Probability Kinematics and Commutativity

Carl G. Wagner*
Department of Mathematics
The University of Tennessee
Knoxville, TN 37996–1300

Abstract

The so-called "non-commutativity" of probability kinematics has caused much unjustified concern. When identical learning is properly represented, namely, by identical Bayes factors rather than identical posterior probabilities, then sequential probability-kinematical revisions behave just as they should. Our analysis is based on a variant of Field's reformulation of probability kinematics, divested of its (inessential) physicalist gloss.

1 Introduction

The much remarked "non-commutativity" of probability kinematics (Domotor 1982, Skyrms 1986, van Fraassen 1989, Döring 1999, Lange 2000) has evoked varying degrees of concern. In this paper it is shown that when identical learning is properly represented, namely, by identical Bayes factors rather than identical posterior probabilities, then sequential probability-kinematical revisions behave just as they should.

Our analysis, which unifies and extends results in Field 1978, Diaconis and Zabell 1982, and Jeffrey 1988, is based on a variant of Field's reformulation of probability kinematics, divested of its (inessential) physicalist gloss. In § 2 a brief review of probability kinematics is presented. In § 3 we extend Field's Theorem to countable partitions, showing that the uniformity

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

of Bayes factors in the representation of identical learning is sufficient for the commutativity of probability kinematics. In § 4 it is shown that under mild restrictions such uniformity is also necessary for commutativity. In § 5 we discuss the methodological and philosophical implications of these theorems.

The notational conventions of this paper are as follows: If q is a revision of the probability measure p and A and B are events, the *Bayes factor* (or *odds factor*) $\beta_{q,p}(A : B)$ is the ratio

$$\beta_{q,p}(A : B) := \frac{q(A)}{q(B)} \bigg/ \frac{p(A)}{p(B)} \tag{1.1}$$

of new-to-old odds, and the *probability factor* (or *relevance quotient*) $\pi_{q,p}(A)$ is the ratio

$$\pi_{q,p}(A) := q(A)/p(A) \tag{1.2}$$

of new-to-old probabilities. When q comes from p by conditionalization on the event E, then (1.1) is simply the *likelihood ratio* $p(E|A)/p(E|B)$. More generally,

$$\beta_{q,p}(A : B) = \pi_{q,p}(A)/\,\pi_{q,p}(B), \tag{1.3}$$

a simple, but useful, identity.

2 Probability Kinematics

Let $(\Omega, \mathbf{A}, p)$ be a probability space, and suppose that $\mathbf{E} = \{E_i\}$ is a countable family of pairwise disjoint events, with $p(E_i) > 0$ for all i. A probability measure q is said to come from p by *probability kinematics* on $\mathbf{E}$ (Jeffrey 1965, 1983) if there exists a sequence (e_i) of positive real numbers summing to one, such that

$$q(A) = \sum_i e_i p(A|E_i), \qquad \text{for all } A \in \mathbf{A}.\text{[1]} \tag{2.1}$$

If $\mathbf{E} = \{E\}$, then $q(A) = p(A|E)$, and so probability kinematics is a generalization of ordinary conditionalization.

Formula (2.1) is clearly equivalent to the conjunction of the conditions

$$q(E_i) = e_i, \qquad \text{for all } i, \text{ and} \tag{2.2}$$

$$q(A|E_i) = p(A|E_i), \qquad \text{for all } A \in \mathbf{A} \text{ and for all } i. \tag{2.3}$$

Then (2.1) defines the appropriate revision of p in the light of new evidence when the total evidence, *old as well as new*,[2] prompts us to revise the probabilities of the events E_i as specified by (2.2), but we learn nothing new about the relevance of any E_i to other events. Condition (2.3), which captures the latter state of affairs, is called the *rigidity* (or *sufficiency*) *condition*.

