Commuting Probability Revisions: The Uniformity Rule

Carl G. Wagner (wagner@math.utk.edu)*
University of Tennessee, Knoxville

Abstract. A simple rule of probability revision ensures that the final result of a sequence of probability revisions is undisturbed by an alteration in the temporal order of the learning prompting those revisions. This *Uniformity Rule* dictates that identical learning be reflected in identical ratios of certain new-to-old odds, and is grounded in the old Bayesian idea that such ratios represent what is learned from new experience alone, with prior probabilities factored out. The main theorem of this paper includes as special cases (i) Field's theorem on commuting probability-kinematical revisions and (ii) the equivalence of two strategies for generalizing Jeffrey's solution to the old evidence problem to the case of uncertain old evidence and probabilistic new explanation.

1. Introduction

This paper explores a simple rule of probability revision that ensures that the final result of a sequence of probability revisions is undisturbed by an alteration in the temporal order of the learning prompting those revisions. This *Uniformity Rule* dictates that identical learning be reflected in identical ratios of certain new-to-old odds, and is grounded in the old Bayesian idea (Good 1985; Jeffrey 1992) that such ratios represent what is learned from new experience alone, with prior probabilities factored out.

The connection between uniform ratios of new-to-old odds and commuting probability revisions, detailed below in Theorem 2.1, can already be glimpsed in the well-known theorem of Field (1978) on commuting probability-kinematical revisions. Both Field's theorem, and its extension to countable partitions, are corollaries of Theorem 2.1. In Wagner 1997, 1999, 2001, Jeffrey's (1991, 1995) solution to the old evidence problem is generalized to the case of uncertain old evidence and probabilistic new explanation. The equivalence of two strategies for carrying out this generalization is also a corollary of Theorem 2.1.

Terminology and notation here are as follows: A σ-algebra $\mathbf{A}$ of subsets of Ω is *purely atomic* if the family $\mathbf{A}^*$ of atomic events[1] in $\mathbf{A}$ is countable, and constitutes a partition of Ω. Every finite σ-algebra is purely atomic, whatever the cardinality of Ω, and if Ω is countable,

* Research supported by the National Science Foundation (SES–9984005)

then every σ-algebra on Ω is purely atomic (Renyi 1970, Theorems 1.6.1, 1.6.2). If q is a revision of the probability measure p, and A, B, and C are events, then the *probability factor* (or *relevance quotient*) $\pi_p^q(A)$ is the ratio

$$\pi_p^q(A) := q(A)/p(A) \tag{1.1}$$

of new-to-old probabilities, the *Bayes factor* $\beta_p^q(A : B)$ is the ratio

$$\beta_p^q(A : B) := \frac{q(A)}{q(B)} \Big/ \frac{p(A)}{p(B)} \tag{1.2}$$

of new-to-old odds, and the *conditional Bayes factor* $\beta_p^q(A : B|C)$ is the ratio

$$\beta_p^q(A : B|C) := \frac{q(A|C)}{q(B|C)} \Big/ \frac{p(A|C)}{p(B|C)} \tag{1.3}$$

of new-to-old conditional odds. When $q(\cdot) = p(\cdot|E)$, then (1.2) is simply the likelihood ratio $p(E|A)/p(E|B)$. More generally

$$\beta_p^q(A : B) = \pi_p^q(A) \Big/ \pi_p^q(B) \ , \tag{1.4}$$

a simple, but useful, identity.

It is assumed throughout this paper that all probabilities are strictly coherent, i.e. that all nonempty events have positive probability. With the addition of certain technical conditions, however, the theorems presented here hold for arbitrary probabilities. The next section delineates the connection between uniform Bayes factors and commutativity. In sections 3 and 4 the Uniformity Rule is applied to observation- and explanation-based probability revisions. Section 5 examines three commonly encountered indices of probability change that might appear to offer alternatives to Bayes factors, and demonstrates how each fails to furnish an appropriate measure of what is learned from new experience.

2. Bayes Factors and Commutativity

Consider the probability revision schema

$$\begin{array}{ccc} p & \longrightarrow & Q \\ \Big\downarrow & & \downarrow \\ & & r \\ q & \longrightarrow & R \end{array} \ , \tag{2.1}$$

representing two possible sequential revisions of p. In the first, p is revised to q, which is then revised to R. In the second, p is revised to

Q, which is then revised to r. Suppose that the revisions of p to q, and of Q to r, are prompted by identical learning, and that the revisions of p to Q, and of q to R, are prompted by identical learning. Then it surely ought to be the case that $r = R$. Call this the *Commutativity Principle*.

The following theorem demonstrates that the Commutativity Principle is satisfied for purely atomic σ-algebras when identical learning is represented by identical Bayes factors at the level of atomic events.

