Probability Disassembled

John D. Norton

Center for Philosophy of Science and

Department of History and Philosophy of Science

University of Pittsburgh

www.pitt.edu/~jdnorton

ABSTRACT

While there is no universal logic of induction, the probability calculus succeeds as a logic of induction in many contexts through its use of several notions concerning inductive inference. They include *Addition*, through which low probabilities represent disbelief as opposed to ignorance; and *Bayes property*, which commits the calculus to a 'refute and rescale' dynamics for incorporating new evidence. These notions are independent and it is urged that they be employed selectively according to needs of the problem at hand. It is shown that neither is adapted to inductive inference concerning some indeterministic systems.

1

1. Introduction

No single idea about induction[1] has been more fertile than the idea that inductive inferences may conform to the probability calculus. For no other proposal has proven anywhere near as effective at synthesizing a huge array of disparate intuitions about induction into a simple and orderly

---

[1] The terms 'induction' and 'inductive inference' are used here in the broadest sense of any form of ampliative inference. They include more traditional forms of induction, such a enumerative induction and inference to the best explanation, which embody a rule of detachment; as well as confirmation theories, such as in traditional Bayesianism or Hempel's satisfaction criterion, which lack such a rule and merely display confirmatory relations between sentences. Since the assumptions of *Framework* (Section 3 below) lack a rule of detachment, the positive analysis of this paper uses the latter approach.

system. No single idea about induction has wrought more mischief than the insistence that *all* inductive inferences *must* conform to the probability calculus. For it has obliged probabilists to stretch their calculus to fit it to cases to which it is ill suited and to devise many ingenious but ill-fated proofs of its universal applicability.

This paper offers an alternative to this second idea. It is part of a larger project (Norton [2003], [2005]) in which it is urged that there is no single logic of induction, but many logics each adapted to particular contexts. The goal of the present paper is to understand why the probability calculus works so well as a logic of inductive inference, in the contexts in which it does; and to try to demarcate when it does not. To this end, the paper draws on an extensive, existing literature in presenting an axiom system for the probability calculus. However, unlike traditional axiomatizations, the goal is not to find the most parsimonious system. Instead the individual axioms have been carefully selected so that each expresses an intuitively natural idea about inductive inference that can be used independently. As a result, the ideas are logically stronger than they need be were the only purpose to deduce the probability calculus. These ideas, as developed in the Sections 3-6, are: *Framework*, *Addition*, *Bayes property (=Narrowness + Multiplication)* and *Real values*.

Excepting *Framework,* these ideas are independent of one another. The *Bayes property*, for example, is responsible for the dynamics of conditionalization under Bayes' theorem; it is independent of *Addition* and *Real values*, and may be invoked independently of them. The proposal of this paper is that we should do just this. We should not assume that all these component notions apply in every context in which we may seek to use the probability calculus as a logic of induction. Rather we should determine which, if any, apply in the context at hand and use those only. I will suggest that following this course will help us avoid problems associated with the application of the probability calculus to inductive inference.

How are we to decide which components apply in a given context? A principled basis is supplied by what I call elsewhere a 'material theory of induction' (Norton [2003], [2005]). According to it, induction differs fundamentally from deduction in that inductive inferences are not licensed ultimately by universally applicable inference schemas into which particular content may be inserted. Rather they are licensed by contingent facts. Since different facts obtain in different domains, we should expect different inductive inference forms to be applicable in different domains. If we are reasoning about stochastic systems governed by a theory with

physical chances, the facts of that theory will likely license inductive inference forms involving the probability calculus. In domains in which different facts prevail, these forms may no longer be licensed. Section 8 provides illustrations and argues that neither *Addition* nor *Bayes property* is licensed for inductive inferences concerning some indeterministic systems not governed by physical chances.[2]

2. Failure of Demonstrations of Universality.

There have been numerous attempts to establish that the probability calculus is the universally applicable logic of induction. The best known are the Dutch book arguments, developed most effectively by de Finetti ([1937]), or those that recover probabilistic beliefs from natural presumptions about our preferences (Savage [1972]).[3] Others proceed from natural supposition over how relations of inductive support must be, such as Jaynes ([2003], Ch. 2).

2.1 Working Backwards

These demonstrations are ingenious and generally quite successful, in the sense that accepting their premises leads inexorably to the conclusion that probability theory governs inductive inference. That, of course, is just the problem. The conclusion is established only in so far as we accept the premises. Since the conclusion makes a strong, contingent claim about our world, the demonstrations can only succeed if their premises are at least strong factually.[4] That makes them

---

[2] The idea that one should investigate induction locally has been considered in the literature that gives a probabilistic analysis of induction, but without forgoing the idea that the probability calculus underwrites inductive inference even locally. For an entry to this literature, see Kyburg ([1976]).

[3] Strictly speaking, these arguments purport to establish only that degrees of belief, as made manifest by a person's preferences and behaviors, must conform to the probability calculus on pain of inconsistency. They become arguments for universality if we add some version of a view common in subjectivist interpretations that degrees of belief are only meaningful in so far as they can be manifested in preferences and behaviors.

[4] There is no escape in declaring that good inductive inferences are, *by definition*, those governed by the probability calculus. For any such definition must conform with essentially the same facts

at least as fragile as the conclusion they seek to establish. Since they are usually created by the simple expedient of working backwards from the conclusion, they are often accepted just because we tacitly already believe the conclusion.

For these reasons, all demonstrations of universality are fragile and defeated by a denial of one or more of the premises. A few examples illustrate this general strategy for defeating the demonstrations. Dutch books arguments are defeated simply by denying that some beliefs are manifested in dispositions to accept wagers. Or their results can be altered merely by adjusting the premises we will accept. Dutch book arguments commonly assume that there are wagers for which we are willing to accept either side. That assumption is responsible for the additivity of the degrees of belief the argument delivers. Its denial involves no incoherence, in the ordinary sense. It just leads us to a calculus that is not additive. (See Smith [1961].) Similarly, there is no logical inconsistency in harboring intransitive preferences. They will, however, not sustain a recovery of transitivity of beliefs in Savages' ([1972], §3.2) framework, which is necessary for beliefs to be probabilistic.[5] Finally, Jaynes ([2003], §2.1) proceeds from the assumption that the plausibility of A and B conditioned on C (written '(AB|C)') must be a function of (B|C) and (A|BC) alone, from which he recovers the familiar product rule for probabilities, $P(AB|C)=P(A|BC)P(B|C)$. That this sort of functional relation must exist among plausibilities, let alone this specific one, is likely to be uncontroversial only for someone who already believes that plausibilities are probabilities and has tacitly in mind that we must eventually recover the product rule.[6]

---

in that it must cohere with canonical inductive practice. Otherwise we would be free to stipulate any system we choose as the correct logic of inductive inference.

[5] Savage's framework harbors a circularity. In its barest form, it offers you a prize of $1, say, for each of the three acts f_A, f_B, f_C, if uncertain outcomes A, B or C happen, respectively. You prefer f_A to f_B just in case you think A more likely than B. So your preferences on f_A, f_B and f_C will be transitive just in case you already have transitive beliefs on the possibilities of A, B and C.

[6] A simple illustration of an assignment of plausibilities that violates the functional dependence is 'Plaus.' It is generated by a probability measure P over propositions A, B, ... as a coarsening,

The fragility of these demonstrations is very similar to the failure of attempts to show that Euclid's fifth postulate of the parallels is the only postulate admissible in geometry. These attempts started by denying Euclid's fifth postulate in the context of the other postulates and inferring from the denial some unusual geometric propositions that, we were to suppose, are incoherent. It was eventually realized in the nineteenth century that the denial of Euclid's fifth postulate involved no inconsistency; it merely led us to different geometries.

While I believe all these demonstrations fail in establishing universality, they still have great value. For we learn from them that, in domains in which their premises hold, our inductive inferences must be governed by the probability calculus.

2.2 The Surface Logic.

There is a second sort of argument for universality, mostly suggested indirectly by impressive catalogs of the success of Bayesian analysis at capturing our intuitions about inductive inference. All these intuitions so far have been captured by the probability calculus; so, the thought goes, we should expect this success to continue.

In my view, the success is overrated and does not sustain the probability calculus as the unique logic of induction. In many cases, the success is achieved only by presuming enough extra hidden structures—priors, likelihoods, new variables, new spaces—until the desired intuition emerges. That does not mean that the logic on the surface is probabilistic, but only that this surface logic can be simulated with a more complicated, hidden structure that employs probability measures.

Two examples will illustrate the concern. Take Hempel's original question of whether a non-black, non-raven confirms that all ravens are black. A probabilistic analysis gives an intuitively very comfortable result. But it only succeeds by adding a great deal of new structure to the original problem: populations with different distributions of ravens and black objects and a presumption that we are sampling randomly from them. That changes the problem to a new one amenable to probabilistic analysis. (For a survey, see Earman ([1992]), §3.3.) Consider ignorance, which, I argue below in Section 4.2, is not represented in an additive calculus. It may

---

with only two intermediate values: Plaus(A|B) = 'Low' when $0 < P(A|B) < 1/2$; and Plaus(A|B) = 'High' when $1/2 \leq P(A|B) < 1$

be introduced by associating beliefs with convex sets of probability measures. While additive measures were used to produce them, the sets themselves no longer conform to a logic with the formal property of *Addition* as defined below. Additive measures are merely the device used to generate a new system governed by a different surface logic.