Having revised p to q by the probability-kinematical formula (2.1) above, let us consider a subsequent revision of q to r by the formula

$$(2.4) \qquad r(A) = \sum_j f_j q(A|F_j), \qquad \text{for all } A \in \mathbf{A},$$

where $\mathbf{F} = \{F_j\}$ is a countable family of pairwise disjoint events such that $q(F_j) > 0$ for all j, and (f_j) is a sequence of positive real numbers summing to one. Now imagine reversing the order of the foregoing, first revising p to, say, q' by the formula

$$(2.5) \qquad q'(A) = \sum_j f_j p(A|F_j), \qquad \text{for all } A \in \mathbf{A},$$

and then revising q' to, say, r' by the formula

$$(2.6) \qquad r'(A) = \sum_i e_i q'(A|E_i), \qquad \text{for all } A \in \mathbf{A}.$$

Symbolically,

$$(2.7) \qquad \begin{array}{ccc} p & \xrightarrow{\mathbf{E},e_i} & q \\ {\scriptstyle \mathbf{F},f_j}\Big\downarrow & & \Big\downarrow{\scriptstyle \mathbf{F},f_j} \\ & & r \\ q' & \xrightarrow[\mathbf{E},e_i]{} & r' \end{array} .$$

Unless $\mathbf{E} = \{E\}$ and $\mathbf{F} = \{F\}$, in which case $r'(A) = r(A) = p(A|EF)$, it may well be the case that $r' \neq r$. The possibility of such "non-commutativity" has been the source of much confusion and unjustified concern. In the next two sections we lay the foundations for clarifying this issue with two theorems, delineating their methodological and philosophical implications in §5.

3 Field's Theorem

A proper analysis of the commutativity issue requires consideration of the more general probability-kinematical revision schema

$$(3.1)\qquad \begin{array}{ccc} p & \xrightarrow{\mathbf{E}} & q \\ \Big\downarrow{\scriptstyle \mathbf{F}} & & \downarrow{\scriptstyle \mathbf{F}} \\ & & r \\ q' & \xrightarrow[\mathbf{E}]{} & r' \end{array},$$

where the sequence $\big(r'(E_i)\big)$ may differ from $\big(q(E_i)\big)$, and the sequence $\big(q'(F_j)\big)$ from $\big(r(F_j)\big)$. Field (1978) was the first to identify conditions sufficient to ensure that $r' = r$ in this setting, in the special case where $\mathbf{E}$ and $\mathbf{F}$ are finite. In fact, Field's result holds for all countable families $\mathbf{E}$ and $\mathbf{F}$, but requires a different proof.

Theorem 3.1. *Given the probability revision schema* (3.1)*, if the Bayes factor identities*

$$(3.2)\qquad \beta_{r',q'}(E_{i_1} : E_{i_2}) = \beta_{q,p}(E_{i_1} : E_{i_2}), \qquad \textit{for all } i_1, i_2,$$

and

$$(3.3)\qquad \beta_{q',p}(F_{j_1} : F_{j_2}) = \beta_{r,q}(F_{j_1} : F_{j_2}), \qquad \textit{for all } j_1, j_2,$$

hold, then $r' = r$.

Proof. It follows from (2.1) and (2.2) that for all $A, A' \in \mathbf{A}$

$$(3.4)\qquad q(A)\,/q(A') = \sum_i \frac{q(E_i)}{p(E_i)} p(AE_i) \Bigg/ \sum_i \frac{q(E_i)}{p(E_i)} p(A'E_i).$$

Dividing the numerator and denominator of the right-hand side of (3.4) by $q(E_1)/p(E_1)$, and setting $A' = \Omega$, then yields the formula

$$(3.5)\qquad q(A) = \sum_i B_i p(AE_i) \Bigg/ \sum_i B_i p(E_i),$$

where

$$B_i := \beta_{q,p}(E_i : E_1). \tag{3.6}$$

With $B'_i := \beta_{r',q'}(E_i : E_1)$, $b_j := \beta_{r,q}(F_j : F_1)$, and $b'_j := \beta_{q',p}(F_j : F_1)$, it follows from (3.5), along with analogous formulas for r, q', and r' that

$$r(A) = \sum_{i,j} B_i b_j p(AE_iF_j) \Big/ \sum_{i,j} B_i b_j p(E_iF_j) \tag{3.7}$$

and

$$r'(A) = \sum_{i,j} B'_i b'_j p(AE_iF_j) \Big/ \sum_{i,j} B'_i b'_j p(E_iF_j). \tag{3.8}$$

