) = P(A_3|\Omega) = 1/3$$

$$P(B_1|\Omega) = P(B_2|\Omega) = P(B_3|\Omega) = 1/3$$

The adaptation fails since $A_1 = B_2$ v B_3 so that we end up with a contradiction:

$$1/3 = P(A_1|\Omega) = P(B_2\text{v } B_3|\Omega) = P(B_2|\Omega) + P(B_3|\Omega) = 2/3$$

A similar contradiction follows from $B_1 = A_2\text{v } A_3$.

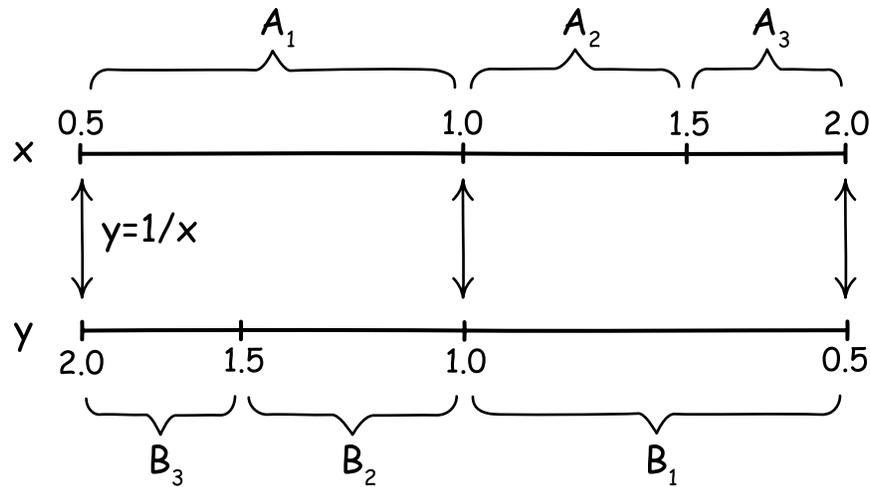Figure 1 illustrates how the atoms in the two spaces are related:

Figure 1. Relations between two outcome spaces for the wine and water example

Instead of trying to impose probabilities where they do not belong, we can apply the invariances to arrive at the completely neutral support (5). We start with a coarsened outcome space with two outcomes: the ratio of wine to water is either greater or less than 1. This space can be represented in two equivalent ways:

$$A_1, A_{2v3} = A_2 \text{ v } A_3 = B_1 \quad \text{or} \quad B_1, B_{2v3} = B_2 \text{ v } B_3 = A_1$$

Following the principle of indifference, we assign equal support to each outcome:

$$[A_1|\Omega] = [B_1|\Omega] = I$$

Implementing invariance under redescription, we disjunctively refine the outcome space and expect the equalities to be preserved. There are two ways to implement the disjunctive refinement. We can refine $B_1 = A_2$ v A_3 and end up with

$$[A_1|\Omega] = [A_2|\Omega] = [A_3|\Omega] = I$$

Since A_2 v $A_3 = B_1$ we have

$$[A_2|\Omega] = [A_3|\Omega] = [A_2 \vee A_3|\Omega] = I$$

Or we can refine $A_1 = B_2$ v B_3 and end up with

$$[B_1|\Omega] = [B_2|\Omega] = [B_3|\Omega] = I$$

Since B_2 v $B_3 = A_1$ we have

$$[B_2|\Omega] = [B_3|\Omega] = [B_2 \vee B_3|\Omega] = I$$

In these relations, we have now recovered much of completely neutral support (5) for the two outcome spaces. Continued application of the invariances recovers the remaining parts. For example, applying either invariance to the disjunctive coarsening A_2, $A_{1v3} = A_1$ v A_3 returns $[A_2|\Omega] = [A_1 \vee A_3|\Omega] = I$.

Returning to the concerns expressed at the start of this section, there is no sense in which one of the outcome spaces A_1, A_2, A_3 or B_1, B_2, B_3 is more refined than the other. The first represents a refinement of B_1 but not A_1; and the second a refinement of A_1 but not B_1. The perfect symmetry in all the formulae gives us no basis for preferring one over the other.

Negation invariance is also implemented automatically. A_1 is the negation of B_1 and conversely. They are both assigned the same support I. The earlier concern that negation invariance might be troubled by an asymmetry in atom counts is also not realized. In the A-outcome space, A_1 is comprised of one atom and B_1 is a disjunction of two atoms. In the B-outcome space, this is reversed: A_1 is a disjunction of two atoms and B_1 is comprised of one atom. Once again, the perfect symmetry in all the formulae gives us no basis for preferring one over the other.

9. Imprecise Probabilities Again

The literature in imprecise probability treats this case of completely neutral support or, as they characterize it often in subjective terms, of complete ignorance. In so far as these treatments are seeking to replicate formally the behavior of completely neutral support (5), they are doing the right thing. The invariances show that (5) is the correct representation for this case. There will be room to quibble about the treatments given in imprecise probability, as I will below, but these quibbles are minor in comparison to the major concern. It is that they are not doing the right thing, if this representation of complete ignorance is intended as part of the case for the universality of the particular scheme employed, now conceived as some kind of a generalized probability theory. For I have already described above in Section 4 how all such efforts are

18

necessarily ill-fated. In so far as the generalized probability theory places restrictions that go beyond logical consistency, these restrictions are contingent and it is inevitable that there are systems that contradict these factual restrictions. The generalized probability theory must fail to apply to these systems, so that its aspirations for universality must fail.

Consider a popular approach to imprecise probability that employs sets of probability measures to represent credal states. Benétreau-Dupin (2015, §3) has given a careful accounting of them and their prospects in regard to complete neutrality of support. This approach depends on the assumption that probability measures can be defined for the system in question. Otherwise sets of the measures cannot be formed. Thus the approach fails when applied to systems with nonmeasurable outcomes, such as are investigated in the chapter "Uncountable Problems," for these outcomes admit no probability measures. It also fails when applied to infinite dimensional outcome spaces, for they admit no non-trivial additive measure, even if the requirement of normalization to unity is dropped. We shall see such an example arising naturally in the chapter "Indeterministic Systems" in conjunction with an indeterminism in Newtonian cosmology.

There are lesser technical issues as well. As Benétreau-Dupin (2015, §3) recounts, it is unclear just which set of probability measures should represent completely neutral support. The natural choice is the set of all probability measures in the outcome space. However that set has the unappealing property for Bayesians that it is preserved under conditionalization so that inductive learning is precluded.

In my view, this representation is needlessly complicated since it assigns no definite probability value to a given outcome. Rather it assigns all values to each outcome and not just as an interval of values, but with each value part of one of infinitely many probability measures. There is no analog of the simple and adequate ignorance strength I. Worse, a set of probability measures violates invariance by negation. This arises because every probability measure is non-decreasing as we pass through chains of deductive consequences, such as

(Contradiction)

entails (A_1)

entails (A_1 v A_2)

entails (A_1 v A_2 v A_3)

entails …

entails (tautology)

Since the contradiction is assigned zero probability and the tautology unit probability, the probabilities of these outcomes must, at some point, be strictly increasing. That is, this strict increase endows the probability measures with a directedness from fewer to greater atom propositions. The operation of negation maps the strengths assigned to propositions with fewer

atoms to those with more atoms; and conversely. That is, it flips the assignments of inductive strengths with respect to this direction. It follows that any assignments of strengths that is directed cannot be preserved by negation. Since all measures have such directedness, a set of measures cannot be preserved under negation. The invariance is violated. Norton (2007, §6) explores this failure as a failure of a duality required by the representation of completely neutral support. Benétreau-Dupin (2015, §3) has sketched a dissenting view.