Once again there is a geometric analogy. We can recover many non-Euclidean geometries by considering curved surfaces embedded in a higher dimensioned Euclidean space. That does not mean that Euclidean geometry is the universal geometry. It is not the geometry intrinsic to the surface. However we learn that Euclidean geometry can be used as a tool to generate that geometry, as could other geometries.

3. Framework

The system of properties for confirmation relations to be described here draws on the extensive literature in axioms for the probability calculus already developed. See especially Cox ([1961]) and, for surveys, see Fine ([1973]) and Fishburn ([1973]).

3.1 The Properties

The framework assumes a set of propositions A_1, A_2, … closed under the familiar Boolean operations ~ (negation), ∨ (disjunction) and & (conjunction). Where the context calls for it, the set will be assumed to be closed under countable disjunction. The universal proposition is $\Omega = A_1 \vee A_2 \vee$ …. Implication ⇒ is stronger than material implication; $A \Rightarrow B$ means that that propositions are so related[7] that ~A∨B must always be true; that is, ~A∨B = Ω. The universal proposition, Ω, is implied by every proposition in the algebra and is always true. The proposition, ∅, implies every proposition and is always false.

The symbol [A|B] represents the degree to which proposition B confirms proposition A and is undefined when B=∅. The relation on these degrees

$$[A|B] \leq [C|D]$$

(or equivalently [C|D] ≥ [A|B]) is interpreted informally as 'D confirms C at least as strongly as B confirms A.' It satisfies:

---

[7] For example, if we associate propositions with the sets of worlds in which they are true, then A ⇒ B obtains just if A's worlds are a subset of B's.

F. *Framework*

F1. *Partial order*. The relation ≤ is a partial order. That is, for any admissible[8] propositions

    A, B, C, D, E and F:

F1a. *Reflexivity*. [A|B] ≤ [A|B]

F1b. *Antisymmetry*. If [A|B] ≤ [C|D] and [A|B] ≥ [C|D] then [A|B] = [C|D]

F1c. *Transitivity*. If [A|B] ≤ [C|D] and [C|D] ≤ [E|F] then [A|B] ≤ [E|F]

Antisymmetry allows us to define < and > in the usual way.[9] We also suppose:

F2. For all admissible propositions A and B:

F2a. $[\varnothing|\Omega] \leq [A|B] \leq [\Omega|\Omega]$

F2b. $[\varnothing|\Omega] < [\Omega|\Omega]$

F2c. $[A|A] = [\Omega|\Omega]$ and $[\varnothing|A] = [\varnothing|\Omega]$;

and

F3. *Universal comparability*. For all admissible propositions A, B, C and D

    [A|B] ≤ [C|D] or [A|B] ≥ [C|D];

and

F4. *Monotonicity*. For all admissible propositions A, B and C,

    if A⇒B⇒C, then [A|C]≤[B|C].

3.2 Boundaries

While these properties are natural, they nonetheless have significant content and it is far from clear that they will be applicable to all cases of inductive inference. Two properties are especially vulnerable, *F3. Universal comparability* and *F1c. Transitivity*, as is possibly *F4. Monotonicity.*

3.2.1 Universal Comparability

We cannot presume, as Keynes ([1921], Ch.3) correctly urged, that all degrees of confirmation are comparable. A tacit expectation of universal comparability is natural as long as we think of degrees of confirmation as real valued. The expectation rapidly evaporates once we use more complicated structures. Imagine for example that the degrees are real intervals in [0,1] with the

---

[8] Here and henceforce, 'admissible' precludes formation of the undefined [·|∅].

[9] [A|B] < [C|D] and [C|D] > [A|B] just in case [A|B] ≤ [C|D] but not [A|B] = [C|D].

size of the interval betokening something about the bearing of evidence. Take two intervals [0.01, 0.99] and [0.49, 0.51]. If they must be comparable, the only relation that respects the symmetry of dispositions about the midpoint 0.5 is that they are equal. But that contradicts the presumption that the size of the interval represents some sort of difference in the degrees of confirmation.

However, even if degrees of confirmation are real valued, it does not follow that they are comparable. For two degrees to be comparable in the relevant sense, they must measure essentially the same thing. The mere fact that two scales employ real values is not enough to assure this. One hundred degrees Celsius on the mercury thermometer scale and on the ideal gas thermometer scale are equivalent since they measure the same thing, temperature. They are none of equivalent to, less than or greater than one hundred degrees Baumé of specific gravity.

Propositions can bear evidentially on one another in many ways and the range of variation is sufficiently great that we can surely not always presume comparability of the degrees, even if both are measured on the same numerical scale. Consider the hypothesis H that the half-life of radioactive decay of Radium 221 is 30 seconds and the evidence E that some Radium 221 atom did decay in a time period of 30 seconds. The two degrees, [E|H] and [H|E], are very different. In the first we take certain laws of physics, with their characteristic constants, as fixed and distribute belief over possibilities (decay in 30 seconds, decay in 40 second, etc.). Those laws provide physical chances for the possibilities and the bearing of H on E is detailed for us completely as a matter of physical law.[10] In the second, we take an experimental fact as fixed and must now distribute belief over the possibility of different half lives for Radium 221. No physical law can fix the bearing of E on H, for now the range of possibilities must involve denial of physical laws; there is only one correct value for the half life. Even exactly how we are to conceive that range is unclear. Will we try to hold all of physics fixed and just imagine different half-lives for Radium 221? Or should we recall that the physical properties of Radium 221 are fixed by quantum physics and chemistry, so that differences in half-lives must be reflected in difference throughout those theories. And how should those differences be effected? As alterations just to fundamental constants like h and c? Or in alterations to Schrödinger's

[10] Or, more cautiously, Lewis' ([1980]) 'principal principle' in effect enjoins us to endow our degrees of confirmation with the properties of a physical chance.

equation itself? My point is not that we cannot answer these questions, but that answering them engages us in a very different project that is a mixture of science and speculative metaphysics. The way H bears on E in [E|H] is very different from the way E bears on H in [H|E].[11]

So, if we expect the degrees of confirmation simply to measure the bearing of evidence, as an objectivist about probability like Keynes would, then we should not expect the two sets of degrees always to be comparable. A subjectivist about probabilities has no easy escape. Of course the subjectivist simply supposes comparability and stipulates real valued prior probabilities that lead to real values for both [E|H] and [H|E] upon conditionalization. The hope is that the subjectivist's assignments will eventually betoken something more than arbitrary numbers as the accumulation of evidence 'washes out the priors' and leads to a convergence of values for all subjectivists. If the very idea that the two degrees are comparable entered originally as a supposition without proper grounding, the convergence does not remove its arbitrariness. Oranges are not apples, even if we end up agreeing on how many apples make an orange.

3.2.2 Transitivity

The prevalence of real values for degrees of confirmation can also mislead us into expecting their transitivity universally. That expectation fades once we entertain the possibility that these degrees have more complicated structures.[12] For example, that some hypothesis H entails true evidence E is generally taken to confirm H. Some hypotheses, however, are routinely assessed as being more deserving of support if they manifest certain virtues in the context of the successful deduction. These virtues include: simplicity, scope, fecundity and explanatory power, with the latter engendering the account of induction known as 'inference to the best explanation.' So three

---

[11] Humphreys ([1985]) uses related illustrations to object to the propensity interpretation of probability. For example, if proposition S asserts that a person is a smoker and C that the person has an undiscovered lung cancer, then the causal propensity of a smoker to have an undiscovered lung cancer is expressed by the direct probability P(C|S). Yet precisely because this causal propensity is uni-directional, the inverse probability P(S|C) does not express a causal propensity of people with undiscovered lung cancers to smoke.

[12] The discussion of Section 4 below raises the possibility of degrees of confirmation with a two-dimensional structure, where lower degrees representing some mix of disbelief and ignorance.

hypotheses H_1, H_2 and H_3 may score differently with regard to three virtues V_1, V_2 and V_3. Allowing for three values, 'high,' 'medium' and 'low,' we may end up with the following assignments:[13]

	V_1	V_2	V_3
$[H_1|E]$	high	medium	low
$[H_2|E]$	medium	low	high
$[H_3|E]$	low	high	medium

Table 1. Intransitive degrees

Following a simple rule that the majority wins, $[H_1|E] > [H_2|E]$, since $[H_1|E]$ outscores $[H_2|E]$ in two of three virtues. Similarly $[H_2|E] > [H_3|E]$ and $[H_3|E] > [H_1|E]$, which violates transitivity. Indeed, if we assign equal importance to the three virtues and require a rule of comparison to rank solely on the basis of the values in the table, then any rule that yields $[H_1|E] > [H_2|E]$ must also generate the intransitivity. For there is a cyclic symmetry in the values in that $[H_1|E]$ relates to $[H_2|E]$ in the same way as $[H_2|E]$ relates to $[H_3|E]$ and $[H_3|E]$ relates to $[H_1|E]$.