Since (3.2) implies (indeed, is equivalent to) $B'_i = B_i$ and (3.3) implies (indeed, is equivalent to) $b'_j = b_j$, it follows that $r' = r$. □

Remark 3.1. Field's proof of Theorem 3.1, with **E** and **F** finite and $\mathbf{E} = \{E_1, \ldots, E_m\}$, involved reformulating (2.1) as

$$q(A) = \sum_{i=1}^{m} G_i p(AE_i) \Big/ \sum_{i=1}^{m} G_i p(E_i), \tag{3.9}$$ [3]

where G_i is the geometric mean

$$G_i := \left(\prod_{k=1}^{m} \beta_{q,p}(E_i : E_k) \right)^{1/m}. \tag{3.10}$$

Remark 3.2. By (1.3), the Bayes factor identities (3.2) and (3.3) are equivalent, respectively, to the probability factor proportionalities

$$\pi_{r',q'}(E_i) \propto \pi_{q,p}(E_i), \quad \text{for all } i, \text{ and} \tag{3.11}$$

$$\pi_{q',p}(F_j) \propto \pi_{r,q}(F_j), \quad \text{for all } j. \tag{3.12}$$ [4]

Remark 3.3. Jeffrey (1988) showed that if **E** and **F** are finite and r and r' are defined by (3.1), then the probability factor identities

$$\pi_{r',q'}(E_i) = \pi_{q,p}(E_i), \quad \text{for all } i, \text{ and} \tag{3.13}$$

$$\pi_{q',p}(F_j) = \pi_{r,q}(F_j), \quad \text{for all } j, \tag{3.14}$$

imply that $r' = r$. In light of the above remark, Jeffrey's result is a corollary of Theorem 3.1.

In the special case of schema (3.1) represented by (2.7), Bayes factor and probability factor identities are equivalent to each other, as well as to an especially salient pair of conditions known as *Jeffrey-independence conditions*, namely,

$$q'(E_i) = p(E_i), \qquad \text{for all } i, \text{ and} \tag{3.15}$$

$$q(F_j) = p(F_j), \qquad \text{for all } j. \tag{3.16}$$

Theorem 3.2. *If in* (3.1) *it is the case that* $\big(r'(E_i)\big) = \big(q(E_i)\big)$*, then the conditions* (3.2)*,* (3.13)*, and* (3.15) *are equivalent. If it is the case that* $\big(q'(F_j)\big) = \big(r(F_j)\big)$*, then the conditions* (3.3)*,* (3.14)*, and* (3.16) *are equivalent.*

Proof. Straightforward algebraic verification. □

Remark 3.4. Diaconis and Zabell (1982), who coined the term "Jeffrey independence," proved in the context of (2.7), with $\mathbf{E}$ and $\mathbf{F}$ finite, that conditions (3.15) and (3.16) imply that $r = r'$. In view of Theorem 3.2, this result is also a corollary of Theorem 3.1. Note that the p-independence of $\mathbf{E}$ and $\mathbf{F}$ (i. e., $p(E_iF_j) = p(E_i)p(F_j)$, for all i, j) entails Jeffrey independence.

Remark 3.5. Formula (3.7), and consequently Theorem 3.1, can be generalized to arbitrary finite sequences of probability-kinematical revisions. We leave the details as an exercise for interested readers.

4 A Partial Converse of Field's Theorem

The Bayes factor identities (3.2) and (3.3) are not in general necessary for r' and r in the revision schema (3.1) to coincide. For example, if $\mathbf{F} = \mathbf{E} = \{E_i\}$ and $r(E_i) = r'(E_i)$ for all i, then $r' = r$, no matter what values are assigned to $q(E_i)$ and $q'(E_i)$. However, Field's Theorem does admit of a partial converse.

To motivate the statement of conditions under which (3.2) and (3.3) are necessary for commutativity, it is useful to reiterate the conditions under which the probabilities in schema (3.1) are well-defined. In order to implement the formulas for q, q', r, and r', we must have, respectively, $p(E_i) > 0$, $p(F_j) > 0$, $q(F_j) > 0$, and $q'(E_i) > 0$, for all i and j, or, equivalently, that

$$\forall i\ \exists j\ :\ p(E_iF_j) > 0, \qquad \text{and} \tag{4.1}$$

$$\forall j\ \exists i\ :\ p(E_iF_j) > 0. \tag{4.2}$$ [5]

The identities (3.2) and (3.3) turn out to be necessary for commutativity under a mild strengthening of the well-definedness conditions (4.1) and (4.2).