Theorem 2.1. *Suppose that the probabilities in the revision schema* (2.1) *are defined on a purely atomic σ-algebra $\mathbf{A}$, with $\mathbf{A}^*$ denoting the set of atomic events in $\mathbf{A}$. If the Bayes factor identities*

$$\beta_Q^r(A:B) = \beta_p^q(A:B), \qquad \forall A, B \in \mathbf{A}^*, \qquad \textit{and} \tag{2.2}$$

$$\beta_q^R(A:B) = \beta_p^Q(A:B), \qquad \forall A, B \in \mathbf{A}^* \tag{2.3}$$

hold, then $r = R$.

Proof. Trivially, (2.3) is equivalent to

$$\beta_Q^R(A:B) = \beta_p^q(A:B), \qquad \forall A, B \in \mathbf{A}^*, \tag{2.4}$$

which, with (2.2), implies that

$$\beta_Q^r(A:B) = \beta_Q^R(A:B), \qquad \forall A, B \in \mathbf{A}^*. \tag{2.5}$$

Hence, $r = R$, since the r-odds and the R-odds on A against B are both gotten by multiplying the corresponding Q-odds by the same factor. □

Remark 2.1. From (1.4) it follows that (2.2) is equivalent to the probability factor proportionality

$$\pi_Q^r(A) \propto \pi_p^q(A), \qquad \forall A \in \mathbf{A}^*, \tag{2.6}$$

the proportionality constant being $\pi_Q^r(B) \big/ \pi_p^q(B)$, for arbitrary $B \in \mathbf{A}^*$. Similar remarks apply to (2.3).

Remark 2.2. From (2.6) and Theorem 2.1, we get the explicit formula

$$r(A) = R(A) = \frac{q(A)Q(A)}{p(A)} \Bigg/ \sum_{A \in \mathbf{A}^*} \frac{q(A)Q(A)}{p(A)}, \quad \forall A \in \mathbf{A}^*. \tag{2.7}$$

Thus, if p, q, Q, r, and R are well-defined and in place, and (2.2) and (2.3) hold, then, necessarily, the sum in the denominator of the right-hand side of (2.7) converges. If only p, q, and Q are in place at the outset and the aforementioned sum converges, then (2.7) defines probabilities

r and R satisfying (2.2) and (2.3). So (2.2) furnishes a recipe for constructing a probability r that would be the appropriate revision of Q if, in the probabilistic state Q, one were to undergo learning identical to that which prompted the revision of p to q. And (2.3) functions analogously in the construction of R. However, it is easy to construct examples where the sum in (2.7) fails to converge. Then there exists no probability r satisfying (2.2) and hence no probability R satisfying (2.3). In short, in the conceptual state reflected in q (respectively, Q), it is impossible to experience learning identical to that prompting the revision of p to Q (respectively, of p to q). This phenomenon is discussed further in section 5.

Remark 2.3. The commutativity issue arises not only for probability revisions, but also in belief revision theory. For a discussion of the latter in terms of ranking functions, see Spohn 1988.

Call the rule dictating that identical learning in purely atomic spaces be represented by identical Bayes factors at the level of atomic events the *Uniformity Rule*. By Theorem 2.1, adherence to this rule ensures satisfaction of the Commutativity Principle. In the next two sections we describe applications of this theorem to observation- and explanation-based probability revisions. In each of these cases the Uniformity Rule admits of natural, coarser-grained avatars.

3. Observation-Based Revision

Let $(\Omega, \mathbf{A}, p)$ be a probability space and let $\mathbf{E} = \{E_i\}$ be a countable partition of Ω, with each $E_i \in \mathbf{A}$. A probability measure q on $\mathbf{A}$ is said to come from p by *probability kinematics* on $\mathbf{E}$ (Jeffrey 1965, 1983) if

$$q(A) = \sum_i q(E_i)p(A|E_i), \qquad \forall A \in \mathbf{A}. \tag{3.1}$$

Formula (3.1) is easily seen to be equivalent to the *rigidity condition*

$$q(A|E_i) = p(A|E_i), \qquad \forall i,\ \forall A \in \mathbf{A}, \tag{3.2}$$

which is itself equivalent to the condition

$$q(A)/p(A) = q(E_i)/p(E_i), \qquad \forall i,\ \forall A \in \mathbf{A}:\ A \subset E_i. \tag{3.3}$$

Formula (3.1) embodies a two-stage revision of p in the light of new (typically, observational) evidence: (i) based on the total evidence, old as well as new, assign new probabilities $q(E_i)$ to members of the partition $\mathbf{E}$, and (ii) extend q to $\mathbf{A}$ by (3.1) after ascertaining that (3.2) holds, i.e., that nothing new has been learned about the relevance of any E_i to other events.