3.2.3 Monotonicity

Monotonicity prohibits evidence from confirming a proposition more strongly than its deductive consequences. Yet, as Tversky and Kahneman ([1982]) showed in psychological experiments, people are easily led to violate this prohibition. If she is described appropriately, subjects will judge it more probable that Linda is a bank teller and a feminist than that Linda is a bank teller. Tversky and Kahneman interpret this to mean that people conflate probability with representativeness. Might there be a calculus of confirmation that violates monotonicity in that degrees of confirmation measure, in part, goodness of fit, in which Linda the bank teller and feminist would be a better fit to the evidence than Linda the bank teller? That could arise in a system of inductive inference with a rule of detachment that forces us to select among well-confirmed hypotheses, using quantities [H|E] as scores. On evidence E = 'the coin did not fall

---

[13] These virtues are discussed further in Section 5.3.3 below.

heads,' it may score H = 'the coin fell tails' higher than H' = 'the coin fell tails or on edge.' For if we must choose just one hypothesis to detach from E, it would in ordinary circumstances be H and not H', even though H entails H'.

4. Addition

4.1 The Property: Disbelief versus Ignorance

The range of degrees of confirmation for some proposition A spans from the maximal $[A|\Omega]=[\Omega|\Omega]$ to the minimal $[A|\Omega]=[\varnothing|\Omega]$. Do these extreme values correspond to justification of complete belief in A and complete disbelief in A? Or do they correspond to complete belief in A and complete ignorance concerning A? The signal feature of a probability measure is that it is an *additive* measure and we shall see that this property is derived from choosing the first option:

> *Underlying intuition of Addition*: The range of degrees of confirmation span justification of complete belief and complete disbelief.

This first option is characterized by a reciprocal relationship between degrees of confirmation for A and for its negation, ~A. Complete disbelief in A corresponds to complete belief in ~A. As the degree of confirmation $[A|\Omega]$ weakens from the maximum $[\Omega|\Omega]$ that justifies complete belief, then the degree of confirmation $[\sim A|\Omega]$ must strengthen accordingly from the minimal $[\varnothing|\Omega]$ that justifies complete disbelief. We should expect, under the above intuition, that this reciprocal relation between degrees of confirmation will also hold when we divide any proposition B into two, exhaustive and mutually exclusive logical parts, A&B and ~A&B; and that it will obtain when we conditionalize on any background C. The map that takes us from [A&B|C] to [~A&B|C] will, in general, differ according to [B|C], since the maximum degree that can be assigned to [A&B|C] or [~A&B|C] is set by [B|C] under *F4. Monotonicity*. So there is a family of functions, $f_{[B|C]}(\cdot)$. We express the above intuition by requiring:

> *A'. Addition*. For any propositions A and B and any admissible C, there exists a function
>
> $$[\sim A\&B|C] = f_{[B|C]}([A\&B|C])$$
>
> where f is strictly increasing in [B|C][14] and strictly decreasing in [A&B|C].[15]

---

[14] That is, for each y, if x'>x, then z'>z, where $z'=f_{x'}(y)$ and $z=f_x(y)$.

To convert this form of *Addition* into a more familiar one, we note that, since f is strictly increasing in [B|C], the function f can be inverted in this argument.[16] That is, there exists a function g, such that [B|C] = g([A&B|C], [~A&B|C]) where g is strictly increasing in both [A&B|C] and [~A&B|C]. This last function is presented in a more familiar way as an addition operator in a property equivalent to *A' Addition*:

A. Addition. For any admissible proposition Z and mutually contradictory propositions X and Y, there exists an addition operator ⊕ such that

$$[X \vee Y | Z] = [X|Z] \oplus [Y|Z]$$

where ⊕ is strictly increasing in both [X|Z] and [Y|Z].

This second form justifies the name *Addition*, since it displays the sense in which the degree of confirmation of a proposition is fixed by the 'adding up' of degrees of confirmation of its logical parts.

Properties that ⊕ must carry for compatibility with the Framework F. are readily deduced from the logical properties of propositions, such as X∨Y= Y∨X, U∨V∨W = (U∨V)∨W = U∨(V∨W), X∨∅=X and X∨~X=Ω:

$$[X|Z] \oplus [Y|Z] = [Y|Z] \oplus [X|Z]$$

$$[U|Z] \oplus [V|Z] \oplus [W|Z] = ([U|Z] \oplus [V|Z]) \oplus [W|Z] = [U|Z] \oplus ([V|Z] \oplus [W|Z])$$

$$[X|Z] \oplus [\varnothing|Z] = [X|Z]$$

$$[X|\Omega] \oplus [~X|\Omega] = [\Omega|\Omega]$$

4.2 Boundaries

The obvious limitation of any calculus of inductive inference that employs *A. Addition* is that will be unable to incorporate directly degrees of confirmation that support ignorance, as opposed to disbelief. To see the difference, imagine that we have an atom of Radium-221 with a half life of 30 seconds. On that background evidence, we assign familiar probabilities to the outcomes of radioactive decay in the next 30 seconds (D) or no decay in the next 30 seconds (~D):

---

[15] That is, for each x, if y'>y, then z'<z, where z'=f_x (y') and z=f_x(y).

[16] If it were not invertible, there would be unequal values x and x' such that $f_{x'}$(y) =f_x(y), which would violate the strict increase of f in x.

$$P(D)=0.5 \quad P(\sim D)=0.5 \quad P(D\vee\sim D)=1$$

Now consider an atom of the first radioactive element to be mentioned in the first text book of nuclear chemistry that will be published in the year 2100. One might reasonably protest that the evidence given supports no real belief towards either of the corresponding propositions D', that an atom of this element decays in a 30 second time period, or ~D'. Assigning a probability of 0.5 to each seems excessive. We certainly do not believe the consequence often associated with probabilities that, in situations like this, decay will happen roughly one in two times. But then if we assign a probability of less than 0.5 to one of D' or ~D' to flag our uncertainty, because of the additivity of probabilities, we must assign a probability greater than 0.5 to the other over which we are equally uncertain.

The Shafer-Dempster theory of belief functions (Shafer [1976], pp. 23-24) was devised to accommodate just such a situation. In it, we may represent ignorance by assigning beliefs as:

$$Bel(D)=0 \quad Bel(\sim D)=0 \quad Bel(D\vee\sim D)=1$$

where the last assignment reflects our certainty in the logical truth D∨~D. Or, if on the evidence of present day text books, we incline slightly to believing that the text books of 2100 will favor discussion of very short lived elements, we may shift our belief just a little toward D:

$$Bel(D)=0.1 \quad Bel(\sim D)=0 \quad Bel(D\vee\sim D)=1$$

These belief functions Bel are non-additive. They violate the property A. *Addition*; Bel(D∨~D) is not a strictly increasing function of Bel(D) and Bel(~D).

While this example does not show precisely what formal properties are to be associated with ignorance,[17] it does display how deviations from *A. Addition* do allow some sort of representation of ignorance.

It is common to form convex sets of probability distributions as a way of representing ignorance in probabilistic analysis. If P_x is the distribution that assigns $P_x(D) = x$ and $P_x(\sim D)=1-x$, then we may represent complete ignorance over D as the convex set spanning the two extreme cases of P_0 and P_1; that is, the set $\{P_x: 0\leq x\leq 1\}$. This proposal has the very real advantage of allowing us to deal with ignorance in a systematic way. However it is not helpful for the present project of understanding formally the various components ideas that have led to

---

[17] That project is reserved for Norton (unpublished).

the success of probability theory as a calculus of induction. First, the representation is not literally correct. That is, ignorance is not the maintaining of all possible beliefs at once; it is the maintaining of none of them. So we should regard the device of convex sets as a way of simulating ignorance through a convenient fiction. And it is an arbitrary one, since it corresponds to a uniform distribution over all beliefs in that each member in the set enters equally. We could certainly define a non-uniform distribution and have another way of approaching ignorance. The real difficulty in the present context is that use of these convex sets diverts us from the question of what formal property should replace *A. Addition* if our calculus is to allow representations of ignorance directly. Indeed, as noted in Section 2.2 above, since the device employs additive measures to simulate a new surface logic, we may even end up overlooking that *A. Addition,* or some analog of it, must be violated in this surface logic.

5. Bayes Property

5.1 The Property

The characteristic of the Bayesian approach to induction is that the import of new evidence is incorporated into the probability distributions by conditionalization and that the dynamics of this incorporation is governed by Bayes' theorem. In developing what is called the 'Bayes property' here, we shall see here that these dynamics can be inferred from a simple model of how hypotheses are confirmed by their true deductive consequences.