Other popular approaches employ some form of superadditivity of a measure.[15] An early version arises in the theory of Shafer-Dempster belief functions. The "vacuous belief function" (Shafer, 1976, p.22) assigns unity to the tautology and zero to every other proposition, including the contradiction. This vacuous belief function does respect both invariances of Section 7 above. However it has the awkward feature of assigning the same value of zero also to the contradiction. That means that its individual values do not distinguish complete disbelief, which we must have in the contradiction, from complete ignorance, which is presumed for all the contingent propositions.

Walley's (1991) related approach represents a credence in each outcome by two numbers, a lower and an upper probability. So-called "vacuous upper and lower probabilities" (p. 92) assign a zero lower probability and a unit upper probability to all contingent propositions. "They seem to be," Walley writes, "the only reasonable models for 'complete ignorance'…" He continues to note that this representation accords with appropriate invariance properties: it is invariant under refinements and coarsenings of the outcome space. Walley's representation, considered in isolation from the rest of his system, is unobjectionable. It is equivalent to the representation of completely neutral support (5). Wherever the strength I appears in (5), for example, Walley has the functionally equivalent pair of upper and lower probabilities, 0,1. Unlike the Shafer-Dempster vacuous belief function, contingent propositions are distinguished from the contradiction, in that both lower and upper probabilities of zero are assigned to the contradiction.

While this particular representation of complete ignorance is successful, there are other problems with Walley's system, however. Notably, he derives his quantities in the de Finetti tradition as previsions associated with betting scenarios through which some sort of universality of applicability is suggested. My concerns over this betting approach will be addressed below.

More generally, more exotic problems in inductive inference will present continuing challenges to aspirations of universality for all systems of imprecise probability. We will see some in coming chapters. It is not clear now how these systems can accommodate the different

---

[15] A measure is superadditive if the value assigned to a disjunction of mutually incompatible outcomes A v B is greater than the sum of the values assigned to A and B individually.

sectors of the logic native to the infinite lottery. As detailed in a chapter to follow, the sectors are divided into finite sets of outcomes, infinite-co-infinite sets of outcomes and infinite-co-finite sets of outcomes, each with their own distinctive structures. If some form of probabilistic account is to be preserved, the three sectors would appear to need both infinitely small and infinitely large probabilities. Still more serious is the challenge of recovering the logic native to quantum systems, as sketched in the chapter "Quantum Inductive Inference," for the basic structure of that logic is not a real valued function but an operator in a Hilbert space.

As long as theories of imprecise probability implement an inductive logic, they will place contingent constraints on the domains to which they can apply. That means that we should always expect new systems to arise to which their inductive logic does not apply. The cycle of extension and counterexample can continue without end, unless the theory of imprecise logic is so weakened by generalization that it places no factual restriction on the domains to which it applies. However then the theory has ceased to implement an inductive logic. It will implement only the requirement of logical consistency.

10. All Proofs of the Necessity of Probabilities are Circular

One of the more appealing aspects of the Bayesian approach is that its proponents have systematically taken on the burden of demonstrating that their approach is the uniquely correct one. The efforts at proof go back at least as far as the "Dutch book" arguments of Ramsey (1926) and de Finetti (1937) and their expansion by Savage (1954). Other approaches include the identification of necessary conditions and their consequences through representation theorems, such as developed by Cox (1961) and Jaynes (2003). More recent approaches, such as recounted in Pettigrew (2016), focus directly on the notion of accuracy as measured by scoring rules.

The literature is energetic. Existing approaches are subject to continuing amendment and expansion; and new approaches are offered. Optimists will see this as a proper and ever-improving response to a worthy problem of the first order. My reaction is more pessimistic. The ferment is the inevitable outcome when a literature sets itself an unattainable task. No proposal proves sustainable, but there is always the hope that a new approach might escape the problems that beset the last one.

That the goal is unattainable follows from the material approach to inductive inference. The proofs seek to establish the necessity of probabilities, that is, that objective degrees of inductive support or subjective degrees of belief must be probabilities. It has already been argued in Section 3 above that this is a contingent proposition. It may be true or false. It is not a necessary truth, demonstrable by pure logic alone.

It then follows directly that all proofs of the universal necessity of a probabilistic inductive logic, objective or subjective, must be circular. For all such proofs are logical deductions. They start with premises and from them deduce the conclusion sought. It is a basic fact of deductive logic that these premises must be at least as strong logically as the conclusion sought. Since the necessity of probabilities is not an a priori truth of logic, it follows that the premises of any demonstration of the necessity of probabilities must already contain exactly that necessity as contingent propositions, in one form or another.[16]

Seen in this perspective, there is a simple procedure for undoing all purported proofs of the necessity of probabilities: one merely needs to explore the premises of the proof and uncover the disguised presumption of probabilities. No matter how natural and comfortable the proof's starting points, no matter how congenial and convincing they may appear initially, the proof will depend on contingent premises that presuppose precisely what is to be proved. One then sees that the proofs are, in the best case, no better than merely positing probabilities in the first place. In a worse case, the premises are logically stronger, so one must assume more than the necessity of probabilities in order to derive the necessity of probabilities. In that case one is better off positing probabilities directly in the first place.

There is a dominance argument implicit in these last observations. If our interest is to minimize risk of error in an attempt to vindicate probabilities, we are never better off using one of these proofs. That is, merely directly positing probabilities weakly dominates, in a game theoretic sense, any justification by a proof. The only possible gain from the proof is a psychological one: one might find the premises posited by the proof more intuitively congenial, even if they risk being logically stronger than the goal.

This way of approaching the proofs casts a different light on the activity of the vindicators of probabilities. For nearly a century, their efforts have produced a flourishing literature that never quite produces the final, definitive demonstration. Rather the community of vindicators finds itself forever dissatisfied with the latest vindication. Sometimes new avenues are explored. Sometimes, the dissatisfaction results in a regress: a quest for further demonstrations that would establish the premises of the most recent demonstration. This regress cannot end well, for each further demonstration merely faces the same challenge anew: it must

---

[16] While I know of no such efforts, one might seek to show the universal necessity of a probabilistic inductive logic through a demonstration that itself employs inductive inferences. These efforts would face a dilemma. If the inductive inferences used are not probabilistic, it is conceded at the outset that some inductive inferences are not probabilistic. If the inductive inferences are probabilistic, then it must be shown that this particular probabilistic demonstration of the necessity of probabilities is not viciously circular.

find new, contingent premises from which to derive the old ones; and then it will have to justify these new premises. If the dissatisfaction is deep enough, it will seek another approach. This is a regress of reasons that cannot end in the proof sought.