Consider the probability revision schema

$$
\begin{array}{ccc}
p & & Q \\
\mathbf{E}\big\downarrow & & \big\downarrow \text{identical learning} \\
q & & r
\end{array}
\tag{3.4}
$$

defined over a purely atomic σ-algebra $\mathbf{A}$, where q comes from p by probability kinematics on $\mathbf{E}$, and r is a revision of Q prompted by learning identical to that which prompted the revision of p to q. As shown by the following theorem, the fine-grained atomic representation of this identical learning dictated by the Uniformity Rule is equivalent here to a natural, coarser-grained representation.

Theorem 3.1. *Suppose that the probabilities in the revision schema* (3.4) *are defined on a purely atomic σ-algebra* $\mathbf{A}$. *Then the probability r satisfies the atomic-level Bayes factor identities*

$$\beta_Q^r(A:B) = \beta_p^q(A:B), \qquad \forall A, B \in \mathbf{A}^* \tag{3.5}$$

if and only if r comes from Q by probability kinematics on $\mathbf{E}$, *and*

$$\beta_Q^r(E_i:E_k) = \beta_p^q(E_i:E_k), \qquad \forall i, k. \tag{3.6}$$

Proof. Sufficiency. Let $A, B \in \mathbf{A}^*$ and suppose that $A \subset E_i$ and $B \subset E_k$. Then by (1.4), (3.3), and (3.6),

$$\beta_Q^r(A:B) = \beta_Q^r(E_i:E_k) = \beta_p^q(E_i:E_k) = \beta_p^q(A:B). \tag{3.7}$$

Necessity. By Remark 2.1, (3.5) implies that $r(A) \propto q(A)Q(A)/p(A)$, for all $A \in \mathbf{A}^*$. So if $E \in \mathbf{A}$,

$$r(E|E_i) = \sum_{A \subset EE_i}^{*} \frac{q(A)Q(A)}{p(A)} \Bigg/ \sum_{A \subset E_i}^{*} \frac{q(A)Q(A)}{p(A)} = Q(E|E_i), \tag{3.8}$$

(with $\sum^*$ indicating summation over $A \in \mathbf{A}^*$) since $q(A)/p(A) = q(E_i)/p(E_i)$ for $A \subset E_i$. So r comes from Q by probability kinematics on $\mathbf{E}$. To show (3.6), choose any $A, B \in \mathbf{A}^*$ such that $A \subset E_i$ and $B \subset E_k$. Then by (1.4), (3.3), and (3.5),

$$\beta_Q^r(E_i:E_k) = \beta_Q^r(A:B) = \beta_p^q(A:B) = \beta_p^q(E_i:E_k). \tag{3.9}$$

□

Consider now the revision schema

$$
\begin{array}{ccc}
p & \xrightarrow{\mathbf{F}} & Q \\
\Big\downarrow{\scriptstyle \mathbf{E}} & & \big\downarrow{\scriptstyle \mathbf{E}} \\
 & & r \\
q & \xrightarrow[\mathbf{F}]{} & R
\end{array}, \tag{3.10}
$$

illustrating two possible sequential revisions of p. In the first, p is revised to q by probability kinematics on $\mathbf{E} = \{E_i\}$, and then q is revised to R by probability kinematics on a partition $\mathbf{F} = \{F_j\}$. In the second, probability kinematics is carried out first on $\mathbf{F}$, yielding Q, and then on $\mathbf{E}$, yielding r. The partitions $\mathbf{E}$ and $\mathbf{F}$ may or may not coincide, and the sequences $(r(E_i))$ and $(R(F_j))$ may or may not differ, respectively, from the sequences $(q(E_i))$ and $(Q(F_j))$. Nothing is assumed about the σ-algebra on which the probabilities in question are defined. In particular, it may fail to be purely atomic. The following theorem gives conditions sufficient to ensure that $r = R$.

Theorem 3.2. *Let the probabilities in* (3.10) *be defined on an arbitrary* σ*-algebra* $\mathbf{A}$. *If*

$$\beta_Q^r(E_{i_1} : E_{i_2}) = \beta_p^q(E_{i_1} : E_{i_2}), \qquad \forall i_1, i_2, \qquad \textit{and} \tag{3.11}$$

$$\beta_q^R(F_{j_1} : F_{j_2}) = \beta_p^Q(F_{j_1} : F_{j_2}), \qquad \forall j_1, j_2, \tag{3.12}$$

then $r = R$.