> *Underlying intuition of Bayes' property.* An hypothesis accrues inductive support from evidence just if it has a disjunctive part that entails the evidence.
> (narrowness) The presence of other disjunctive parts logically incompatible with the evidence does not affect the level of support.
> ('refute and rescale') Evidence bears on hypotheses that entail it by refuting those logically incompatible with it and uniformly redistributing support over those that remain; this uniform redistribution is carried out everywhere in the same way and preserves the relative ranking of hypotheses that entail the evidence.[18]

---

[18] That Bayesian inference depends on such a simple model is well recognized. See, for example, Hawthorne ([1993]).

We will re-express this intuition in more formal terms as two properties. To arrive at the first, note that an hypothesis H that is logically compatible with evidence E can be divided into two disjunctive parts, H&E and H&~E. The first entails the evidence E. So we have:

N. Narrowness. For any proposition A and any admissible B,

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

To develop the intuition of 'refute and rescale' dynamics, consider three propositions A, B and C, where

$$A \Rightarrow B \Rightarrow C$$

Proposition A begins with support $[A|C]$, where this degree may vary from a minimum of $[\varnothing|C]$ to a maximum of $[B|C]$, set by conformity to *F4. Monotonicity.* After conditionalizing on B, it has support $[A|B\&C] = [A|B]$, where this degree may now vary from a minimum of $[\varnothing|B]$ to a maximum of $[B|B] = [\Omega|\Omega]$. So the effect of conditionalizing on B is represented by a map $f_{[B|C]}$ that rescales the support accorded to A from the old to the new range:

$$[A|B] = f_{[B|C]}([A|C])$$

where neither B nor C may be $\varnothing$. (See Figure 1.) The subscript on f is needed since the function must map the extremal values as $[\varnothing|B] = f([\varnothing|C])$ and $[B|B] = f(B|C))$, so that a different map is needed for each distinct value of $[B|C]$. That does not fix the action of $f_{[B|C]}$ on intermediate values. In a more general context, one might posit different functions $f_{[B|C],C}$ that are specific to the environment of each proposition C. That would amount to supposing that the rescaling differs according to the content of the proposition C. The requirement above that the redistribution of support 'is carried out everywhere in the same way' is intended to preclude this. That is, there is a unique family of rescaling maps $f_{[B|C]}$ for the whole set of propositions, sensitive only to the degrees $[B|C]$ and $[A|C]$ and not to anything further in the content of the propositions A, B and C.[19]

---

[19] That is, if we have propositions $A \Rightarrow B \Rightarrow C$ and $A' \Rightarrow B' \Rightarrow C'$, where, for admissible B, B', C and C', $[A|C] = [A'|C']$ and $[B|C] = [B'|C']$, then $[A|B] = f_{[B|C],C}([A|C]) = [A'|B'] = f_{[B'|C'],C'}([A'|C'])$.
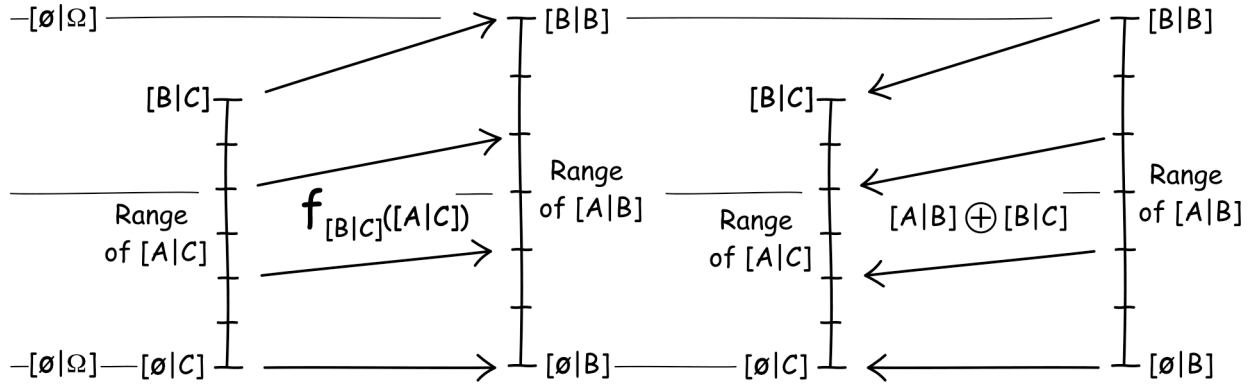
Figure 1. Conditionalization as rescaling degrees of confirmation

The maps $f_{[B|C]}$ must also 'preserve the relative ranking of hypotheses that entail the evidence.' So if [A'|C] > [A|C], then [A'|B] > [A|B]. It follows that $f_{[B|C]}$ ([A|C]) is strictly increasing in [A|C]. Therefore $f_{[B|C]}$ ([A|C]) is invertible in [A|C]. The inverse of this function,

$$[A|C] = f_{[B|C]}^{-1}([A|B])$$

can be written in a more familiar way as a product operator

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

which must be strictly increasing in [A|B] since $f_{[B|C]}$ ([A|C]) is strictly increasing in [A|C].

That the operator should also be strictly increasing in [B|C] for all values of [A|B] excepting [∅|B] is the import of the requirement above that the redistribution be 'uniform.' An increase in [B|C], when [A|B] has the maximal value [B|B], is reflected by an exactly equal increase in [A|C], since [B|C] = [B|B] ⊗ [B|C]. An increase in [B|C], when [A|B] has the minimal value [∅|B], is reflected by no change in [A|C], since then [∅|C] = [∅|B] ⊗ [B|C]. The requirement of uniformity amounts to asking that the increase in [A|C] for intermediate values of [A|B] should be uniformly interpolated between these two extreme values. Or it would amount to this if there were a way to represent 'uniformly interpolated' with the structures defined so far. But there is not. However, whatever it may amount to, minimally, it must require some increase in [A|C] for all intermediate values of [A|B]. That is sufficient to support the strict increase of ⊗ in [B|C] unless [A|B] is [∅|B].

Collecting these properties, we have:

M. Multiplication. For any proposition A and admissible propositions B and C such that A ⇒ B ⇒ C, there exists a multiplication operator ⊗ such that

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

where $\otimes$ is strictly increasing and thus invertible in both arguments (excepting [B|C], when [A|B]=[∅|B]).

This operator is the analog of the normal product operator of the probability calculus, where, for these A, B and C, $P(A|C) = P(A|B) \cdot P(B|C)$.

The two properties combined form:

B. Bayes Property.

N. Narrowness and *M. Multiplication*

We can readily deduce the expected rules from this combined property. The analog of the product rule of probability theory is

$$[A\&B|C] = [A\&B|B] \otimes [B|C] = [A|B] \otimes [B|C] \tag{1}$$

Combined with *A. Addition* we have the analog of the rule of total probability

$$[A|C] = [A\&B|C] \oplus [A\&{\sim}B|C] = ([A|B] \otimes [B|C]) \oplus ([A|{\sim}B] \otimes [{\sim}B|C]) \tag{2}$$

5.2 Bayes' Theorem

The analog of Bayes' theorem is derived in the usual way from the product rule. For an hypothesis H and evidence E:

$$[H\&E|\Omega] = [H|E] \otimes [E|\Omega] = [E|H] \otimes [H|\Omega] \tag{3}$$

The terms can be labeled in the obvious way in analogy with the usual, probabilistic form of Bayes' theorem as: 'posterior' ([H|E]), 'expectedness' ([E|Ω]), 'likelihood' ([E|H]) and 'prior' ([H|Ω]). Since the operator $\otimes$ is strictly increasing and invertible in both arguments (excepting one case), the posterior [H|E] can be recovered by inverting $\otimes$ and the theorem can be used in the usual way to recover familiar intuitions. Other terms equal, the posterior [H|E] will have a maximum value when H⇒E, for then the likelihood [E|H] = [Ω|Ω], which is the maximum value.[20] Similarly, other factors equal, an increase in the prior [H|Ω] will lead to a corresponding increase in the posterior [H|E]. And an hypothesis that successfully entails evidence of lower expectedness [E|Ω] will have a higher posterior. This much, and many more familiar results like

---

[20] The likelihood [E|H] = [E&H|H] by N. and, since H=E&H when H⇒E, we have [E|H] = [H|H], which is the maximal [Ω|Ω] by F2b.

18

them, are recoverable without assuming *A. Addition*. If it is assumed then a further form of Bayes' theorem can be recovered by substituting for the expectedness using the rule (2):

$$[E|\Omega] = ([E|H] \otimes [H|\Omega]) \oplus ([E|{\sim}H] \otimes [{\sim}H|\Omega]).$$

5.3 Boundaries

While we may find the simplicity of the 'refute and rescale' dynamics appealing, that simplicity proves to be its fundamental limitation. The dynamics are sensitive only to entailment relations. As we shall see below, that forces the inductive character of the inferences to be inserted by our selection of priors. That burden overtaxes the priors since they will also be called upon to represent initial states of ignorance at the same time as they must supply essential inductive content. And worse, that inductive content is decided in significant measure as a matter of stipulation. For these reasons, prior probabilities have inevitably become the traditional locus of problems in probabilistic analysis; they are called upon to make up for the deficiencies of the 'refute and rescale' dynamics.