Theorem 4.1. *Let r and r' be defined by the probability revision schema* (3.1)*, and suppose that*

$$\forall i_1 \ \forall i_2 \ \exists j \ : \ p(E_{i_1}F_j)p(E_{i_2}F_j) > 0, \quad \textit{and} \tag{4.3}$$

$$\forall j_1 \ \forall j_2 \ \exists i \ : \ p(E_iF_{j_1})p(E_iF_{j_2}) > 0. \tag{4.4}$$

If $r' = r$, then the Bayes factor identities (3.2) *and* (3.3) *hold.*

Proof. Since q comes from p by probability kinematics on $\mathbf{E}$, the rigidity condition (2.3) implies that

$$q(E_iF_j) = q(E_i)p(F_j|E_i), \qquad \text{for all } i, j. \tag{4.5}$$

Similarly,

$$q(E_iF_j) = q(F_j)r(E_i|F_j), \qquad \text{for all } i, j, \tag{4.6}$$

$$q'(E_iF_j) = q'(F_j)p(E_i|F_j), \qquad \text{for all } i, j, \text{ and} \tag{4.7}$$

$$q'(E_iF_j) = q'(E_i)r'(F_j|E_i), \qquad \text{for all } i, j, \tag{4.8}$$

and so

$$q(E_i)p(F_j|E_i) = q(F_j)r(E_i|F_j), \qquad \text{for all } i, j, \text{ and} \tag{4.9}$$

$$q'(E_i)r'(F_j|E_i) = q'(F_j)p(E_i|F_j), \qquad \text{for all } i, j. \tag{4.10}$$

Given arbitrary i_1 and i_2, let j be such that (4.3) holds. It then follows *inter alia* from the relevant rigidity conditions that $r(E_{i_2}F_j) > 0$ and $r'(E_{i_2}F_j) > 0$. Setting $i = i_1, i_2$ in (4.9) and solving for $q(E_{i_1})$ and $q(E_{i_2})$ then yields the formula

$$\beta_{q,p}(E_{i_1} : E_{i_2}) = \frac{p(E_{i_2}F_j)}{p(E_{i_1}F_j)} \frac{r(E_{i_1}F_j)}{r(E_{i_2}F_j)}. \tag{4.11}$$

Similarly, (4.10) yields the formula

$$\beta_{r',q'}(E_{i_1} : E_{i_2}) = \frac{p(E_{i_2}F_j)}{p(E_{i_1}F_j)} \frac{r'(E_{i_1}F_j)}{r'(E_{i_2}F_j)}. \tag{4.12}$$

Hence if $r' = r$, then the Bayes factor identity (3.2) holds.

Given arbitrary j_1 and j_2, let i be such that (4.4) holds. It then follows from (4.9) and (4.10) by an argument similar to the above that

$$\beta_{q',p}(F_{j_1} : F_{j_2}) = \frac{p(E_i F_{j_2})}{p(E_i F_{j_1})} \frac{r'(E_i F_{j_1})}{r'(E_i F_{j_2})}, \quad \text{and} \tag{4.13}$$

$$\beta_{r,q}(F_{j_1} : F_{j_2}) = \frac{p(E_i F_{j_2})}{p(E_i F_{j_1})} \frac{r(E_i F_{j_1})}{r(E_i F_{j_2})}. \tag{4.14}$$

Hence if $r' = r$, then the Bayes factor identity (3.3) holds. □

Remark 4.1. Note that when $\mathbf{F} = \mathbf{E}$, conditions (4.3) and (4.4) fail to hold, thus allowing the possibility (illustrated in the example cited at the beginning of this section) that $r' = r$ even though (3.2) and (3.3) fail to hold. On the other hand, (4.3) and (4.4) always hold when $\mathbf{E}$ and $\mathbf{F}$ are qualitatively independent (for all i, j, $E_i E_j \neq \varnothing$) and p is strictly coherent ($p(A) > 0$ for all nonempty events A).