We can now see that the continuing pain is not the result of some maddening inability of the vindicators to find just the right premises for their demonstrations. Rather it is the inevitable result of the awkward fact that there are no premises truly adequate to the task. The best a vindicator can do is to proceed from an assumption that probabilists will find intuitively appealing, since the assumption is equivalent to or logically stronger than the presumption of probabilities. That same assumption will appear arbitrary and even uncongenial to someone who is antecedently unconvinced of the necessity of probabilities.

The future of the vindication project is easy to predict. Like the circle squarers and angle trisectors of old, the vindicators will be trapped perpetually in the frustrating cycle of promising avenues; proofs that finally seem to succeed; and then the unhappy recognition that the latest proof falls just a little a short. If they persist, it must be so. The escape from the trap lies in the recognition that there is no necessity to probabilities.

Might one worry that this mode of objection is made too easily? If it works, might it not be able refute any demonstration of any proposition in philosophy? This worry is easily set aside, for this mode of objection applies only when a deductive proof is offered for a contingent proposition. Then it is cogent and should be applied.

For example, consider a theist who offers a deductive demonstration of the necessity of God's existence, such as Anselm's ontological argument. What reply can be given by a skeptic who holds the assertion of God's existence to be a contingent proposition? The skeptic would proceed as I have with the contingency of probabilities. Assuming its steps are valid, the skeptic would look at the premises of the theist's demonstration and expect to find contingent premises that are logically at least as strong as the necessity of God's existence. The theist would be untroubled by the display of these contingent premises. They would merely be a reformulation, possibly in logically stronger form, of what the theist already believes. The skeptic however would object that, precisely because of this, the demonstration is no demonstration at all, but is assuming what is to be proved.

11. Illustrations of Circularity

The last section established as a generality that all vindications of probabilities will prove circular and it predicted a manifestation of the failure in a regress of reasons. The exercise now is to find these circularities in the standard vindications. The literature in vindications is so large that it is impractical to cover it in any details. Therefore I have chosen to examine a recent,

presently popular vindication in greater detail. The scoring rule or accuracy based vindication is driven by a single, intuitively appealing idea. It suggests that there is a unique way to distribute our beliefs such that we cannot improve their accuracy, whichever circumstance may prove the true one. That distribution is probabilistic. Treating this case adequately proved to be a lengthy undertaking and has required a separate chapter, which follows this one. That chapter illustrates how the presumption of probability resides in the particular choice of the scoring rule used to measures accuracy. That choice must be carefully fine-tuned, else the approach fails to return probabilities. We shall then see how efforts to protect the fine-tuning from suggestions of circularity trigger just the doomed regress predicted above.

Here we shall take a briefer look at two other attempts to vindicate probabilities. We shall see that each presumes the very thing sought.

12. The Dutch Book Argument

12.1 The Betting Scenario

The Dutch book argument or arguments, if we separate out various forms of them, derive initially from Ramsey (1926) and de Finetti (1937). They have been a mainstay of the subjective Bayesian approach for decades.[17] The argument begins with the assertion that beliefs must be manifested operationally. The method chosen is to offer agents various bets and to determine their beliefs from which bets they accept and refuse. The argument then takes on a normative[18] burden: if—and only if—the beliefs manifested do not conform with the probability calculus, then it is possible, the argument goes, to offer the agent a combination of bets that results in a sure loss. That combination is the "Dutch book." Beliefs that allow this sure loss are disparaged as incoherent and reflect the supposed irrationality of non-probabilistic beliefs.

The central structure of the argument is the wager offered. The stake S is the sum of money or some other valuable of similar type associated with the bet. How it is distributed is decided by the presently unknown truth or falsity of some proposition A. In a bet "on" A, the agent pays a price qS for the possibility of gaining $S>0$ if A turns out to be true. Otherwise, if A is false, the agent simply loses the price. In sum:

---

[17] For recent surveys of a very extensive literature, see Hájek (2009) and Vineberg (2016).

[18] The normative element is essential. The system so narrowly constrains an agent's possible responses that it is a dismal means of ascertaining beliefs non-coercively.

Proposition A is true.	Agent gains $S - qS$
Proposition A is false	Agent gains $-qS$

Table 1. Payoffs of a bet "on" A ($S>0$) and "against" A ($S<0$)

This arrangement is reversed for a bet "against" A. It is most simply implemented by selecting a negative S and using the same payoffs as in Table 1. The full analysis requires an additional assumption to which we will return shortly:

> *Existence of a fair bet.* For any proposition A, for each agent, there is a "fair" betting quotient q such that the agent is willing to accept either side of the bet: "on" A or "against" A. This betting quotient measures the agent's strength of belief in A.

The main result is that failing to conform the betting quotients to the axioms of the probability calculus allows a Dutch book to be made against the agent (Dutch book theorem); and that conforming the betting quotients to the axioms makes it impossible for the Dutch book to be made (converse Dutch book theorem). A simple illustration does not even require the notion of a fair bet. Avoidance of sure loss immediately precludes $q>1$. For if $q>1$, a bet on any proposition A leads to a loss of $S(1-q)<0$ if A turns out to be true; and a loss of $-qS<0$ if A turns out to be false.

12.2 The Dubious Presumption

How does this construction presume probabilities? The principal, tendentious presumption is laid out in plain sight at the very start. In requiring agents *always* to express their beliefs in terms of monetary bets, accepted or refused, it forces agents to represent their beliefs on a single numerical scale. The betting quotient q has to be a real number, else the payoffs $S - qS$ and $-qS$ cannot be formed. Since there is so much more detail to come, it is easy to treat this presumption as an unimportant preliminary and to skip past it. That is mistaken if one wants to understand which are the strongest assumptions underlying the Dutch book argument. Once proponents of the argument can get us to accept that beliefs are measurable on a real number scale, most of their work is done.

The arguments for this presumption are weak in relation to the strength of what it asserts. De Finetti (1937, p. 139) pretends the assumption is innocuous. It is, he says:

> …the trivial and obvious idea that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event.

Ramsey (1927, p. 166) recognized that this idea is not nearly so innocent. "It is a common view," he conceded "that belief and other psychological variables are not measurable…" He recognized

that something stronger is needed and asserted that, without some measurement protocol, meaninglessness threatens (p. 167):

> …degree of a belief is just like a time interval; it has no precise meaning unless we specify more exactly how it is to be measured.

Here Ramsey echoes operationist sentiments of growing popularity in the 1920s. Bridgman (1927) was then writing his manifesto of operationism. It used special relativity, including its treatment of time, as his motivating example (Ch.1). Perhaps Ramsey's remark on time alluded to this example. At the same time in psychology, behaviorists were urging the elimination of invisible thoughts and ideas in favor of observable behaviors.

Nearly a century later, operationism and behaviorism have long fallen from favor. They proved unable to deliver accounts that match the richness of complex physical theories and mental content. The deepest problem with operationism was its core assertion. It is, in Bridgman's (1927, p.5, his emphasis) words:

> In general, we mean by any concept nothing more than a set of operations; *the concept is synonymous with the corresponding set of operations.*

It is a false assertion. Concepts are not synonymous with the operations that measure them. Time is not the ticking of a clock; or length the laying out of a ruler; or mass the extension of a spring in a weighing scale; or electric current the deflection of a needle in an ammeter. Correspondingly, belief is not the behavior of accepting or refusing bets. To reflect a widespread objection[19] to the Dutch book approach: beliefs have cognitive goals concerning learning the truth; betting behaviors have pragmatic goals in the maximizing of one's fortune. Supposing otherwise risks oversimplifications comparable to those that doomed operationist analyses elsewhere. Here we might imagine the (possibly fictional) enlightened Buddhist who has no material desires. Such a figure would, under these operationist strictures, be incapable of holding beliefs.

