Jeffrey Conditioning and External Bayesianity

Carl Wagner
Department of Mathematics
The University of Tennessee
Knoxville, TN 37996-1300, USA
wagner@math.utk.edu

Abstract. Suppose that several individuals who have separately assessed prior probability distributions over a set of possible states of the world wish to pool their individual distributions into a single group distribution, while taking into account jointly perceived new evidence. They have the option of (i) first updating their individual priors and then pooling the resulting posteriors or (ii) first pooling their priors and then updating the resulting group prior. If the pooling method that they employ is such that they arrive at the same final distribution in both cases, the method is said to be *externally Bayesian*, a property first studied by Madansky (1964). We show that a pooling method for discrete distributions is externally Bayesian if and only if it commutes with Jeffrey conditioning, properly parameterized. This result provides further evidence that identical new learning should be represented by identical ratios of certain new to old odds.

1. Combining Probability Distributions.

In what follows, Ω denotes a countable set of possible states of the world, assumed to be mutually exclusive and exhaustive. A function $p: \Omega \rightarrow [0,1]$ is a *probability mass function* (pmf) iff $\sum_{\omega \in \Omega} p(\omega) = 1$. The *support* of a pmf p is the set $\mathrm{Supp}(p) := \{ \omega : p(\omega) > 0 \}$. Each pmf p gives rise to a *probability measure* (which, abusing notation, we also denote by p) defined for each set $E \subseteq \Omega$ by $p(E) := \sum_{\omega \in E} p(\omega)$.

Denote by Δ the set of all pmfs on Ω, and let n be a positive integer. Let Δ^n denote the n-fold Cartesian product of Δ, with

$$\Delta^{n+} := \{ (p_1,\ldots,p_n) \in \Delta^n : \ \cap_i \mathrm{Supp}(p_i) \neq \varnothing \} .$$

A *pooling operator* is any function $T: \Delta^{n+} \rightarrow \Delta$.[1] Given $(p_1,\ldots,p_n) \in \Delta^{n+}$, the pmf $T(p_1,\ldots,p_n)$ may, depending on the context, represent

(i) a rough summary of the current pmfs $p_1,\ldots,p_n$ of n individuals;

(ii) a compromise adopted by these individuals in order to complete an exercise in group decision making;

(iii) a "rational" consensus to which all individuals have revised their initial pmfs $p_1,\ldots,p_n$ after extensive discussion;

(iv) the pmf of a decision maker external to a group of n experts (who may or may not have assessed his own prior over Ω before consulting the group) upon being apprised of the pmfs $p_1,\ldots,p_n$ of these experts;

(v) a revision of the pmf p_i of a particular individual i upon being apprised of the pmfs of individuals $1,\ldots,i\text{-}1,i+1,\ldots,n$, each of whom he may or may not consider to be his "epistemic peer."

The appropriate restrictions to place on the pooling operator T will naturally depend on the interpretation of $T(p_1,\ldots,p_n)$. For example, it might seem reasonable for T to *preserve unanimity* ($T(p,\ldots,p) = p$) [2] in cases (i),(ii), and (v) above, but perhaps not in cases (iii) and (iv). There is an extensive literature on this general subject (see, for example, the article of Genest and Zidek (1986) for a summary and appraisal of work done through the mid-1980s). There has also been a recent surge of interest in problems associated with interpretation (v) above, as part of a field of inquiry that has come to be termed the "epistemology of disagreement."

Our interest here is not in adjudicating which restrictions on pooling are appropriate in which situations, but rather in a formal analysis of one proposed group rationality condition, known as *external Bayesianity*, and its connections with Jeffrey conditioning.(We employ in what follows the language of interpretations (ii) and (iii) above, since it is in those cases where external Bayesianity seems most compelling. But our formal results apply to the other interpretations as well, for whatever interest that may have.) We shall see that a pooling operator for pmfs is externally Bayesian if and only if it commutes with Jeffrey conditioning, properly parameterized. This result provides further evidence for the claim, advanced in Wagner (2002, 2003), that identical new learning should be reflected in identical ratios of certain new to old odds.

2. Externally Bayesian Pooling Operators.

Consider the situation in which n individuals who have assessed pmfs $p_1,\ldots,p_n$ over Ω subsequently undergo identical new learning as a result of jointly perceived new evidence. Should they first update their individual priors based on the jointly perceived new information, and then pool the posteriors? Or should they pool their priors, and then update the result of pooling based on this information?