5.3.1 Dogmatism of the Priors

It is well known in probabilistic analysis that once zero or unit probability has been assigned to an hypothesis' prior probability, conditionalization on new evidence compatible with it cannot alter those probabilities. The same problem arises in a system with *B. Bayes property*. Learning from experience will never lead it inductively to alter judgments of maximum or minimum belief, unlike humans.

For any hypothesis H and evidence E, we have from Bayes' theorem (3) the paired relations

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

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

where the second relation arises from setting H=Ω and noting that $[\Omega|E] = [E|E] = [\Omega|\Omega]$ from N. and F2c. Even if H is not Ω, once we set the prior $[H|\Omega]$ to $[\Omega|\Omega]$, compatibility of the paired relations forces the posterior $[H|E] = [E|E] = [\Omega|\Omega]$. A prior set to certainty is immovable inductively.

For any hypothesis H and any admissible evidence E, from the product rule (1), we have the paired relations

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

$$[\varnothing|\Omega] = [\varnothing|E] \otimes [E|\Omega]$$

where the second relation arises from setting H=$\varnothing$. If H is not $\varnothing$, if we set the prior [H|Ω] = [$\varnothing|\Omega$], it follows from F4. that [H&E|Ω] = [$\varnothing|\Omega$]. Compatibility of the paired relations forces the posterior to [H|E] = [$\varnothing$|E] = [$\varnothing|\Omega$]. A prior set to maximum disbelief is immovable inductively.

We can see how this last example arises directly from the excessive simplicity of the 'refute and rescale' dynamics. Those dynamics are sensitive only to the fact that both H&E and $\varnothing$ are each able to entail the evidence E. So, if they are given the same priors, they must then have the same posteriors. Since $\varnothing$ must remain at the minimal level of confirmation on any evidence, H&E is condemned to the same fate. A more sophisticated dynamics would be able to recognize and exploit the difference between $\varnothing$ vacuously entailing E and H&E entailing E.[21]

5.3.2 Impossibility of Prior Ignorance

We have seen that *A. Addition* precludes lower degrees of confirmation from representing ignorance as opposed to disbelief. It also turns out that *B. Bayes property* precludes priors that truly represent ignorance and does so independently of *A. Addition*. To see this, note that the property entails that, for any propositions H and E, where [E|Ω] is not [$\varnothing|\Omega$]:

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

This relation is invertible in [H|E]. That is:

[H|E] is fixed by the priors [H&E|Ω] and [E|Ω],

(unless [E|Ω] is [$\varnothing|\Omega$]). What this means is that the degree [H|E]—whether it is high or low and in which precise measure—is already encoded in the prior [·|Ω]. The prior [·|Ω] amounts to a massive catalog of all possible relations of inductive support between all pairs of propositions. It must decide in advance just how we will redistribute support once we learn E, no matter what E may be (as long as [E|Ω] is not [$\varnothing|\Omega$]).

---

[21] Analogously, the fixity of maximum support arises since the dynamics does not distinguish the trivial entailment E $\Rightarrow$ E from the non-trivial H $\Rightarrow$ E, where H is strictly stronger, logically, than E (that is, for some X, E = HvX, where H&X is $\varnothing$).

There is a large literature devoted to 'ignorance priors,' 'uninformative priors' or 'informationless priors' in probability theory. (See, for example, Jaynes, 2003, Ch. 12.) It is generally recognized that these terms are misnomers; the priors are really only as uninformative as the probability calculus allows and are typically tailored to being that uninformative about one particular fact, such as a parameter value. Were they really to achieve ignorance in the sense of a complete null state, the result would be a catastrophe for any system whose dynamics conforms to *B. Bayes property*. For all the system can do is to take a prior already rich in inductive information and refine it by the dynamic of 'refute and rescale.'

5.3.3 Accommodation of Virtues

An important limitation of the 'refute and rescale' dynamics is that it cannot differentially reward two hypotheses for their success in entailing the same true evidence. If hypotheses H_1 and H_2 entail the evidence E and we conditionalize on E, the resulting changes in degrees of confirmation will be the same for each. For, in this case, *M. Multiplication* becomes

$$[H_i|E] \otimes [E|\Omega] = [H_i|\Omega]$$

If the two priors $[H_1|\Omega]$ and $[H_2|\Omega]$ agree, then so must the posteriors $[H_1|E]$ and $[H_2|E]$, because of the invertibility of the operator $\otimes$.

There is strong indication that this outcome renders systems with the *B. Bayes property* too insensitive to differences in the way hypotheses may entail evidence. The dogmatism of the priors above arose because the system is unable to distinguish the non-vacuous entailment of evidence E by some hypothesis from the vacuous entailment of E by the contradiction $\varnothing$. Some logics of induction, such as that illustrated in Section 8.3 below, must differentially reward hypotheses H_1 and H_2.

Moreover, standard lore does not automatically accord equal confirmatory boosts to the two hypotheses H_1 and H_2. One is often favored over the other because the first entails the evidence is some virtuous way: with great simplicity or explanatory power; or because the second does it with some deficiency: it is adhoc or grue-ified. Might there be some system of inductive inference that could distinguish some entailments as virtuous and others as deficient? The principle obstacle is that the virtues—notably simplicity and explanatory power—are so poorly understood that even the outlines of such a system are obscure.

The problem of accommodating these virtues and vices into a probabilistic analysis is not new. (For a helpful entries into this literature, see Howson and Urbach, 1996, Ch. 7.) While the problem has been addressed with many ingenious stratagems, they must all come down to one idea only. The only way a system that conforms to *B. Bayes property* can differentiate H_1 and H_2 is to reward virtue with a high prior and punish vice with a low prior.

The effect of this need is that any system conforming to *B. Bayes property* must urge that the standard lore is mistaken in distinguishing virtuous entailments. For example, the standard lore is that the success of an ad hoc hypothesis in entailing some remarkable evidence gives it no boost in confirmatory support, for the success is achieved unvirtuously by cooking the books.[22] Under 'refute and rescale' dynamics, this same conclusion must be arrived at in a two-step calculation that must itself be cooked to yield the null outcome. It says, contrary to the lore, that the ad hoc hypothesis does accrue exactly as large a boost in confirmatory support as enjoyed by the hypothesis that virtuously entails the same evidence. However the gains of that boost are exactly cancelled by a prior that has been cooked to just the very low value needed.

While this stratagem of explicating virtues and vices in terms of high and low priors has had some notable successes in the probability literature, it faces a fundamental limitation. The assigning of a prior is global; it is done once. Yet, in the lore, the import of virtues and vices is local and may differ as hypotheses are subject to evidential scrutiny in different contexts, which in turn may call for differing priors. For example, the wave theory of light gives an especially simple and elegant explanation of interference. Its account of the rule of stellar aberration, however, proves to be quite tortured, once one looks at it closely—so much so, that it was a major achievement of late 19th century electrodynamics to be able to show that the wave theory could accommodate the totality of the rule satisfactorily. (See Norton [forthcoming].) The situation reverses for a corpuscular theory of light. It gives a simple and elegant explanation of stellar aberration; but, in so far as Newton's corpuscular theory was able to give any account of the interference effect of 'Newton's rings' using his fits of easy reflection and refraction, it was

---

[22] Or at least this is clearly so for the 'bad' cases, such as the supposition of a creationist geology that the world was created in 4004 BC complete with the fossil record of all geological eras intact. See Howson and Urbach ([1996], pp. 154-57) for cases of 'good' ad hoc hypotheses that do deserve support.

certainly not virtuous.[23] The one prior must somehow reward virtue in one context and punish vice in another.

Or we may be in a situation in which we cannot adjust priors to reward a virtue. In 1905, Einstein used his light quantum hypothesis to produce remarkably simple explanations of some of the observed properties of radiation. We should like to reward the light quantum hypothesis for not just entailing the evidence, but for explaining it virtuously. Yet, in 1905, after the nineteenth century overthrow of corpuscular theory and the resounding success of the wave theory of light, any investigation of the properties of light must begin with a low prior on any corpuscular hypothesis.