Proof. It suffices to show that $r(G) = R(G)$ for every $G \in \mathbf{A}$. Let $\mathbf{B}$ be the σ-algebra generated by $\mathbf{E} \cup \mathbf{F} \cup \{G\}$. Clearly, $\mathbf{B}$ is purely atomic, the family $\mathbf{B}^*$ of atomic events of $\mathbf{B}$ comprising all nonempty events of the form E_iF_jG or $E_iF_j\bar{G}$. From (3.11), (3.12) and the sufficiency part of Theorem 3.1, it follows that the hypotheses of Theorem 2.1 are satisfied for $\mathbf{B}$. Hence $r = R$ on $\mathbf{B}$, and so $r(G) = R(G)$. □

Remark 3.1. Theorem 3.2 was proved for finite partitions $\mathbf{E}$ and $\mathbf{F}$ in Field 1978 and for countable partitions in Wagner 2002. Field's proof involved, inter alia, reformulating (3.1), where $E = \{E_1, \ldots, E_m\}$, as

$$q(A) = \sum_{i=1}^{m} G_i p(AE_i) \Big/ \sum_{i=1}^{m} G_i p(E_i), \tag{3.13}$$

where G_i is the geometric mean

$$G_i := \left(\prod_{k=1}^{m} \beta_p^q(E_i : E_k) \right)^{1/m}, \tag{3.14}$$

and Wagner's proof reformulating (3.1) as

$$q(A) = \sum_i B_i p(AE_i) \bigg/ \sum_i B_i p(E_i) \ , \tag{3.15}$$

where

$$B_i := \beta_p^q(E_i : E_1). \tag{3.16}$$

Remark 3.2. It is perhaps worth noting that, conversely, Theorem 2.1 can be derived from Theorem 3.2. For if p and q are *any* strictly coherent probability measures on a purely atomic σ-algebra $\mathbf{A}$, then q comes from p by probability kinematics on the partition $\mathbf{A}^*$ comprising all atomic events of $\mathbf{A}$ (the conditional probabilities in the rigidity condition (3.2) take only the values 0 and 1 in this case). It would be a mistake to conclude from this, however, that these theorems are equally fundamental. First, the proof of Theorem 2.1 is simple and transparent. Second, construing arbitrary probability revisions as probability-kinematical, while technically correct, is intuitively artificial (the explanation-based revisions of section 4 below are a perfect example). Most decisive, however, is the fact that in their fullest generality, these theorems are no longer mutually derivable. The extension of Theorem 3.2 beyond the realm of strictly coherent probabilities can still be derived from the corresponding extension of Theorem 2.1, but not conversely. Hence Theorem 2.1 is the fundamental result.

4. Explanation-Based Revision

Probability kinematics involves revisions arising from the assignment of certain new unconditional probabilities, while maintaining certain conditional probabilities. We consider here the orthogonal situation in which revisions arise from the assignment of certain new conditional probabilities, while maintaining certain unconditional probabilities.

Specifically, let p be a probability on the algebra $\mathbf{A}$ generated by hypothesis H and evidence E, whence $\mathbf{A}^* = \{HE, H\bar{E}, \bar{H}E, \bar{H}\bar{E}\}$. Suppose that theoretical analysis reveals a new explanatory connection between H and E, and $\bar{H}$ and E, prompting a revision of p to the probability Q uniquely determined by the conditions

$$Q(E|H) = u, \tag{4.1}$$

$$Q(E|\bar{H}) = v, \qquad \text{and} \tag{4.2}$$

$$Q(H) = p(H). \tag{4.3}$$

Now consider the probability revision schema

$$\begin{array}{ccc} p & \xrightarrow{\quad u,v \quad} & Q \\ & & \\ q & \xrightarrow{\qquad\quad} & R \end{array} \quad , \tag{4.4}$$

identical

learning

where p and Q are as above, and R is a revision of q prompted by the same explanatory learning that prompted the revision of p to Q. As shown by the following theorem, the atomic representation of this identical learning dictated by the Uniformity Rule is equivalent here to three natural conditional Bayes factor identities.

Theorem 4.1. *In the revision schema* (4.4) *the probability R satisfies the atomic-level Bayes factor identities*

$$\beta_q^R(A : B) = \beta_p^Q(A : B), \qquad \forall A, B \in \mathbf{A}^* \tag{4.5}$$

if and only if it satisfies the conditional Bayes factor identities

$$\beta_q^R(E : \bar{E}|H) \ = \ \beta_p^Q(E : \bar{E}|H), \tag{4.6}$$

$$\beta_q^R(E : \bar{E}|\bar{H}) \ = \ \beta_p^Q(E : \bar{E}|\bar{H}), \qquad \textit{and} \tag{4.7}$$

$$\beta_q^R(H : \bar{H}|E) \ = \ \beta_p^Q(H : \bar{H}|E). \tag{4.8}$$

Proof. It is straightforward to verify that (4.6) is equivalent to the case $A = HE$ and $B = H\bar{E}$, (4.7) to the case $A = \bar{H}E$ and $B = \bar{H}\bar{E}$, and (4.8) to the case $A = HE$ and $B = \bar{H}E$ of (4.5). It is simply a matter of tedious algebra to check that these three cases of (4.5) imply that (4.5) holds for all $A, B \in \mathbf{A}^*$. $\square$