Remark 4.2. As noted in Remark 3.4, if in the probability revision schema (3.1), $\big(r'(E_i)\big) = \big(q(E_i)\big)$ and $\big(q'(F_j)\big) = \big(r(F_j)\big)$, then the Jeffrey independence conditions (3.15) and (3.16) imply that $r' = r$. Interestingly, given (4.3) and (4.4), if for some i, $r'(E_i) \neq q(E_i)$, or for some j, $q'(F_j) \neq r(F_j)$, then Jeffrey independence not only fails to ensure that $r' = r$, but actually ensures that $r' \neq r$. This follows from Theorem 4.1 and the easily verified fact that (3.15) and (3.16), along with (3.2) and (3.3), imply that $\big(r'(E_i)\big) = \big(q(E_i)\big)$ and $\big(q'(F_j)\big) = \big(r(F_j)\big)$. In particular, the p-independence of $\mathbf{E}$ and $\mathbf{F}$, since it implies both (4.3) and (4.4) and (as noted in Remark 3.4 above) Jeffrey independence, also ensures that $r' \neq r$ unless $\big(r'(E_i)\big) = \big(q(E_i)\big)$ and $\big(q'(F_j)\big) = \big(r(F_j)\big)$.

Remark 4.3. Diaconis and Zabell (1982) proved for $\mathbf{E}$ and $\mathbf{F}$ finite that in the special case of (3.1) represented by (2.7) Jeffrey independence is necessary for $r' = r$. In view of Theorem 3.2, this result is a corollary of Theorem 4.1.

5 Sequential Probability Kinematics: All is Cool

That r' may fail to coincide with r in the probability-kinematical revision schema (2.7) has been cause for concern among several commentators (see

Lange 2000 for some sample quotations). Their concern appears to be based on implicit acceptance of two principles relating to the general revision schema (3.1), reproduced below:

$$\begin{array}{ccc} p & \xrightarrow{\mathbf{E}} & q \\ {\scriptstyle \mathbf{F}}\Big\downarrow & & \Big\downarrow{\scriptstyle \mathbf{F}} \\ & & r \\ q' & \xrightarrow[\mathbf{E}]{} & r' \end{array} .$$

I. If what is learned from the experience prompting the revisions of p to q, and of q' to r', is the same, and if what is learned from the experience prompting the revisions of q to r, and of p to q', is the same, then it ought to be the case that $r' = r$.

II. Identical learning underlying the revisions of p to q and of q' to r' ought to be reflected in the posterior probability identities

$$r'(E_i) = q(E_i), \qquad \text{for all } i, \tag{5.1}$$

and identical learning underlying the revisions of q to r and of p to q' in the identities

$$q'(F_j) = r(F_j), \qquad \text{for all } j. \tag{5.2}$$

The first of these principles is unexceptionable. To paraphrase van Fraassen (1989), two persons who undergo identical learning experiences on the same day, but in a different order, ought to agree in the evening if they had exactly the same opinions in the morning. But the second principle is mistaken, losing sight of the fact that posterior probabilities assigned to events in the families **E** and **F** are based on the total evidence, old as well as new, and thus incorporate elements of the relevant priors.[6]

What we need is a numerical representation of what is learned from new evidence alone, with prior probabilities factored out. It is a staple of Bayesianism that ratios of new-to-old odds furnish the correct representation of the desired type (Good, 1950, 1983; Jeffrey 2000).[7] Accordingly, Principle **II** needs to be modified by replacing (5.1) and (5.2), respectively, by the Bayes factor identities (3.2) and (3.3). So modified, Principle **II** is both sufficient (Theorem 3.1), and in a substantial number of cases, necessary (Theorem 4.1) for the satisfaction of Principle **I**.