Finally, there is a related problem arising directly from *N. Narrowness*. That property allows evidence E to support an hypothesis H only through support of a disjunctive part H_1 that entails E. The other disjunctive parts are H_2, H_3, …, where $H = H_1 \vee H_2 \vee H_3 \vee$ … and $(H_2 \vee H_3 \vee …) \& E = \varnothing$. They have no effect on the support accrued to H. The property N. denies that there can be a synergy between the disjunctive parts, such that we should assign a different boost to the entire hypothesis than to the part, or to different disjunctive hypotheses that share the same disjunctive part that entails the evidence. Yet such synergies seem to have a place in the lore of confirmation. Kepler's hypothesis H_{Kep} that Mars orbits the sun in a particular ellipse gains some support from the evidence E of Tycho's observations of Mars and the sun. *N. Narrowness* requires us to accord just the same support on evidence E to the disjunctive hypothesis, $H_{disj} = H_{Kep} \vee H_2 \vee H_3 \vee ... \vee H_n$, where H_2, H_3, ... , H_n are hypotheses asserting other trajectories. As long as the hypotheses disjoined in H_{disj} form an inchoate set, this seems

---

[23] To anticipate the rejoinder, I fully expect that this example and most others can be accommodated in a Bayesian system by adding in enough distinctions, variables, likelihoods and priors, just as Ptolemy's geocentric system was able to accommodate any celestial motion by adding in enough epicycles and equants. That did not mean, however, that he had the right theory.

reasonable enough. However, at the level of accuracy of Tycho's data,[24] H_{Kep} is also a disjunctive part of another hypothesis. If we restrict Newton's theory of gravitation to two masses, one the size of the sun and the other Mars, the resulting hypothesis H_{New} predicts a large number of possible orbits.[25] The hypothesis, H_{New}, is a disjunction of hypotheses asserting them. The set disjoined is far from inchoate; its members are uniquely picked out as the set of orbits that satisfy Newton's inverse square law of gravity for these masses. In effect the hypothesis H_{New} just asserts that the orbit of Mars conforms to Newton's law.

The natural intuition is that H_{New} somehow expresses a deeper truth than H_{disj}, which merely disjoined H_{Kep} with a haphazard collection of alternatives. So we might expect that the synergistic disjunction in H_{New} deserves more support on the evidence than the inchoate disjunction of H_{disj}. *N. Narrowness* prohibits us from rewarding H_{New} for this synergy among its parts; it requires that the evidence E must support H_{disj} and H_{New} equally. The disparity becomes more striking the larger we conceive to be the set of haphazardly chosen orbits disjoined in H_{disj}. The usual strategy, of course, is to attempt to reward H_{New} in advance by assigning much greater priors to the disjunctive sets of hypotheses delimited by simple differential equations, such as appear in Newton's theory. However no assignment of priors can serve this end. As long as *N. Narrowness* is preserved, H_{disj} and H_{New} must be accorded the same support on evidence E, whatever their priors.

Once we discard the idea that any calculus of inductive inference *must* conform to *B. Bayes property*, we can begin to reflect upon what a replacement rule may bring. It may reward synergies; it may differentially reward virtuous and unvirtuous entailment of evidence; it may

---

[24] To simplify the example, I adopt the fiction that Tycho's data picks out just one orbit from each disjunctive set and neglect the motion of the sun around the sun-Mars center of mass that is entailed by Newton's theory.

[25] It predicts many more than the countably many disjuncts presumed by the *F. Framework*. To circumvent this difficulty, define $H_{New-Kep}$ as hypothesizing all the orbits admitted by Newton's theory in this case, excluding H_{Kep}. Then H_{New} retains the disjunctive form $H_{New} = H_{Kep} \lor H_{New-Kep}$.

not be so dogmatic that assignments of complete certainty and disbelief are immovable; and it may be rich enough to admit true null states as priors.

6. Real Values

The properties developed so far are necessary properties if degrees of confirmation are to be probabilities. They are not sufficient. They do not preclude value sets that cannot be mapped one-one onto a closed interval of reals in way that preserves ranking. The traditional counter-example (Jeffrey, 1961, pp. 19-20) is a family of hypotheses $H_{x,y}$, with real valued parameters x and y, where $[H_{X,Y}|\Omega] > [H_{x,y}|\Omega]$ just in case X>x, or, if X=x, Y>y. While ingenious 'Archimedean Axioms' have been devised to bridge the gap, none seem as illuminating in terms of fundamental ideas about inductive inference as the direct statement of the gap itself:

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

The obvious limitation of a system with this property is that it cannot accommodate inference problems that require larger value sets, such as infinitely great or infinitesimally small degrees (or at least not without non-standard reals). We can readily contrive problems that require such extensions. For example, consider the problem of picking a real number in [0,1] 'at random.' That the number is in the interval [0,0.5] is finitely more probable than in the interval [0,0.4], which is infinitely more probable than in the discrete set {0, 0.1, 0.2}, which is finitely more probable than in the set {0, 0.1}.

7. Sufficiency and Independence

The properties *F. Framework*, *A. Addition*, *B. Bayes property* and *R. Real values* are necessary if degrees of belief are to be probabilities. That they are sufficient follows from theorems in Aczel ([1966], pp. 319-24). That is, they are sufficient in the sense that, for each connected region[26] of

---

[26] A connected region is a set of propositions such that for any two propositions V and W in the set, there exist other propositions C_1, C_2, … , C_n in the set such that all of V&C_1, C_1&C_2, …, W&C_n are in the set.

the set of propositions, there exists a rescaling of the real values assigned to the degrees by *R*. *Real values* such that the rescaled values obey the probability calculus.

The independence of *A. Addition*, *B. Bayes property* and *R. Real values* from one another is obvious. That independence is important here since it is urged that we should implement these properties selectively, according to the problem at hand. There is some further independence of *A. Addition* and *B. Bayes property* from *F. Framework*. The most interesting is their independence from *F1c. Transitivity*. For that shows that A. and B. may obtain not just when the degrees of belief are not reals, but also when they are not even partially ordered. The demonstration of the consistency of A. and B. with a non-transitive value set is achieved by displaying an example that has all three.[27]

8. Illustrations

It is urged here that we should not seek the one, true combination of properties that yields the universally true logic of induction. Rather, in accord with the material theory of induction (Norton [2003], [2005]), we should invoke just those properties in each domain warranted by the material facts prevailing in each domain. So each domain will prove to have its own characteristic logic of induction. Some illustrations follow.

---

[27] Values are pairs (r,θ) of reals, where $0 \leq r \leq 1$ and $0 \leq \theta < 360$. The quantity θ will behave like an angle variable whose value always remains in $[0,360)$, so two θ's are added or subtracted *modulo* 360 (written '$-_m$' and '$+_m$'). The ranking is defined by $(r',\theta') > (r,\theta)$ when $r' > r$ or, if $r' = r$ and neither are 0 or 1, $0 < \theta' -_m \theta < 180$. Also for any $0 < r < 1$, $(r',\theta') = (r,\theta)$ if $\theta' -_m \theta = 180$. This ranking is intransitive: $(0.5, 0) > (0.5, 240) > (0.5, 120) > (0.5, 0)$. The maximum and minimum values are $(1,\theta)$ and $(0,\theta)$, where, for these two cases, all $(1,\theta)$ are taken to be same and all $(0,\theta)$ are taken to be same, for all θ values. The addition operator $\oplus$ is implemented as $(r',\theta') \oplus (r,\theta) = (r'+r, \theta' +_m \theta)$. The multiplication operator $\otimes$ is $(r',\theta') \otimes (r,\theta) = (r'.r, \theta' +_m \theta)$, when neither r' nor r is 1; as $(r',\theta') \otimes (1,\theta) = (r', \theta')$; or as $(1,\theta') \otimes (r,\theta) = (r, \theta)$. The operator $\oplus$ is strictly increasing in both arguments, as is $\otimes$, excepting in the latter case when either argument is $(0,\theta)$.

8.1 All Properties Retained

If the circumstances are governed completely enough by stochastic, physical laws, we will have sufficient material facts to warrant all the properties that comprise the probability calculus. Imagine, for example, that we randomly sample an atom of naturally occurring Uranium and seek evidence for its half-life. The evidence is that it does not undergo radioactive decay over the period of a week. To what degree does that evidence confirm each of the three half-lives possible for this atom? The known distribution of isotopes in naturally occurring Uranium fixes the physical chances for our sampling each of them. They are, by atoms in natural uranium: U-234 is 0.0054%, U-235 is 0.72% and U-238 is 99.275%. These chances fix our prior probabilities that the atom is the corresponding isotope with the characteristic half-life. Those half-lives are: U-234, 244,500 years; U-235, 703,800,000 years and U-238, 4,468,000,000 years. These half-lives, in conjunction with the rule of radioactive decay, give the physical chances for each isotope persisting for a week without decay. These physical chances provide the likelihoods that figure in the obvious, fully probabilistic analysis.

The scenarios imagined in Dutch book arguments give us another case in which we would use the full probability calculus. If we are in a casino, gambling, such that all the conditions in those scenarios obtain, then we ought to reason on the outcomes of the various games by means of the probability calculus.