A basic principle of scientific inference asserts that if hypothesis H is known to imply the less-than-certain proposition E, the subsequent discovery that E is true confirms (i.e. raises the probability of) H. There is a simple Bayesian account of such confirmation, for from $p(E|H) = 1 > p(E)$ it follows that $p(H|E) > p(H)$. Suppose, however, that we *first* become certain of E and *subsequently* discover, quite apart from this certainty, that H implies E. There are numerous examples of this in the history of science, one of the best known being Einstein's explanation of the previously observed "anomalous" advance in the perihelion of Mercury by means of the general theory of relativity (see

Weinberg 1992). Just as it does when explanation precedes observation, the explanation of a *previously known* fact E by H ought to confirm H, but how?

This problem, posed by Glymour (1980), has been variously termed the *historical old evidence problem* (Garber 1983), the *problem of new old evidence* (Eells 1990), the *problem of the confirmation event* (Zynda 1995), the *problem of new explanation* (Jeffrey 1995), and the *problem of logical learning* (Joyce 1999).[2] In this case the prior q has the property that $q(E) = 1$, so it is ineffectual to condition q on E (since $q(H|E) = q(H)$) and, in any case, not to the point, since what is required is a revision of q based on the discovery that H implies E, not the discovery that E is true.

Accordingly, one proposed solution to this problem, due to Garber (1983), extends the algebra on which probabilities are defined to include the proposition $H \vdash E$ that H implies E. Under certain conditions, $q(H|H \vdash E) > q(H)$. On the other hand, Jeffrey (1991, 1995) has proposed a solution that retains the original algebra, but revises probabilities by an entirely new method called *reparation.* Central to Jeffrey's approach is the imaginative reconstruction of a probability distribution that predates both our certainty about E and our discovery that H implies E. The explanation-based revision of this ur-distribution then serves as a paradigm for all explanation-based revisions.

In Wagner 1997, 1999, 2001, Jeffrey's solution was shown to generalize in a natural way to cases in which observation raises our confidence in E without rendering it certain, and the subsequent explanation of E by H is probabilistic rather than implicational. As we demonstrate below, this generalization can be given a simple and unified formulation in terms of the Uniformity Rule.

Specifically, suppose that q is our current probability distribution on $\mathbf{A}$, the algebra generated by E and H. Empirical investigation (*observation*) has given us a certain measure of confidence in E, as reflected in the value $q(E)$. We subsequently discover, quite apart from this observation, theoretical considerations that, taken alone, indicate that the truth of H would confer probability u on E, and the truth of $\bar{H}$ would confer probability v on E (*explanation*). Taken alone, these considerations do not alter the probability of H. How should q be revised in light of this new probabilistic explanation?

Following Jeffrey, we resurrect a notional *ur-distribution* p, predating both observation and explanation. It is assumed that $p(A) > 0$ for all $A \in \mathbf{A}^*$.[3] Generalizing Jeffrey's assumption that $q(\cdot) = p(\cdot|E)$, we assume here that q comes from p by probability kinematics on $\mathbf{E} = \{E, \bar{E}\}$. In effect, it is in the conceptual state captured by p that we make the aforementioned theoretical discovery, a discovery that would warrant the revision of p to Q, as defined by (4.1)–(4.3). Two candidates for the desired revision q now present themselves:

(4.9)
$$\begin{tikzcd}
p \arrow[r, "\substack{\text{explanation}\\ u,v}"] \arrow[dd, "\text{observation } \mathbf{E}"'] & Q \arrow[d, "\text{observation}"] \\
 & r \\
q \arrow[ur] \arrow[r, "\text{explanation}"'] & R
\end{tikzcd}$$

This revision is to be based on explanatory learning identical to that which we imagined prompting the revision of the ur-distribution p to Q. Accordingly, the Uniformity Rule would dictate revising q to R, as defined by (4.5), or, equivalently (by Theorem 4.1), by (4.6)–(4.8). On the other hand, we might continue the imaginative exercise of exchanging the temporal order of observation and explanation, taking as the explanation-based revision of the observation-based revision q of p the appropriate observation-based revision of the explanation-based revision Q of p. This revision of Q, call it r, is to be based on the same observational learning that we imagined prompting the revision of p to q. Accordingly, the Uniformity Rule would dictate that r be defined by (3.5), or, equivalently (by Theorem 3.1), that r come from Q by probability kinematics on E, with

$$\beta_Q^r(E : \bar{E}) = \beta_p^q(E : \bar{E}). \tag{4.10}$$

Fortunately, one need not choose between r and R since, by Theorem 2.1, these probabilities are identical.