Remark 5.1. Field (1978) came close to our modification of Principle **II**, but made sensory stimulus, *eo ipso*, the source of Bayes factors, with identical stimuli prompting probability revisions by identical Bayes factors. A counterexample of Garber (1980) shows this physicalist view to be untenable: Imagine that you glance briefly in dim light at an object known to be blue or green, resulting in your becoming slightly more confident than you were before that the object is blue. Repeated glances producing the identical sensory stimulus will then result in your approaching certainty that the object is blue. Our formulation, in which identical *learning* gives rise to identical Bayes factors, is immune to Garber's counterexample. We learn nothing new from repeated glances and so all Bayes factors beyond the first are equal to one.[8]

Remark 5.2. If $\mathbf{F} = \mathbf{E}$ in the probability revision schema (2.7), then $r' = q$ and $r = q'$, and so it is always the case that $r' \neq r$, except in the uninteresting case in which $q' = q$. Lange (2000) has furnished a lucid example suggesting that this never involves a violation of Principle $\underline{\text{I}}$ since the relevent revisions are not based on identical learning.[9] Note that this claim follows from the principle that identical learning ought to be reflected in identical Bayes factors: Since $q' \neq q$, either $q' \neq p$ or $q \neq p$. In the former case, the identity $\beta_{r',q'}(E_{i_1} : E_{i_2}) = \beta_{q,p}(E_{i_1} : E_{i_2})$ can not hold for all i_1, i_2, since $r' = q$. In the latter, the identity $\beta_{r,q}(E_{i_1} : E_{i_2}) = \beta_{q',p}(E_{i_1} : E_{i_2})$ cannot hold for all i_1, i_2, since $r = q'$.

Remark 5.3. Note that in Theorem 3.1 the probabilities p, q, q', r, and r' are assumed to be well-defined and in place at the outset. Then $r' = r$ if the Bayes factor identities (3.2) and (3.3) hold. Suppose that only p, q, and q' were in place at the outset. Does (3.2) then furnish a recipe for *constructing* a probability r' that would be the appropriate revision of q' if in the probabilistic state q' one were to learn precisely what prompted the revision of p to q? And does (3.3) function analogously in the construction of a probability r? Only if, in the first instance,

$$(5.3) \qquad \sum_i q(E_i)\, q'(E_i)/\, p(E_i) < \infty$$

and in the second,

$$(5.4) \qquad \sum_j q(F_j)\, q'(F_j)/\, p(F_j) < \infty,$$ [10]

since a probability r' satisfies (3.2) if and only if

$$r'(E_i) = \frac{q(E_i)q'(E_i)}{p(E_i)} \Big/ \sum_i \frac{q(E_i)q'(E_i)}{p(E_i)} \tag{5.5}$$

and a probability r satisfies (3.3) if and only if

$$r(F_j) = \frac{q(F_j)q'(F_j)}{p(F_j)} \Big/ \sum_j \frac{q(F_j)q'(F_j)}{p(F_j)} \, . \tag{5.6}$$

But this raises an intriguing question, with which we conclude this paper. If (5.3), and hence (5.4), fails to hold, as may be the case,[11] does this mean that the learning prompting the revision of p to q (respectively, p to q') cannot identically occur in the probabilistic state q' (respectively, q)?

Notes

1. In the standard exposition of probability kinematics the family $\mathbf{E} = \{E_i\}$ is taken to be a *partition* of Ω, so that, in addition to pairwise disjointness of the events E_i, one has $E_1 \cup E_2 \cup \cdots = \Omega$. Standardly, (e_i) is a sequence of *non-negative* real numbers summing to one, and it is assumed that "zeros are not raised", i. e., that $p(E_i) = 0$ implies that $e_i\big(= q(E_i)\big) = 0$. Finally, it is stipulated that $0 \cdot p(A|E_i) = 0$ if $p(E_i) = 0$, so that $e_i p(A|E_i)$ is well-defined even when $p(A|E_i)$ isn't. Given the standard format, our family $\mathbf{E}$ simply comprises those E_i in the partition for which $e_i > 0$. Conversely, our format yields the standard one by associating to our family $\mathbf{E} = \{E_1, E_2, \ldots\}$, if it fails to be a partition, the partition $\{E_0, E_1, E_2, \ldots\}$, where $E_0 := \Omega - (E_1 \cup E_2 \cup \ldots)$, and setting $e_0 = 0$. When one deals with sequential probability-kinematical revisions in the standard format, conventions involving the values of expressions involving benign sorts of undefinedness multiply, and can easily obscure non-benign cases of undefinedness (see, especially, §4 below, where certain positivity conditions are crucial). Our format minimizes the possibility of such confusion.