While concepts are not operations, there is still some value in asking how something might be measured, as long as we do not infer too hastily to meaninglessness if the operations prove elusive. What operations might measure strength of belief? Here we face the awkward realization that only one operation has been proposed: measurement through monetary bets accepted or refused. Why must we accept just this? Why is money the measure of belief?

The answer is that nothing forces the acceptance in general. However there are quite specific circumstances in which we are forced to make money the measure of belief. The most obvious arises if we are wagering in a casino or racetrack. Others arise if we are buying or selling insurance. We must determine what premium is appropriate now as insurance against

---

[19] Recounted in Weirich (2010, p. 246).

some uncertain, future harm whose gravity will be measured monetarily.[20] In the financial futures market, one can buy a contract that allows purchase of some asset at a fixed price at some later date. For example, an airline, fearing an increase in jet fuel prices, might buy a contract that enables purchase of jet fuel at present prices, but at a later date. Whether the later purchase will be made depends on the unknown of whether jet fuel prices will rise or fall. Thus pricing the contract requires the same sort of judgments over uncertainties as insurance and wagering.

Something close to a fair bet is also realized in these circumstances. In casinos and racetracks, every wager bought is sold by someone. In insurance, every policy, bought by someone, is sold by someone else. In the futures market, every contract bought by one trader is sold by another. These transactions would comprise fair bet if we neglect a small spread between the buying and selling prices and the house's small margin in casino gambling.

Viewed materially, if we are in any of these circumstances, then pragmatic goals will force us to reason inductively as the framework of the Dutch book argument requires. The facts of the circumstances warrant the resulting logic. It provides a nice illustration of how the material theory of induction is applied. When we move to other circumstances, however, when facts of this type are missing, nothing warrants an inductive logic in which strengths of support must conform to defensive gambling strategies. Viewed material, Dutch book argumentation fails to establish the universal rationality of probabilistic inference precisely because the factual presumptions of the Dutch book scenarios do not hold universally.

Hájek (2008) has proposed an amusing device that we can use to underscore the dependence of the logic on the background conditions. If an agent has incoherent betting quotients, a benevolent bookie can offer the agent a combination of bets that assures a gain, a "Czech book." If, for example, the agent's betting quotient q is greater than one for some proposition A, the bookie would offer the agent a bet against A. Since $S<0$, the agent will make a gain $S(1-q)>0$ if A turns out to be true; and a loss of $-qS>0$ if A turns out to be false. That is, if one finds oneself in the clutches of a benevolent bookie, coherent betting quotients would be the only thing preventing the benevolent bookie providing you an assured gain!

12.3 The Rationality of Refusing to Bet

A sharper expression of the coercive presumption of the betting scenario is provided by a common response to it: it can be an expression of rationality for agents simply to refuse to bet. In the abstract, this refusal may seem like a crafty evasion. That it need not be is easier to see if we consider how a bookie might seek to force a Dutch book on someone whose beliefs conform

---

[20] Starting in 1931, de Finetti had worked for an insurance company. Might this explain why he found it trivial that monetarily rewarded betting behavior measures belief?

with completely neutral support (5). Consider the case of three mutually exclusive outcomes A_1, A_2, A_3, such as arises in the wine and water problem. An agent whose credences conform with (5) would judge all of the following to be equally supported:

$$A_1, A_2, A_3, \text{not-}A_1 = A_2 \lor A_3, \text{not-}A_2 = A_1 \lor A_3, \text{not-}A_3 = A_1 \lor A_2$$

Assume *per impossibile* that there is some bet on A_1 that is acceptable to the agent as fair. Since the agent regards not-A_1 as equally supported, the agent will accept as fair a bet on not-A_1 with the same payoffs. Since a bet "on" A_1 is just the same as a bet "against" not-A_1, one can see that the two net payoffs, $S(1-q)$ and $-qS$ must be numerically equal, but different in sign.[21] For simplicity, assume that the bet "on" A_1 pays a net of 1, if A_1 is true, and -1, if A_1 is false. The agent will judge similar bets fair for A_2 and A_3.

The three bets on A_1, A_2 and A_3 combined form the Dutch book in Table 2:

	Bet on A_1 pays:	Bet on A_2 pays:	Bet on A_3 pays:	Net payoff
A_1 is true.	+1	-1	-1	-1
A_2 is true.	-1	+1	-1	-1
A_3 is true.	-1	-1	+1	-1

Table 2. Dutch Book for an Agent with Beliefs conforming with Completely Neutral Support

What we cannot conclude from this Dutch book is that the assignments of support by the agent are irrational. They were determined as the only assignments compatible with the invariances of the system in question. If this Dutch book impugns the rationality of these assignments, then all we can conclude is that no rational treatment of systems like von Mises' wine and water is possible.

The obvious alternative is to recognize that someone who harbors assignments of support like those of (5) should not accept bets in accord with the rules specified in the Dutch book gambling scenario. For such an agent's credences are in conflict with the assumptions of the scenario. The irrationality lies not in the assignment of beliefs but in the indiscriminate

---

[21] Suppose that the bet "on" A_1 pays a net of $X>0$, if A_1 is true, and $Y<0$, if A_1 is false. Then the bet "against" A_1 pays a net of $-X<0$, if A_1 is true, and $-Y>0$, if A_1 is false. The bet with the same stakes "on" not-A_1 pays $X>0$, if not-A_1 is true, and $-Y>0$, if not-A_1 is false. These last two bets can only be same if, $X=-Y$.

acceptance of bets devised using those rules. Here I concur fully with the assessment of Bacchus et al. (1990, pp. 504-505) who argue:

> …that an agent ought not to accept a set of wagers according to which she loses come what may, if she would prefer not to lose, is a matter of deductive logic and not of propriety of belief.

12.4 Circularity in the Notion of a Fair Bet

Consider again a key assumption in the Dutch book argument: for any proposition A, an agent can find a fair bet with payoffs comprising those of Table 1; and the associated betting quotient q is the strength of belief in A. This may seem like a benign preliminary before the real work of assembling a Dutch book begins. However it is not. That assumption in effect already has the axioms of the probability calculus built into it and an excursion in repeated betting shows it.

To see it, consider some set of atomic propositions $A_1, A_2, \ldots, A_n$ and their Boolean combinations over which an agent distributes belief. Imagine that there are repeated scenarios in which there is a like set of propositions over which the agent distributes the same beliefs. Call the corresponding propositions of the form A_1 in each scenario "like propositions"; and so on for the remaining $A_2, \ldots, A_n$.

The obvious example is provided by the two propositions that a tossed coin shows heads (A_1) or that it shows tails (A_2). The repeated scenarios are then just independent tossing of many coins. For another example, we might consider the propositions that someone named in a telephone directory was born on Monday (A_1), or born on Tuesday (A_2), or born on some Boolean combination of days, such as (not-Monday and not-Friday) = (not-A_1 & not-A_5.) We create scenarios with identical beliefs over like propositions by scanning down a list of names in a telephone directory and asking after the birthday of each person named.