If we denote the common value of the Bayes factors in (4.10) by β, then (Wagner, 2001)

$$\frac{R(H)}{R(\bar{H})} = \frac{p(H)}{p(\bar{H})}\frac{[(\beta-1)Q(E|H)+1]}{[(\beta-1)Q(E|\bar{H})+1]}, \qquad \text{and} \tag{4.11}$$

$$\frac{q(H)}{q(\bar{H})} = \frac{p(H)}{p(\bar{H})}\frac{[(\beta-1)p(E|H)+1]}{[(\beta-1)p(E|\bar{H})+1]}. \tag{4.12}$$

In short, the new odds on H are simply gotten from the old odds on H, as given by (4.12), by replacing the ur-likelihoods $p(E|H)$ and $p(E|\bar{H})$ by $Q(E|H) = u$ and $Q(E|\bar{H}) = v$. From (4.11) and (4.12) one can derive several simple and intuitively reasonable conditions that ensure that H is confirmed $(R(H) > q(H))$. See Wagner 1999, 2001 for details. In particular, in the special case treated by Jeffrey, where $Q(E|H) = 1$,

$Q(E|\bar{H}) = p(E|\bar{H})$ and $q(\cdot) = p(\cdot|E)$, H is always confirmed, since

$$\beta_q^R(H : \bar{H}) = \frac{1}{p(E|H)} > 1. \tag{4.13}$$

Formula (4.13) neatly accounts for the strong confirmation that the general theory of relativity (H) received from Einstein's discovery that H implied the known advance in the perihelion of Mercury (E). For it would surely have been the case that, prior to observing E or learning that H implied E, one would have assigned the peculiar phenomenon E very small conditional probability, given H.

5. Discussion

We conclude by examining three familiar indices of probability change that might appear to offer alternatives to Bayes factors, namely, the difference $\delta_p^q(A) := q(A) - p(A)$, the normalized difference $D_p^q(A) = (q(A) - p(A))/p(A)$, and the probability factor (or relevance quotient) $\pi_p^q(A) := q(A)/p(A)$. In what follows we evaluate these indices with respect to three criteria that clearly should be satisfied by any representation of new learning:

I. The representation should ensure satisfaction of the Commutativity Principle

II. The representation should ensure that learning identical to that prompting a probability-kinematical revision should prompt a probability-kinematical revision on the same partition.

III. The representation of new learning should not unduly restrict the set of priors amenable to revision in response to such learning.

Since $D_p^q(A) = \pi_p^q(A) - 1$, the indices D and π clearly stand or fall together, and so it will suffice in what follows to restrict attention to d and π.

As some readers may already have observed, both d and π satisfy I, i.e., both for $\nu = d$ and for $\nu = \pi$, Theorem 2.1 continues to hold when the Bayes factor identities (2.2) and (2.3) are replaced, respectively, by

$$\nu_Q^r(A) = \nu_p^q(A), \qquad \forall A \in \mathbf{A}^*, \text{ and} \tag{5.1}$$

$$\nu_q^R(A) = \nu_p^Q(A), \qquad \forall A \in \mathbf{A}^*. \tag{5.2}$$

These results are, however, less impressive than they seem, as will be clear from our discussion of criterion III below.

As for criterion II, this crucial part of Theorem 3.1 holds when the Bayes factor identities (3.5) are replaced by (5.1) with $\nu = \pi$, but not with $\nu = d$.[4] But here again the fact that π satisfies II is less impressive than it seems.

Criterion III is motivated by observations made in Remark 2.2, which emphasized that our real interest in an identity such as, say, (5.1) lies in its potential, given p, q, and Q, to provide a recipe for constructing a probability r that is the appropriate revision of Q based on learning identical to that prompting the revision of p to q. As noted in Remark 2.2, this potential may fail to be realized. Again taking (5.1) as an illustration, it may be that for certain probabilities Q there exists no probability r satisfying (5.1). To the extent that this is true for a substantial number of priors Q, ν-analogues of Theorems 2.1 and 3.1 will in a substantial number of cases involve implications that are vacuously true.

This is exactly the case for $\nu = d$ and $\nu = \pi$. Since $d_Q^r(A) = c$ implies that $-c \leqslant Q(A) \leqslant 1 - c$, it follows that the existence of single $A \in \mathbf{A}^*$ such that $Q(A) > 1 - d_p^q(A)$ precludes the existence of any probability r satisfying (5.1) for $\nu = d$. Since $\pi_Q^r(A) = c$ implies that $Q(A) \leqslant 1/c$, it follows that the existence of a single $A \in \mathbf{A}^*$ such that $Q(A) > 1 \Big/ \pi_p^q(A)$ precludes the existence of any probability r satisfying (5.1) for $\nu = \pi$.