Probability kinematics may arise through ordinary conditionalization. Suppose, for example, that $\mathbf{A}$ is the σ-algebra generated by $\mathbf{E} = \{E_i\}$ along with arbitrary hypotheses $H_1, H_2, \ldots$, and an event $E \subset \bigcup E_i$ such that $p(EE_i) > 0$ for all i. Let $\mathbf{A}'$ be the σ-subalgebra of $\mathbf{A}$ generated by $\mathbf{E}$ and the H_i. If q comes from p by conditioning on E and for all $A \in \mathbf{A}'$ and for all i, A and E are conditionally p-independent, given E_i ($p(A|EE_i) = p(A|E_i)$), then on $\mathbf{A}'$, q comes from p by probability kinematics on $\mathbf{E}$. The whole point of probability kinematics is of course that the experience prompting revision of the probabilities of events in the family $\mathbf{E}$ often fails to be representable as the occurrence of an event E. But it is sometimes useful, when attempting to explore certain aspects of probability kinematics, to entertain such a fictional "phenomenological event" E. At least when $\mathbf{A}$ is finite, this is always a formal possibility (see Diaconis and Zabell 1982, Theorem 2.1, for the whole story).

2. This is an exceedingly important point. As will be seen in §5 below, the mistaken apprehension that the probabilities $q(E_i)$ are based solely on new evidence is responsible for much of the confusion surrounding the "noncommutativity" of probability kinematics.

3. Actually, Field expresses G_i in the form e^{α_i}, where $\alpha_i := \log G_i$, interpreting the α_i as "input parameters" associated with a given sensory stimulus. Each instance of this stimulus prompts a probability revision of the type (3.9) involving these input parameters. See Remark 5.1 in §5 for further discussion of this idea.

4. As an illustration, let us show the equivalence of (3.11) and (3.2). The proportionality (3.1) asserts the existence of a positive constant c such that, for all i, $\pi_{r',q'}(E_i) = c\pi_{q,p}(E_i)$. Then (3.2) follows immediately from (1.3). Conversely, it follows from (3.2) with $i_1 = i$ and $i_2 = 1$, along with (1.3), that, for all i, $\pi_{r',q'}(E_i) = c\pi_{q,p}(E_i)$, where $c = \pi_{r',q'}(E_1)/\pi_{q,p}(E_1)$.

5. From (4.1) it follows that $p(E_i) > 0$ for all i, and from (4.2) that $p(F_j) > 0$ for all j. Since $q(F_j) = \sum_i e_i p(F_j|E_i)$, with all $e_i > 0$, it follows from (4.2) that $q(F_j) > 0$ for all j. Since $q'(E_i) = \sum_j f_j p(E_i|F_j)$, with all $f_j > 0$, it follows from (4.1) that $q'(E_i) > 0$ for all i. If, on the other hand, (4.2) fails to hold, then by the above formula for $q(F_j)$, there exists a j such that $q(F_j) = 0$. And if (4.1) fails to hold, then by the above formula for $q'(E_i)$, there exists an i such that $q'(E_i) = 0$.

6. Suppose, for example, that a ball is chosen at random from an urn containing 9999 green balls and one blue ball, and you get to examine it fleetingly in a dim light. It would be folly to assess the probability that the ball is blue based only on your sensory impression, ignoring the composition of the urn. Indeed, unless you have solid grounds for regarding the sensory impression that your glance produces as being much more likely if the ball is blue than if it is green, your prior probability should undergo little, if any, revision.

7. Notice that several familiar measures of probability change lack the requisite feature of effacing all traces of the prior. For example, knowing nothing about p and q except, say, that $q(E) - p(E) = 1/4$, one can conclude that $p(E) \leqslant 3/4$, and knowing nothing except, say, that $q(E)/p(E) = 2$, one can conclude that $p(E) \leqslant 1/2$. Suppose, on the other hand, that $\mathbf{E} = \{E_1, \ldots, E_m\}$ and nothing is known about p and q except that $\beta_{p,q}(E_i : E_1) = \beta_i$, where (β_i) is a given sequence of positive real numbers with $\beta_1 = 1$. From this information nothing whatsoever can be inferred about p. For given *any* such (β_i) and *any* prior p with $p(E_i) > 0$ for all i, there exists a probability q such that $\beta_{q,p}(E_i : E_1) = \beta_i$, namely, $q(E_i) = \beta_i p(E_i) / \sum_i \beta_i p(E_i)$.