Since the agent has the same belief in the truth of each of the like propositions in the corresponding sets, the agent can execute the same bet on each like proposition. That is the agent's betting quotient for proposition A_i in each scenario is the same value q_i for what the agent judges to be a fair bet with the same fixed stake S_i in each case.[22] Assuming that there are W_i wins and N-W_i losses among N bets, we find for the bets on proposition A_i:

$$\text{total payoff}_i = W_i(1-q_i)S_i - (N-W_i)q_iS_i = (W_i - Nq_i)S_i$$

---

[22] What follows is an analysis concerning the atomic propositions. An analogous analysis can be applied to the propositions that are Boolean combinations of them.

From this we compute the average payoff per wager in terms of the frequency $r_i = W_i/N$ with which the propositions turn out to be true:

$$\text{average payoff}_i = (W_i - Nq_i)S_i/\text{N} = (r_i - q_i)S_i$$

To proceed we need to separate two cases. These frequencies r_i may or may not stabilize to definite limiting values as N grows indefinitely large. In the first case, we can define the limiting frequency as

$$p_i = \lim_{N \to \infty} r_i$$

It would be natural to identify this limiting frequency p_i with the probability of truth among the propositions A_i, for, if there is such a probability, the law of large numbers assures us that, with probability one, it will be revealed as this limit. However to arrive at the results that interest us, we do not need to do this. We can simply treat the p_i as parameters that have the specific property of importance here. Since they are derived from relative frequencies, they conform with the axioms of the probability calculus. That is, they are non-negative, additive for mutually exclusive outcomes and normalize to unity. For example, the limiting frequencies of truth p_i among the repetitions of the atomic propositions A_i always sum to unity:

$$p_1 + p_2 + \ldots + p_n = 1.$$

We can now see that the following two propositions are equivalent where the same set of betting quotients q_i is indicated in each proposition:

(a) There are fair betting quotients q_i such the agent fares equally well by making all the bets over A_i "on" bets with $S_i>0$; or by making all the bets over A_i "against" bets with $S_i<0$.

(b) There are betting quotients q_i that equal the limiting frequency of truth p_i among the propositions A_i (so that these betting quotients conform with the axioms of the probability calculus).

To infer from (a) to (b), note that "fares equally well" means that "on" and "against" betting yields the same results concerning payoffs. It follows that the limiting average payoff must be unaltered when we merely change the sign of S_i from positive to negative, where

$$\text{Limiting average payoff}_i = (p_i - q_i)S_i$$

If we interpret the parameters p_i as probabilities, this limiting average payoff is just the expected payoff per bet. Now, the limiting average payoff is linear in S_i. So it can only remain unchanged under an alternation of the sign of S_i if it is zero. That is $(p_i - q_i)S_i = 0$. It follows immediately that $p_i = q_i$. That is, we have inferred (b). The reverse inference from (b) to (a) follows by taking the steps of the inference in reverse order: $p_i = q_i$ entails that both total and average payoffs are zero, so that bets "on" and "against" are equally attractive.

In the second case, there is no stable limit to the frequencies r_i as N grows indefinitely large. This is an uncommon case, but it can arise. We shall see it, for example, arising for outcomes of drawings from an infinite lottery in the chapter "Infinite Lottery Machines." It is the case that is unfavorable to probabilities and thus we might not expect that the assumption of the existence of fair betting quotients might still drive the quotients toward conformity with the axioms of the probability calculus. However they still do so in the following sense.

Since the frequencies r_i have no limiting value, there is no unique value for them unless we specify the specific number of repetitions N. Once it is specified, fairness of the bet on proposition A_i is implemented if the agent can pick a betting quotient q_i that matches the actual frequency of truth r_i among the set of like propositions A_i in the N repetitions. For then the average payoffs are the same for both "on" and "against" bets. Any other value of q_i will favor either the bets "on" or "against" the like propositions according to whether $q_i > r_i$ or $q_i < r_i$.

As the number of repetitions varies, the particular target set of frequencies of truth r_i will vary. But what will not vary is that the target set for the betting quotients q_i is always a set of frequencies. Frequencies obey the axioms of the probability calculus, but with the added restriction that they are rational number valued. Thus we have weaker analogs of the equivalent propositions (a) and (b) for the same set of betting quotients q_i in each proposition:

(a') For some fixed set or repetitions N, there are fair betting quotients q_i such the agent fares equally well by making all the bets over A_i "on" bets with $S_i > 0$; or by making all the bets over A_i "against" bets with $S_i < 0$.

(b') For some fixed set or repetitions N, there are betting quotients q_i that equals the frequency of truth r_i among the propositions A_i (so that these betting quotients conform with the axioms of the probability calculus).

The proof of the equivalence of (a') and (b') is analogous.

In sum, for both cases, the assumption that there are fair betting quotients in the context of repeated betting scenarios is equivalent to assuming that these betting quotient behave like frequencies, that is, they conform with the axioms of the probability calculus. Thus one should not think that the assumption of fair betting quotients is an innocent background assumption. It does not merely provide a context in which Dutch book argumentation can prove that credences must conform with the axioms of the probability calculus. Rather conformity with those axioms is already tacitly presumed by them. All the Dutch book argumentation does is to make that conformity visible.

This outcome may be untroubling to someone who already believes that credences must be probabilistic. Why be troubled with a demonstration that just clarifies the probabilist's

commitments? If, however, you are someone like me who does not believe that credences must be probabilistic, you will find this result damning. What was supposed to be a demonstration of the incoherence of non-probabilistic beliefs turns out to be an exercise in circularity. Probabilities are demonstrable merely because they were introduced covertly in an assumption of the argument at the outset.

12.5 The Regresses Begin

The prediction of the general analyses in earlier sections is that recognition of weaknesses in an attempted proof of probabilities leads to a regress. One form is a successive weakening of what is sought to be proved. This form of regress is well underway for Dutch book arguments. For it has been long recognized in the literature that the assumption of fairness is arbitrary and can be discarded without compromise to the rationality of the enterprise. This recognition is at least a half century old, extending as far back as Smith (1961). It is the basis of the analysis of Walley's (1991) treatise *Statistical Reasoning with Imprecise Probabilities.*

Following Walley (1991, p. 28), it may be quite prudent for an agent to refuse to admit any bet over some proposition A as fair. Rather, the agent may be willing to accept a bet "on" A with a maximum betting quotient of q_{lower} and be willing to accept a bet "against" A with a minimum betting quotient of q_{upper}. If the two are equal, then they comprise a fair bet over A. If $q_{lower} < q_{upper}$, then the agent is more cautious in the agent's betting behavior. No Dutch book can be made against such an agent. The agent's belief in A is no longer a single probability, but an interval bounded by a lower probability equal to q_{lower} and an upper probability equal to q_{upper}.

That betting quotients $q_{lower} < q_{upper}$ betoken caution becomes most evident in the extreme case in which $q_{lower} = 0$ and $q_{upper} = 1$. In this extreme case, the agent is willing only to accept individual bets for which no loss is possible.[23] The interval of probabilities is maximally large, bounded by 0 and 1. For this reason, Walley (1991, p. 66) associates this state with vacuity or maximum ignorance.