These two examples illustrate how the notorious problem of selecting the right interpretation of the probability calculus[28] is greatly ameliorated in a material theory of induction. The appropriate interpretation will vary from domain to domain; we are absolved from the impossible burden of finding the one, universally correct interpretation that fits every case. In the case of sampling Uranium, all the propositions over which we reason are related by physical chances governed by the probability calculus, so we are able to set our degrees of confirmation by those chances. An objective interpretation will fit these probabilistic degrees best. They represent something like the relative frequency of truth among many physical systems relevantly similar to the present one. In the case of the casino, however, the degrees have a very different

---

[28] For recent discussion, see Gillies ([2000]), Galavotti ([2005]) and Mellor ([2005]). Gillies ([2000], Ch 8, 9) describes and advocates a pluralism over interpretations of the probability calculus.

meaning. They are now internal accounting factors that, if employed in the appropriate way, have the pragmatic value of protecting us from a harm.

8.2 Bayes Property *Only Retained*

A slight adjustment of the Uranium sampling problem above produces a problem in which we will dispense with *A. Addition* and *R. Real Values*. Instead of sampling atoms, imagine that we are given N atoms of some radioactive element of unknown half-life that we wish to determine. Our evidence is that over time t, n of the N atoms decay. By presumption, we have no idea of the half-life of the element. That is, we have no idea of the size of the time constant τ in the rule of radioactive decay, which tells us that the physical chance of decay of one atom over time t is

$$c(t) = (1-\exp(-t/\tau)). \tag{4}$$

and τ relates to the half-life $t_{1/2}$ as $t_{1/2}=\tau \ln 2$. If we were to attempt a probabilistic analysis, our prior probability would be the uniform prior over all values of τ, from 0 to infinity. That is, the probability density

$$p(\tau) = \text{constant} > 0 \tag{5}$$

This is an 'improper prior,' since it cannot be normalized to unit total probability. Many statisticians have been tempted to use such priors, since they can yield useful results. Yet they have been tormented by them, since they violate the probability calculus and only sometimes yield normalizable posterior probabilities. (Rosenkrantz [1981], §4*.2)

We are inclined to employ an improper prior precisely because the material facts of the inference problem do not call for its additivity. So the material theory of induction allows us to dispense with additivity. Compare this with the case of sampling Uranium above. Our uncertainty over the half-lives resulted from a sampling process, governed by physical chances. So, in conforming our degrees of confirmation to the chances, these degree had to be additive. In the new problem, our uncertainty does not result from any physical process governed by chances. It is just plain ignorance. Therefore, as we saw in the discussion of *A. Addition* above, we should not require our prior beliefs to be additive probabilities and, therefore, need not be troubled by the impropriety of the prior (5). Since the principal properties of *A. Addition* and *B. Bayes property* are independent, all other aspects can remain essentially the same. We will still learn about the half-lives from observed decay times, governed by the physical chances of (4). So as

before, we will still expect the dynamics of confirmation to be governed by *B. Bayes property*, with the likelihoods provided by physical chances.

To conform with the countable set of propositions supposed in *F. Framework*, we will replace the continuous range of τ values with a countable set of intervals. The propositions τ_i assert that the time constant τ for the element lies in the small interval $i\Delta\tau$ to $(i+1)\Delta\tau$ for small $\Delta\tau$ and $i=0, 1, 2, \ldots$ Let E be the evidence than n of the N atoms decay in some fixed time t. Bayes' theorem (3) asserts

$$[\tau_i|E] \otimes [E|\Omega] = [E|\tau_i] \otimes [\tau_i|\Omega] \tag{6}$$

For the remaining analysis, we will also dispense with *R. Real values*, since it is not needed for the outcome. To express our ignorance over the value of τ, we will assume the prior $[\tau_i|\Omega]$ has some fixed value greater than $[\varnothing|\Omega]$, independent of τ_i. Similarly we will assume only that the expectedness has some value $[E|\Omega] > [\varnothing|\Omega]$.[29] The physical chances of n decays among N atoms over time t when the time constant τ is within in the interval τ_i are approximated arbitrarily well for small enough $\Delta\tau$ by:

$$c(E|\tau_i) = [N!/(n!\ (N-n)!)]\ (\exp(-t/\tau))^{N-n}\ (1-\exp(-t/\tau))^n\ \Delta\tau \tag{7}$$

The likelihood $[E|\tau_i]$ will be set by these physical chances in the sense that $[E|\tau_i]$ will be a strictly increasing function of $c(E|\tau_i)$. It is a familiar result for the binomial expression of (7) that, for t, n and N fixed, $c(E|\tau_i)$ has a maximum value when $\exp(-t/\tau_{max}) = (N-n)/N$. That is,

$$\tau_{max} = t/\ln(N/(N-n)) \tag{8}$$

So the likelihood $[E|\tau_i]$ has a maximum at τ_{max}. Finally, since the expression $[\tau_i|E] \otimes [E|\Omega]$ in Bayes' theorem (6) is invertible in the posterior $[\tau_i|E]$, it follows that the posterior $[\tau_i|E]$ has a maximum at τ_{max}.

That is, on the evidence E, the time constant with the highest degree of confirmation is τ_{max}. This is the result we would expect. For example, if N/2 atoms decay in time t, then we would expect t to be a good estimator of the half-life $t_{1/2} = \tau \ln 2$. In this case, (8) becomes $t = \tau_{max} \ln 2$. It is also evident from its derivation that (8) is a maximum likelihood estimator of τ. Finally, the analysis can be repeated, replacing τ with a function of τ. For example, we can re-

---

[29] We choose values other than $[\varnothing|\Omega]$ for these two quantities to preserve the invertibility of $\otimes$.

express the law of radioactive decay (4) using $\lambda = 1/\tau$ and adopt a prior indifferent to values of λ. We arrive at the same estimator τ_{max} of (8).

8.3 Induction without Additivity *and* Bayes Property

The last example gave a principled reason for dispensing with the additivity of the prior. Otherwise the analysis was not so different from the familiar probabilistic one. That is so, since we were inferring inductively about propositions governed by physical chances. We recover an analysis that is very different from the familiar ones if we consider systems whose uncertainties are not governed by physical chances.

Many of our physical theories, including Newtonian physics, allow indeterministic systems. These are systems whose present states do not fix their future states and—the key fact of importance here—our physical theories provide *no physical chances* for the different futures. They tell us only that they are possible. (See, for example, Alper et al. [2000]; Norton [1999].) One of the simplest Newtonian examples is 'the dome,' described more fully in Norton ([2003a], §3). A point mass sits motionless at the apex of a dome with circular symmetry and is able to slide frictionlessly over it. See Figure 2. If the shape of the surface is chosen appropriately, Newton's equations admit solutions in which the mass remains at rest indefinitely at the apex, or for which the mass remains at rest for some arbitrary time T and then spontaneously accelerates in any radial direction.[30]

---

[30] An appropriate shape is $h = (2/3g)r^{3/2}$, where r is the radial distance in the surface of the dome and h the vertical distance below the apex; g is the acceleration due to gravity. For a unit mass on the surface, Newton's laws entail an outwardly directed acceleration field, $F = d^2r/dt^2 = r^{1/2}$. This equation is solved by r(t)=0, for all t; and by a spontaneous excitation at T: $r(t) = 0$, for $t \leq T$ and $r(t) = (1/144)(t-T)^4$, for $t \geq T$.
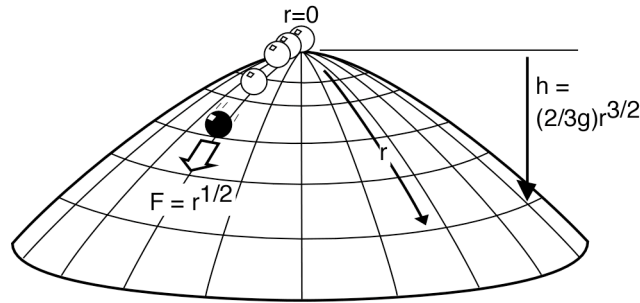
r=0

h =
(2/3g)$r^{3/2}$

r

F = $r^{1/2}$

Figure 2. The dome: an indeterministic system[31]

How are we to represent our uncertainty over the time T of spontaneous acceleration? The natural answer is to treat the problem as analogous to radioactive decay and assume that the timing of spontaneous acceleration is governed by the law of radioactive decay (4). The difficulty is that this law requires a time constant τ. A τ of a millisecond has a very different meaning physically for whether the excitation is likely to happen soon than does a τ of a millennium. Nothing in the physics supplies such a time constant. Indeed any proper probability distribution must employ some sort of parameter to govern the rate at which the integrated distribution approaches unity for large times. Yet no parameters are provided by the physics. Or one may try an improper distribution that merely sets the probability of spontaneous acceleration proportional to the size of the time interval in question. That still goes beyond the physics, since it asserts that spontaneous motion is twice as probable in a time interval that is twice as large. The physics knows nothing of this. It merely asserts that spontaneous motion in both time intervals is possible. There is no notion of 'twice as possible.'