8. A way to see this is to entertain an imaginary phenomenological event E capturing the visual content of the glance (see note 1, *supra*). Then if $q(\cdot) = p(\cdot|E)$, and $r(\cdot) = q(\cdot|E)$, $\beta_{q,p}\big(B(\text{lue}) : G(\text{reen})\big) = p(E|B)/p(E|G) > 1$, by assumption. But $\beta_{r,q}(B : G) = q(E|B)/q(E|G) = p(E|EB)/p(E|EG) = 1/1 = 1$.

9. Lange speaks of experience rather than learning, but we mean the same thing by these terms, namely, *considered* experience (in light of ambient memory and prior probabilistic commitment), rather than the isolated sensory experience that Field saw as the source of Bayes factors. We have adopted our terminology from Jeffrey 2000 to minimize confusion with the latter sort of experience.

10. Actually, (5.3) and (5.4) are equivalent, since the sums in question are each equal to

$$\sum_{i,j} \frac{q(E_i)}{p(E_i)} \frac{q'(F_j)}{p(F_j)} p(E_i F_j).$$

11. Let $\Omega = \{1, 2, \ldots\}$, $E_i = \{2i-1, 2i\}$, and $F_j = \{4j-3, 4j-2, 4j-1, 4j\}$ for $i, j = 1, 2, \ldots$. Let $p(2i-1) = p(2i) = 7/2 \cdot 8^i$, let q come from p by probability kinematics on $\{E_i\}$ with $q(E_i) = 1/2^i$, and let q' come from p by probability kinematics on $\{F_j\}$ with $q'(F_j) = 2/3^j$. Then, since $F_j = E_{2j-1} \dot\cup E_{2j}$,

$$\sum_j \frac{q(F_j) q'(F_j)}{p(F_j)} = \sum_j \frac{2}{21} \left(\frac{16}{3}\right)^j = \infty.$$

References

[1] Diaconis, Persi and Sandy L. Zabell (1982), "Updating Subjective Probability", *Journal of the American Statistical Association* 77(380): 822–830.

[2] Domotor, Zoltan (1980), "Probability Kinematics and Representation of Belief Change", *Philosophy of Science* 47: 384–403.

[3] Döring, Frank (1999), "Why Bayesian Psychology is Incomplete", *Philosophy of Science* 66: S379–S389.

[4] Field, Hartry, (1978), "A Note on Jeffrey Conditionalization", *Philosophy of Science* 45: 361–367.

[5] Garber, Daniel, (1980), "Field and Jeffrey Conditionalization", *Philosophy of Science* 47: 142–145.

[6] Good, Irving J. (1950), *Probability and the Weighing of Evidence.* London: Charles Griffin.

[7] Good, Irving J. (1983), *Good Thinking.* Minneapolis: University of Minnesota Press.

[8] Jeffrey, Richard C. (1965), *The Logic of Decision.* New York: McGraw-Hill; (1983), 2nd ed. Chicago: University of Chicago Press.

[9] Jeffrey, Richard C. (1988), "Conditioning, Kinematics, and Exchangeability", in Brian Skyrms and William L. Harper (eds.), *Causation, Chance, and Credence*, vol. 1. Dordrecht: Kluwer, 221–255.

[10] Jeffrey, Richard C. (2000), "Epistemology Probabilized", manuscript, available at `http://www.princeton.edu/~bayesway`.

[11] Lange, Marc (2000), "Is Jeffrey Conditionalization Defective in Virtue of Being Non-Commutative? Remarks on the Sameness of Sensory Experience", *Synthese* 123: 393–403.

[12] Skyrms, Brian (1986), *Choice and Chance*, 3rd ed. Belmont: Wadsworth Publishing Co.

[13] van Fraassen, Bas (1989), *Laws and Symmetry.* Oxford: Clarendon Press.