To support this discarding of the necessary existence of fair bets, Walley (1991, p. 3) decries what he calls "the Bayesian dogma of precision"—"that uncertainty should always be measured by a single (additive) probability measure…" He continues:

---

[23] The agent is willing to make a bet "on" A only if the bet pays a net of S-0.S = S>0, if A is true; and with a payoff of -0.S=0 if A is false. The agent is willing to accept a bet "against" A only if the bet pays S – 1.S=0 if A is true; and with a payoff of -1S = -S > 0 if A is false, since S<0 for an "against" bet.

For example, de Finetti assumes that for each event of interest, there is some betting rate that you regard as fair, in the sense that you are willing to accept either side of a bet on the event at that rate. This fair betting rate is your personal probability for the event. More generally, we take your lower probability to be the maximum rate at which you are prepared to bet on the event, and your upper probability to be the minimum rate at which you are prepared to bet against the event. It is not irrational for you to assess an upper probability that is strictly greater than your lower probability. Indeed, you ought to do so when you have little information on which to base your assessments. In that case we say that your beliefs about the event are indeterminate, and that (for you) the event has imprecise probability.

This work is motivated by the ideas that the dogma of precision is mistaken, and that imprecise probabilities are needed in statistical reasoning and decision.

There are two ways to understand the import of this relaxation of the conditions of the betting scenarios. The correct way, in my view, is merely to regard the various betting scenarios envisaged as circumstances that may or may not arise in different domains. There is no necessity for their implementation everywhere. All we can say is that *if* an agent is in a circumstance in which the assumptions of the scenario are realized, *then* reasoning inductively according to the prescribed system is their best course. Such circumstances arise, as noted above, in the insurance and futures market. Indeed the slight spread between the buying and selling prices in both cases suggests that Walley's imprecise logic is the appropriate one.

The incorrect way to understand the import of this relaxation is to think of it as a successful purging from the Dutch book analysis of an unwarranted element, the necessary existence of a fair bet, so that the analysis that remains is universally applicable. That would merely be to replace the dogma of precision by the dogma of imprecision. For there is no necessity in the presumption of strict upper and lower limits on the betting quotients or even that having beliefs requires their operational manifestation in betting behavior. With this understanding, we have taken the first step in the regressive weakening of what is sought to be proved, as described in Sections 4 and 9 above.

The other sort of regress arises when we retain what is sought to be proved, but seek to strengthen the grounds used in the proof. This is how I see Savage's (1954) decision theoretic proof of probabilities. It seeks, as does the Dutch book argument, to infer from an agent's preferences to the beliefs that must conform with them and thereby show them necessarily to be probabilistic. Savage does acknowledge (p. 4) inspiration from de Finetti's (1937) work. Now de Finetti simply posits certain betting behaviors and his posits are, as we have seen above, quite

susceptible to challenge. I read Savage's analysis as an attempt to provide a more secure grounding for this approach.

Savage's full theory is based on seven postulates. Since they entail the same contingent result that beliefs are probabilities, they must contain that contingency in one form or another. Once again, careful scrutiny should reveal its presence. Because of the complexity of Savage's system, detailed analysis is precluded here. However we can discern the direction of the analysis by considering just the first postulate. It asserts, in effect, (p. 18) that the relation of preference over acts is a total order, which means that it is antisymmetric and transitive.

Consider the transitivity of preference. It says that if you strictly prefer A to B and B to C, then you must prefer A to C. (Antisymmetry precludes you also strictly preferring C to A.) This form of transitivity is important in the system. It provides an order that, when filtered through the other postulates, orders strengths of beliefs and eventually enables them to be real valued. Savage provides no argument to preclude intransitivity when the postulate is introduced. He merely announces that "the definition of preference suggests …" (p.18). If you are antecedently disposed towards probabilities, it is quite comfortable to accept the suggestion and let the reasoning lead to the result you expect. However if you are not so disposed, you will have seen no good reason in the account that precludes intransitive preferences. Say I prefer eating apple pie to cherry pie and cherry pie to apricot pie. Aside from unsupported declaration in the definition of preference or in notions of rationality, nothing precludes me from preferring apricot pie to apple pie. But that would be an intransitive set of preferences.

It was soon recognized that more was needed if this prohibition on intransitivity was to be sustained. That is, the regress of reasons continues. The instrument that sustains it came to be known as the "money pump" argument, which appears in the literature as early as Davidson et al. (1955, pp. 145-46). Assume an agent harbors intransitive preferences, captured compactly by the obvious notation: $A>B, B>C, C>A$. Presumably the agent could be induced to trade a C to gain a B, while paying some small price, such as \$1; and a B to gain an A, for \$1; and an A to gain a C for \$1. The net effect is that the agent has paid \$3 to be returned to the original C. This, it is supposed, makes the intransitive preferences "irrational."

Once again we have an argument that can only be convincing to someone who already believes that there is some irrationality in intransitive preferences. Someone who does not believe this will have no trouble seeing that the irrationality is not in the intransitivity of the preferences. Rather it lies in the agent engaging in free trading with a second commodity (money) over which the agent's preferences are transitive. That trading behavior is dangerous and should be avoided; and that is all the money pump argument shows. Maher (1993, p. 36) puts it well:

This is such a simple and vivid argument that it is a pity it is fallacious. But fallacious it is. The fallacy lies in a careless analysis of sequential choice. The argument assumes that someone with intransitive preferences will make each choice without any thought about what future options will be available, yet this is not in general a rational way to proceed. For example, …"

13. Necessary Conditions

As another briefer illustration of the inevitable failure of the proofs of necessity of probabilities described in Section 10 above, consider the approach taken by Cox (1961) and Jaynes (2003). The general approach is both elegant and appealing. Necessary conditions are laid down for a structure called "$i|h$" (Cox) or "$A|B$" (Jaynes) which represents the strength of support of the first (i or A) afforded by the second (h or B). From them, by some simple but powerful functional analysis, the computational rules of the probability calculus are derived.

Precisely because these computational rules can be derived, the assumptions used must be logically at least as strong as them. Since the conclusion is contingent, so also are the assumptions. Any hope that the assumptions might somehow be self-evident will fail under scrutiny. Sustaining the proof will then trigger a regress of reasons each of which fails in the sense that the new reasons themselves need further support. We shall see this regress begin with Cox first positing the necessary conditions with short justifications. The inadequacy of the justifications becomes clear. Jaynes then intervenes and provides a stronger justification; and then sometimes when that stronger justification proves inadequate, yet another is provided, but still without arriving at a satisfactory end point.

There are three necessities: that the strengths are real values and what Jaynes calls the sum and product rules. We shall look at each in turn.

First, Cox (1961, p. 1) introduces the idea that the strengths are real valued with some rather casual remarks about their measurability. As an analogy he mentions the measurability of the pitch of a stairway.[24] Jaynes is rightly not satisfied with such a casual development. He makes the requirement explicit as his first desideratum (p. 17)

(I) Degrees of plausibility are represented by real numbers.

Jaynes first seeks to establish that it is satisfied by means of his parable of a robot (pp. 8-9) who will compute with the degrees. He then asserts (p. 17):

---

[24] Cox's (pp. 29-34) later remarks on measurement pertain not to whether the strengths have real valued magnitudes, but whether they can be assessed with precision.