The attempt to represent our uncertainty with a probability measure imposes structure on the problem that is not present in the physics. Therefore, according to a material theory of induction, we cannot use probabilities to represent our uncertainty. What should we use? The physics gives us the structure. It assigns three values to propositions concerning the system: impossible, possible and necessary. Just as we conform our degrees of confirmation to physical chances when they are present, we should take these three as the values possible for our degrees of confirmations. We shall abbreviate them as 'imp,' 'poss' and 'nec.' They are assigned to the propositions $E(T_1, T_2)$, with $T_1 < T_2$, which assert that the time T of spontaneous acceleration lies in $T_1 \leq T < T_2$; and to E_{no}, which just asserts that there is no spontaneous acceleration at any time.

---

[31] Figure from Norton ([2003a], §3).

We can then read the general logic from the physics in the obvious way for pairs of the propositions just defined:[32]

$[A|B]$ = nec   if $B \Rightarrow A$

       = imp   if $B \Rightarrow {\sim}A$

       = poss   otherwise

For example:

$[E(T_1,T_2)|\Omega]$ = poss

$[E_{no}|\Omega]$ = poss

$[E(T_1,T_2)|E(T_3,T_4)]$ = imp, if $[T_1,T_2)$ and $[T_3,T_4)$ do not intersect;

                           = poss, if $[T_3,T_4)$ partially intersects $[T_1,T_2)$;

                           = nec, if $[T_3,T_4)$ is contained in $[T_1,T_2)$.

       Reflecting on results such as these, it becomes apparent that the system conforms to *F. Framework*[33] and *N. Narrowness*. However both *A. Addition* and *M. Multiplication* fail, so that the *B. Bayes property* fails as well.

       To see the failure of *A. Addition*, consider the operator $\oplus'$ that comes closest to the operator $\oplus$ of *A. Addition*. It satisfies $[E_{no} \vee {\sim}E_{no}|\Omega] = [{\sim}E_{no}|\Omega] \oplus' [E_{no}|\Omega]$, which means that

$$\text{nec} = \text{poss} \ \oplus' \ \text{poss} \tag{9}$$

It also satisfies $[E(0,1) \vee E(1,2)|\Omega] = [E(0,1)|\Omega] \oplus' [E(1,2)|\Omega]$, which means that

$$\text{poss} = \text{poss} \ \oplus' \ \text{poss} \tag{9'}$$

Comparing (9) and (9') shows that $\oplus'$ cannot represent a function on the degrees and therefore cannot realize the operator $\oplus$ of *A. Addition*.

       To see the failure of *M. Multiplication*, consider the operator $\otimes'$ that comes closest to the operator $\otimes$ of *M. Multiplication*. For $E(0,1) \Rightarrow E(0,1) \Rightarrow \Omega$, we have $[E(0,1)|\Omega] = [E(0,1)|E(0,1)] \otimes' [E(0,1)|\Omega]$, so that

---

[32] It is sometimes said that a uniform prior over all natural numbers is a contrived fiction, since no physical mechanism could produce it. The dome provides that mechanism.

[33] To ensure that the set of propositions remains countable, consider only integer values of T_1 and T_2 in $E(T_1,T_2)$.

$$\text{poss} = \text{nec} \otimes' \text{poss} \tag{10}$$

However, for $E(0,1) \Rightarrow E(0,2) \Rightarrow \Omega$, we have $[E(0,1)|\Omega] = [E(0,1)|E(0,2)] \otimes' [E(0,2)|\Omega]$, so that

$$\text{poss} = \text{poss} \otimes' \text{poss} \tag{10'}$$

Comparing (10) and (10') shows that $\otimes'$ cannot be the operator $\otimes$ of *M. Multiplication*, since it is not strictly increasing in its first argument, even though the second is not $[\varnothing|\Omega]$. This failure eliminates an essential part of Bayes' theorem. For we can no longer invert the operator $\otimes'$ in the analog of of Bayes' theorem

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

to infer to $[H|E]$. Indeed conditionalizing H on E may yield an $[H|E]$ that is poss or nec, but the theorem will not be able to tell us which.

More directly, this example shows that, when $H \Rightarrow E$, $[H|E]$ is not a function of $[H|\Omega]$ and $[E|\Omega]$. Even though $H_1 = E(0,1)$ and $H_2 = E(0,2)$ have the same priors $[H_1|\Omega] = [H_2|\Omega] = \text{poss}$ and entail the same evidence $E = E(0,2)$, they have different posteriors. For $[H_1|E] = \text{poss}$, but $[H_2|E] = \text{nec}$. That is, the evidence E rewards H_1 and H_2 differentially, contrary to the dynamics governed by *B. Bayes property*, discussed in Section 5.3.3 above.

9. Conclusion

Conceived as a logic of inductive inference, the probability calculus represents the sum of several distinct ideas about inductive inference. Most notable are *A. Addition*, which asserts that low probabilities represent disbelief, and *B. Bayes property*, which requires that new evidence is incorporated by a process of 'refute and rescale.' The principal contention of this paper is that these properties and others like it are not warranted universally and must be invoked independently according to the needs of the problem at hand.

Acknowledgements

References

Aczel, J. [1966]: *Lectures on Functional Equations and their Applications,* New York: Academic Press.

Alper, J. S., Bridger, M., Earman, J. and Norton, J. D. [2000]: 'What is a Newtonian System? The Failure of Energy Conservation and Determinism in Supertasks,' *Synthese*, **124**, pp. 281-293.

Cox, R. T. [1961]: *The Algebra of Probable Inference,* Baltimore: Johns Hopkins Press.

de Finetti, B. [1937]: 'Foresight: Its Logical Laws, Its Subjective Sources,' trans. H. E. Kyburg, in H. E. Kyburg and H. Smokler (*eds*), *Studies in Subjective Probability*, New York: John Wiley & Sons, 1964, pp. 93-158.

Earman, J. [1992]: *Bayes or Bust,* Cambridge, MA: Bradford-MIT.

Fine, T. L. [1973]: *Theories of Probability*, New York: Academic Press.

Fishburn, P. C. [1986]: 'The Axioms of Subjective Probability,' *Statistical Science*, **1**, pp. 335-58.

Galavotti, M. C. [2005]: *Philosophical Introduction to Probability,* Stanford: CSLI Publications.

Gillies, D. [2000]: *Philosophical Theories of Probability*, London: Routledge.

Hawthorne, J. [1993]: 'Bayesian Induction is Eliminative Induction,' *Philosophical Topics*, **21**, pp. 99-138.

Howson, C. and Urbach, P. [1996]: *Scientific Reasoning: The Bayesian Approach*, 2nd ed. Chicago: Open Court.

Humphreys, P. [1985]: 'Why Propensities Cannot be Probabilities,' *Philosophical Review*, **94**, pp. 557-570.

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

Jeffrey, H. [1961]: *Theory of Probability*, 3rd ed. Oxford: Clarendon.

Keynes, J. M. [1921]: *A Treatise of Probability,* London: Macmillan; repr. New York, AMS Press, 1979.

Kyburg, H. E. [1976]: 'Local and Global Induction' in R. J. Bogdan (*ed.*), *Local Induction*, Dordrecht: Reidel, pp. 191-215.

Lewis, D. [1980]: 'A Subjectivist's Guide to Objective Chance,' in R.C. Jeffrey (*ed.*), *Studies in Inductive Logic and Probability*. Berkeley: University of California Press, pp. 263-93.

Mellor, D. H. [2005]: *Probability: A Philosophical Introduction,* London: Routledge.

Norton, J. D. [1999]: 'A Quantum Mechanical Supertask' *Foundations of Physics*, **29**, pp. 1265-1302.

Norton, J. D. [2003]: 'A Material Theory of Induction' *Philosophy of Science*, **70**, pp. 647-70.

Norton, J. D. [2003a]: 'Causation as Folk Science,' *Philosophers' Imprint*, **3**(No.4), <www.philosophersimprint.org/003004/>.

Norton, J. D. [2005]:'A Little Survey of Induction,' in P. Achinstein (*ed*)., *Scientific Evidence: Philosophical Theories and Applications*, Johns Hopkins University Press, 1905, pp. 9-34.

Norton, J. D. [forthcoming]: 'Einstein's Special Theory of Relativity and the Problems in the Electrodynamics of Moving Bodies that Led him to it,' in M. Janssen and C. Lehner (*eds*), *Cambridge Companion to Einstein*, Cambridge: Cambridge University Press.

Norton, J. D. [unpublished]: 'Ignorance and Indifference.'

Rosenkrantz, R. D. [1981]: *Foundations and Applications of Inductive Probability*, Atascadero, CA: Ridgeview.

Shafer, G. [1976]: *A Mathematical Theory of Evidence,* Princeton: Princeton University Press.

Smith, C. A. B. [1961]: 'Consistency in Statistical Inference and Decision,' *Journal of the Royal Statistical Society*, Series B, **23**, pp. 1-25.

Savage, L. J. [1972]: *The Foundations of Statistics,* Second revised ed. New York: Dover.

Tversky, A. and Kahneman, D. [1982]: 'Judgments of and by representativeness,' in D. Kahneman, P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge, UK: Cambridge University Press, pp. 84-98.