In the latter case this phenomenon reaches the point of absurdity when $\mathbf{A}^*$ has just two members. Then, unless $Q = p$, there is no r satisfying (5.1) for $\nu = \pi$.[5] Adopting π as a measure of what is learned from new experience would lead here to the astonishing conclusion that if Q differs from p in the slightest degree, then it is impossible to undergo identical new learning in the conceptual states reflected, respectively, in Q and p.

To summarize, the index d fails to satisfy II, and its satisfying I is vitiated by its failure to satisfy III. The index π satisfies both I and II, but this is vitiated by its failure to satisfy III.[6] As we indicated in Remark 2.2, certain priors may be incompatible with certain kinds of new learning, even when such learning is measured by Bayes factors. But this phenomenon never materializes when there are only finitely many atomic events, and when $\mathbf{A}^* = \{A_1, A_2, \ldots\}$ is infinite, the phenomenon is rather benign.

Suppose, for example, that we are given a prior p and a sequence (β_i) of positive real numbers such that $\beta_1 = 1$. It is straightforward to show that there exists a probability q satisfying

$$\beta_p^q(A_i : A_1) = \beta_i, \qquad i = 1, 2, \ldots \tag{5.3}$$

[7]

if and only if

$$\sum_i \beta_i p(A_i) < \infty, \tag{5.4}$$

in which case

$$q(A_i) = \beta_i p(A_i) \Big/ \sum_i \beta_i p(A_i) , \quad i = 1, 2, \ldots . \tag{5.5}$$

Given the sequence (β_i) there may exist probabilities p for which the convergence condition (5.4) fails to hold. But this is mitigated by the following result:

Theorem 5.1. *Let* $(\beta_1, \beta_2, \ldots)$ *be any sequence of positive real numbers such that* $\beta_1 = 1$ *and let* r *be any real number such that* $0 < r < 1$*. Then there exists a probability* p *such that* $p(A_1) = r$ *and the convergence condition* (5.4) *holds, and so there exists a probability* q *satisfying* (5.3)*.*

Proof. Set $p(A_1) = r$. For $i \geqslant 3$, set $p(A_i) = (1 - r)2^{-i}$ if $\beta_i \leqslant 1$ and set $p(A_i) = (1 - r)2^{-i}/\beta_i$ if $\beta_i > 1$. Then

$$\sum_{i \geqslant 3} p(A_i) \leqslant (1 - r) \sum_{i \geqslant 3} 2^{-i} = (1 - r)/4, \tag{5.6}$$

and so the left hand sum in (5.6) converges to some $s \leqslant (1-r)/4$. Since $r < 1$, $r + s \leqslant r + (1 - r)/4 < 1$, and we may set $p(A_2) = 1 - r - s$. The probability p clearly satisfies (5.4) and so (5.5) defines a revision q of p satisfying (5.3). □

It is clear from the above proof that in the statement of Theorem 5.1 the condition $p(A_1) = r$ could be replaced by the condition $p(A_j) = r$, for a fixed, but arbitrary, j. So the import of this theorem is that while certain Bayes factors may be incompatible with taking certain probability *distributions* as priors, they do not constrain the prior probability of any particular atomic *event*. With this observation we rest our case for Bayes factors and the Uniformity Rule.

Notes

1. An event $A \in \mathbf{A}$ is *atomic* if it is nonempty, and no proper, nonempty subset of A belongs to $\mathbf{A}$.

2. This problem should not be confused with the *ahistorical problem of old evidence* (also called the *problem of old new evidence*, the *problem of the confirmation relation*, and the *problem of evidential relevance*), which asks for a measure of incremental confirmation which, unlike $p(H|E) - p(H)$ and similar measures, remains undisturbed when p is revised to $q(\cdot) = p(\cdot|E)$. We have nothing to say here about this problem, which has been discussed by Skyrms (1983), Joyce (1999), and Eells and Fitelson (2000).

3. In the ur-conceptual state represented by p we know nothing about H, or about $\bar{H}$, that would preclude the truth of E, or of $\bar{E}$. Hence the assumption that p is positive on $\mathbf{A}^*$ is eminently reasonable.