Desideratum (I) is practically forced on us by the requirement that the robot's brain must operate by the carrying out of some definite physical process.

Of course this is incorrect. A robot can represent and compute with all sorts of magnitudes and relational structures. To presume otherwise suggests willful ignorance, if someone has a minimal understanding of computers. To establish that the magnitudes treated are real numbers, we must assume a quite extensive list of specific properties including a transitive order ("greater than") and universal comparability under this order of all the magnitudes.

Tribus (1969, Ch. 1) develops a similar account that includes the parable of the robot. He reports (p. 6) drawing on Cox (1961) and unpublished course notes by Jaynes. His remark (p. 13) is:

The only general way in which objects may be compared with one another is to assign to the objects a real number. The real number system provides the only scale of universal comparability.

It is easy to see that one might let this pass if one already believes that the strengths of support must be probabilities. Otherwise it is baffling that such a claim could be made.

The regress of reasons continues. Jaynes presumably recognized the weakness of the robotic justification and included an Appendix (pp. 656-59) designed specifically to strengthen it. He notes that if an order on the strengths is transitive and universal, then, in the case of a finite outcome space, real valued degrees can be adapted to it. He proceeded to argue rather ineffectively for both transitivity and universality. Counterexamples to transitivity can be readily constructed, as in Norton (2007a, pp. 149-50). Keynes (1921, Ch. III) had long before realized that we must take seriously the possibility of incomparable degrees. More troublesome is that transitivity and universal comparability are insufficient to assure real values strengths.

Cox's second necessity is expressed as (p. 3):

The probability of an inference on given evidence determines the probability of its contradictory on the same evidence.

Cox's justification is brief. He gives a few simple examples (p. 2) and announces that "in this all schools can agree." Jaynes' treatment is similarly hasty and incomplete. He declares (p. 30) "The plausibility that A is false must depend in some way on the plausibility that it is true." He proceeds immediately to conclude the much stronger result that there must be a functional relation of dependence between $A|B$ and not-$A|B$ and even that "common sense requires [the function] to be a continuous monotonic decreasing function…"

Once again, all these suppositions can pass without objection if one already has the goal of additivity of strengths in mind. That is, one might imagine that these necessities are simply reduced descriptions of the rule in the probability calculus that $P(A|B) + P(\text{not-}A|B) = 1$. If one is not antecedently committed to this rule or something like it, these necessities will appear as

unfounded stipulations. One need only consider superadditive measures to find all the conditions laid down by Cox and Jaynes violated.

Cox's third necessity is (p. 4):

The probability on given evidence that both of two inferences are true is determined by their separate probabilities, one on the given evidence, the other on this evidence with the additional assumption that the first inference is true

Its content is more easily grasped if we give it in symbolic form, as does Jaynes (p. 25). The support for the conjunction of A and B on the evidence C, $(AB|C)$ is some function F of two other strengths $(B|C)$ and $(A|BC)$:

$$(AB|C) = F[(B|C) , (A|BC)] \qquad\qquad (6)$$

To someone remote from probability theory, this functional stipulation will appear quite arbitrary. To probabilists, it immediately calls to mind the product rule for forming conjunctions

$$P(AB|C) = P(B|C) . P(A|BC)$$

So, for them, it can pass as reasonable and even natural. Both Cox and Jaynes seek to establish this functional dependence by recalling informal sequences of inferences. If we are to infer to (A and B) from C, we might first establish from the truth of C that B is true. Then we would establish from the truth of (C and B) that A is truth and so also that the conjunction (A and B) is true. (For later reference, represent this as "$C \rightarrow B \rightarrow A \rightarrow AB$.") This sequence is one way that we might proceed deductively. Cox and Jaynes then declare that the functional dependence (6) follows since it mimics the same order of steps.

The inference is quite dubious. Indeed one of the lessons of twentieth century philosophy of science was that transferring properties of deductive inference over to inductive inference regularly produces incorrect rules. For example, if C deductively entails each of A and B separately, then C also deductively entails their conjunction. However the corresponding rule for induction fails. C may strongly support each of A and B separately, but actually refute their conjunction.

Presumably because they recognize the inadequacy of the arguments for (6) so far, Jaynes and Tribus (1969, Ch. 1) embark on a more elaborate demonstration. Its basic supposition is that $(AB|C)$ must be a function of some or all of the following four strengths only:

$$(A|C), (B|C), (B|AC), (A|BC).$$

They then argue that the only possibility is (6) or its equivalent form under relabeling, $(AB|C) = F[(A|C) , (B|AC)]$. Locally the argumentation is quite cogent. For example, $(AB|C)$ cannot depend functionally on just $(A|C)$ and $(B|C)$. For each of A and B may be strongly confirmed by C, but C may either confirm or even refute (A and B).

However this next attempt to buttress the functional dependence of (6) fails. For the assumptions used are still far too strong and likely only unobjectionable if one already accepts

the final result. The lacunae are both narrow and broad. Narrowly, consider the details of the functional dependencies. They infer that $(AB|C)$ can depend functionally on $(A|C)$ and $(B|AC)$; or it can depend functionally on $(B|C)$ or $(A|BC)$. These dependencies are analogous to the two deductive pathways "$C \rightarrow B \rightarrow A \rightarrow AB$" and "$C \rightarrow A \rightarrow B \rightarrow AB$." Since either individually suffices in the deductive case, Jaynes seems to presume that either will also suffice in the inductive case. That certainly does not follow since the analogies between deduction and induction are fragile. They have not ruled out the case that both inductive pathways must enter into the functional dependence, which would mean that $(AB|C)$ is still a function of all four strengths listed.[25]

Taking a broader, synoptic view, the most obvious lacuna is the assumption that $(AB|C)$ must be a function of the four strengths listed. They might be related, but must the relationship be functional? Might there not be a more complicated relationship? Perhaps there is one that involves some auxiliary quantity and the strengths $(A|C)$ are derived from them? Or might there simply be no definite relation at all? This last possibility would then mimic the situation with superadditive measures. They decouple the values of $(A|C)$ and (not-$A|C$), so that there is no functional relation between them. Each value of $(A|C)$ may be compatible with many (not-$A|C$); and conversely.

14. Conclusion

The approach taken in this chapter pursues the two lines of criticism indicated. There are more grounds for hesitation over probabilities. My (Norton, 2011) reviews many of these. Perhaps the best known and most intractable of these problems is the problem of the priors. It arises from the need for a Bayesian analysis always to provide some prior probability, $P(H|B)$, antecedent to the consideration of evidence. The very fact that they must be provided in this way introduces an arbitrariness into the analysis that has been the bane or all forms of Bayesianism. Objective Bayesians try to find good reasons for picking a particular prior, such as through Jaynes' ill-chosen[26] maximum entropy principle. Subjective Bayesians try to avoid the problem

---

[25] Tribus (1969, pp. 16-17) has an argument against this possibility that appears flawed. He seems to argue that it is ruled out since $(B|AC)$ becomes ill-defined when C is not-A. But this sort of difficulty is routinely overcome by allowing that in some special cases the function is ill-defined.