4. Consider probabilities p, q, p', and q' defined on the algebra $\mathbf{A}$ of events generated by H and E as follows

	HE	$\bar{H}E$	$H\bar{E}$	$\bar{H}\bar{E}$
$p:$	.4	.1	.1	.4
$q:$	.64	.16	.04	.16
$p':$	.2	.2	.3	.3
$q':$	.44	.26	.24	.06

While q comes from p by probability kinematics on the partition $\{E, \bar{E}\}$ and $q'(A)-p'(A) = q(A)-p(A)$ for all $A \in \mathbf{A}^* = \{HE, \bar{H}E, H\bar{E}, \bar{H}\bar{E}\}$, q' does not come from p' by probability kinematics on $\{E, \bar{E}\}$ since, for example, $q'(H|E) \neq p'(H|E)$.

5. Suppose, given probabilities p, q, and Q, that there exists an r satisfying (5.1) for $\nu = \pi$, where $\mathbf{A}^* = \{A, \bar{A}\}$. Let $p(A) = \alpha$, $q(A) = \beta$, $Q(A) = \alpha'$, and $r(A) = \beta'$. Then (5.1) implies that (i) $\beta'/\alpha' = \beta/\alpha$ and (ii) $(1-\beta')/(1-\alpha') = (1-\beta)/(1-\alpha)$. From (i) and (ii) we have $\beta' + 1 - \beta' = 1 = \alpha'\beta/\alpha + (1-\alpha')(1-\beta)/(1-\alpha)$, and solving this equation for α' yields $\alpha' = \alpha$, whence $Q = p$.

6. Furthermore, to the extent that π-analogues of Theorems 2.1 and 3.1 are not vacuously true, they are simply corollaries of these theorems, as is clear from Remark 2.1.

7. Note that once the values β_i are specified, the values of all atomic level Bayes factors are determined, since

$$\beta_p^q(A_i : A_j) = \beta_p^q(A_i : A_1) \Big/ \beta_p^q(A_j : A_1) = \beta_i/\beta_j.$$

References

Eells, E.: 1990, 'Bayesian Problems of Old Evidence'. In: C. Savage (ed.): *Scientific Theories*, *Minnesota Studies in the Philosophy of Science* XIV. Minneapolis: University of Minnesota Press. pp. 205–223.

Eells, E. and Fitelson, B.: 2000, 'Measuring Confirmation and Evidence', *The Journal of Philosophy* 97, 663–672.

Field, H.: 1978, 'A Note on Jeffrey Conditionalization', *Philosophy of Science* 45: 361–367.

Garber, D.: 1980, 'Field and Jeffrey Conditionalization', *Philosophy of Science* 47: 142–145.

Garber, D.: 1983, ‘Old Evidence and Logical Omniscience in Bayesian Confirmation Theory’. In: J. Earman (ed.): *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science* X. Minneapolis: University of Minnesota Press. pp. 99–131.

Good, I.: 1985, ‘Weight of Evidence: A Brief Survey.’ In: J. Bernardo et al (eds.): *Bayesian Statistics* 2. North-Holland, pp. 249–370.

Glymour, C.: 1980, *Theory and Evidence*. Princeton: Princeton University Press.

Jeffrey, R.: 1965, *The Logic of Decision*. New York: McGraw-Hill; 1983, 2nd ed. Chicago: University of Chicago Press.

———— : 1991, ‘Postscript 1991: New Explanation Revisited’. In Jeffrey 1992, 103–107.

———— : 1992, *Probability and the Art of Judgement*. Cambridge: Cambridge University Press.

———— : 1995, ‘Probability Reparation: The Problem of New Explanation’, *Philosophical Studies* 77: 97–102.

Joyce, J.: 1999, *The Foundations of Casual Decision Theory*. Cambridge: Cambridge University Press. pp. 200–215.

Renyi, A.: 1970, *Foundations of Probability*. San Francisco: Holden-Day.

Skyrms, B.: 1983, ‘Three Ways to Give a Probability Assignment a Memory’. In: J. Earman (ed.): *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science* X. Minneapolis: University of Minnesota Press, pp. 157–161.

Spohn, W.: 1988, ‘Ordinal Conditional Functions: A Dynamic Theory of Epistemic States’. In: W. Harper and B. Skyrms (eds.): *Causation, Belief Change, and Statistics* II. Dordrecht: Kluwer, pp. 105–134.

Wagner, C.: 1997, ‘Old Evidence and New Explanation’, *Philosophy of Science* 64: 677–691.

———— : 1999, ‘Old Evidence and New Explanation II’, *Philosophy of Science* 66: 283–288.

———— : 2001, ‘Old Evidence and New Explanation III’, *Philosophy of Science* 68 (Proceedings): S165–S175.

———— : 2002, ‘Probability Kinematics and Commutativity’, *Philosophy of Science* 69: 266–278

Weinberg, S.: 1992, *Dreams of a Final Theory*. New York: Pantheon Press.

Zynda, L.: 1995, ‘Old Evidence and New Theories’, *Philosophical Studies* 77: 67–96.