[26] The principle tells us to distribute our prior probabilities as uniformly as the external constraints allow. Thus it is an extended form of the principle of indifference. If there are no constraints other than conformity with the probability calculus, maximizing entropy reduces to

by demoting the prior probability to mere opinion, which can be freely chosen. Theirs has proven to be a poor bargain since, once one allows opinion to be mingled with evidential warrant, they prove virtually impossible to separate.

The necessity for prior probabilities is a form of incompleteness of the inductive logic: those priors always supply inductive content that is beyond the reach of the evidence to be considered subsequently. One might imagine that it is a problem peculiar to the probability calculus so that the best escape is to find another calculus free of the problem. In recent work, I have shown that this escape fails. The sort of incompleteness that troubles the probability calculus must arise in a large class of calculi of induction that include all those we would reasonably entertain. An informal development of this result is provided below in the Chapter "Incompleteness of All Calculi of Inductive Inference."

What I have sought to establish in this chapter is that the probability calculus does not supply a universally applicable logic of inductive inference. The emphasis here is on the universal applicability. I do not doubt the utility of Bayesian analysis in specific domains in which background facts positively warrant it. My hope is that, if they have it, Bayesians can relinquish tacit commitments to the idea that "It's all probabilities." and to the notion that this idea solves the foundational problems of inductive inference. For then we can address these foundational problems anew and, it is to be hoped, find better solutions. Readers, of course, will know that I offer the material theory of induction as my solution to the foundational problem of the nature of inductive inference.

If we are loosening tacit Bayesian commitments, there is a second one that can be relaxed profitably. The general view seems to be that the probability calculus must be accepted or rejected as a whole. In contrast, I have urged in that we can be more selective. Norton (2007a) provides an axiomatization of the probability calculus, using familiar techniques. Its novelty is that it was designed explicitly to identify qualitative properties of support relations that may be employed selectively. The most important result is that there are two components in these qualitative properties, that the two are readily separated and that they can be deployed individually as circumstances demand.

The first is a property I call "addition" which captures the additivity of the calculus. It resides in a reciprocal relation between the support accorded to a proposition and to its negation. Addition is an appropriate property when degrees of support span from positive to negative support. It should be dropped, however, if neutral support is to be represented.

---

choosing the uniform probability distribution required by the original principle of indifference. This principle, as we have seen in Sections 6, 7 and 8, is an insecure basis for reasoning within the probability calculus since it rapidly produces results that contradict the calculus.

The second property, "Bayes property," provides the probability calculus with the updating dynamics characteristic of Bayesian analysis. It depends on a particular mode of updating in which the import of evidence is simply to refute disjunctive parts of the hypothesis logically incompatible with the evidence; and then to redistribute support uniformly.

Many of the successes of Bayesian analysis can be traced back to these properties. Since conditions may favor the use of one but not the other, their utility can only be increased if we decide to employ them separately, for then they can be used more widely. For example, the completely neutral support described in this chapter contradicts additivity, but it is compatible with the Bayes property. Thus an extension of the theory of completely neutral support will permit updating by a Bayesian-style dynamics.

References

Bacchus, Fahiem; H. E. Kyburg, Jr., Henry E.; and M. Thalos, Miriam (1990) "Against Conditionalization," *Synthese*, **85**, pp. 475-506.

Benétreau-Dupin, Yann (2015) "The Bayesian Who Knew Too Much," *Synthese*, **192**, pp. 1527-42.

Bostrom, Nick (2002) *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York: Routledge.

Bradley, Seamus (2016) "Imprecise Probabilities," *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/imprecise-probabilities/

Bridgman, Percy W. (1927) *The Logic of Modern Physics*. New York: MacMillan.

Cox, Richard T. (1961) *The Algebra of Probable Inference*. Baltimore: Johns Hopkins Press.

De Finetti, Bruno (1937) "Foresight: Its Logical Laws, Its Subjective Sources," pp. 134-74 in S. Kotz and N. L. Johnson, eds., *Breakthroughs in Statistics. Volume 1. Foundations and Basic Theory*. New York: Springer Verlag, 1992.

Davidson, Donald; McKinsey, J. C. C.; and Suppes, Patrick (1955) "Outlines of a Formal Theory of Value, I," *Philosophy of Science*, **22**, pp. 140-160.

Eva, Benjamin (manuscript) "Principles of Indifference."

Hájek, Alan (2008) "Arguments for–or against–Probabilism?" *British Journal for the Philosophy of Science*, pp. 793–819.

Hájek, Alan (2009) "Dutch Book Arguments," pp. 173-195 in Paul Anand, Prasanta K. Pattanaik, Clemens Puppe, eds., *The Handbook of Rational and Social Choice: an Overview of New Foundations and Applications*. Oxford : Oxford University Press.

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

Keynes, John Maynard (1921), A Treatise of Probability. London: Macmillan; reprinted, New York: AMS, 1979.

Kolmogorov, Andrey (1950) *Foundations of the Theory of Probability*. Chelsea: New York.

Maher, Patrick (1993) *Betting on Theories*. Cambridge: Cambridge University Press.

Norton John D. (2007) "Disbelief and the Dual of Belief," *International Studies in the Philosophy of Science,* **21**, pp. 231-52.

Norton, John D. (2007a) "Probability Disassembled," *British Journal for the Philosophy of Science*, **58**, pp. 141-171.

Norton, John D. (2008) "Ignorance and Indifference," *Philosophy of Science*, **75**, pp. 45-68.

Norton, John D. (2010) "Cosmic Confusions: Not Supporting versus Supporting Not,"*Philosophy of Science*, **77**, pp. 501-523.

Norton, John D. (2010a) "Deductively Definable Logics of Induction." *Journal of Philosophical Logic*. **39**, pp. 617-654.

Norton, John D. (2011) "Challenges to Bayesian Confirmation Theory," *Philosophy of Statistics, Vol. 7: Handbook of the Philosophy of Science*. Prasanta S. Bandyopadhyay and Malcolm R. Forster (eds.) Elsevier

Pettigrew, Richard (2016) *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Ramsey, Frank P. (1926) "Truth and Probability," in *The Foundations of Mathematics and other Logical Essays*, Ch. VII, pp.156-198, ed. R.B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company, 1931.

Savage, Leonard J. (1954) *The Foundations of Statistics*. John Wiley & sons. Revised ed., New York: Dover, 1972.

Shafer, Glenn (1976) *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.

Smith, Cedric A. B. (1961) "Consistency in Statistical Inference and Decision," *Journal of the Royal Statistical Society. Series B*, **23**, pp. 1-37.

Tribus, Myron (1969) *Rational Descriptions, Decisions and Designs*. New York: Pergamon.

Van Inwagen, Peter (1996) "Why Is There Anything at All?" *Proceedings of the Aristotelian Society* **70** (suppl.), pp. 95–120.

von Mises, Richard (1957), *Probability, Truth and Statistics*. London: George Allen & Unwin.

Vineberg, Susan (2016) "Dutch Book Arguments," *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2016/entries/dutch-book/.

Walley, Peter (1991) *Statistical Reasoning with Imprecise Probabilities.* London: Chapman and Hall.

Weirich, Paul (2011) "The Bayesian Decision-Theoretic Approach to Statistics," pp. 233-61 in Prasanta S. Bandyopadhyay and Malcolm R. Forster, eds., *Handbook of the Philosophy of Science. Volume 7: Philosophy of Statistics.* Amsterdam: Elsevier.