Take the simplest case, where each individual comes to learn that the true state of the world belongs to the subset E of Ω, but nothing that would change the odds between any states ω_1 and ω_2 in E. It would thus be appropriate for each individual i to revise his p_i to, let us call it q_i, by conditioning on E,[3] so that for each $\omega \in \Omega$,

(2.1) $q_i(\omega) = p_i(\omega|E) := p_i(\omega)\,[\omega \in E] / p_i(E)$,

where $[\omega \in E]$ denotes the characteristic function[4] of the set E, evaluated at ω . These revised pmfs might then be combined by means of the pooling operator T. Alternatively, one might imagine first pooling the priors $p_1,\ldots,p_n$ and then conditioning the result on E. If either of these procedures results in the same final distribution, we say that T *commutes with conditioning* (CC). This property may be expressed formally as follows:

CC : For all subsets E of Ω and all $(p_1,\ldots,p_n) \in \Delta^{n+}$ such that $p_i(E) > 0$, i= 1,…,n, and $(p_1(.|E),\ldots,p_n(.|E)) \in \Delta^{n+}$, it is the case that

(2.1) $T(p_1,\ldots,p_n)(E) > 0$,

and

(2.2) $T(p_1(.|E),\ldots,p_n(.|E)) = T(p_1,\ldots,p_n)(.|E)$.

We now wish to generalize condition CC. Given $(p_1,\ldots,p_n) \in \Delta^{n+}$, we say that a function $\lambda : \Omega \to [0, \infty)$ is a *likelihood* for $(p_1,\ldots,p_n)$ iff

$$(2.3) \qquad 0 < \sum_{\omega \in \Omega} \lambda(\omega)\, p_i(\omega) < \infty \ , \ i = 1,\ldots,n,$$

and $(q_1,\ldots,q_n) \in \Delta^{n+}$, where

$$(2.4) \qquad q_i(\omega) := \lambda(\omega)p_i(\omega) / \sum_{\omega \in \Omega} \lambda(\omega)\, p_i(\omega) .$$

In a slight modification of Madansky (1964, 1978), we say that $T : \Delta^{n+} \to \Delta$ is *externally Bayesian* (EB) iff the following condition is satisfied:

EB : If $(p_1,\ldots,p_n) \in \Delta^{n+}$ and λ is a likelihood for $(p_1,\ldots,p_n)$, then

$$(2.5) \qquad 0 < \sum_{\omega \in \Omega} \lambda(\omega)\, T(p_1,\ldots,p_n)(\omega) < \infty,$$

and the following commutativity property holds:

$$(2.6) \quad T\left(\lambda p_1 / \sum_{\omega \in \Omega} \lambda(\omega)\, p_1(\omega), \ldots , \lambda p_n / \sum_{\omega \in \Omega} \lambda(\omega)\, p_n(\omega)\right)$$

$$= \lambda\, T(p_1,\ldots,p_n) / \sum_{\omega \in \Omega} \lambda(\omega)\, T(p_1,\ldots,p_n)(\omega).$$

The term *externally Bayesian* derives from the fact that a group of decision makers having a common utility function but different priors over the relevant states of nature and employing an EB pooling operator will make decisions that appear to an outsider like the decisions of a Bayesian (see Madansky 1964). An example due to Raiffa (1968, pp. 221-6) shows that the use of pooling operators that fail to satisfy EB may lead members of the group to act in strange ways.

The set of externally Bayesian pooling operators is clearly nonempty since "dictatorial" pooling (for fixed i, $T(p_1,\ldots,p_n) = p_i$ for all $(p_1,\ldots,p_n) \in \Delta^{n+}$) satisfies EB. Note, however, that weighted arithmetic means, i.e., pooling operators T defined by

$$(2.7) \qquad T(p_1,\ldots,p_n) := \sum_{1 \le i \le n} w(i)p_i(\omega),$$

where w(1), … ,w(n) is a sequence of nonnegative real numbers summing to one, fail in general to satisfy EB. On the other hand, suppose that we define a pooling operator T by

(2.8) $T(p_1,\ldots,p_n)(\omega) := \prod_{1\le i\le n} p_i(\omega)^{w(i)} / \sum_{\omega\in\Omega} \prod_{1\le i\le n} p_i(\omega)^{w(i)}$,[5]

where $0^0 := 1$. It is easy to verify that such normalized weighted *geometric* means, which have come to be termed *logarithmic pooling operators*, are externally Bayesian, a fact first noted (according to Bacharach (1972)) by Peter Hammond. See Genest, McConway, and Schervish (1986) for a thorough discussion and characterization of externally Bayesian pooling operators for arbitrary, not necessarily discrete, density functions.

Remark 2.1. Setting $\lambda(\omega) = [\omega \in E]$ shows, as suggested above, that EB implies CC.

3. Jeffrey Conditioning and External Bayesianity.

Let p and q be pmfs on the countable set Ω and let $\mathbf{E} = \{E_k\}$ be a set of nonempty, pairwise disjoint subsets of Ω such that $p(E_k) > 0$ for all k. We say that q comes from p by *Jeffrey conditioning* on **E** iff there exists a sequence (e_k) of positive real numbers summing to one such that, for every ω in Ω,

(3.1) $q(\omega) = \sum_k e_k\, p(\omega|E_k) = \sum_k e_k\, p(\omega)\,[\omega \in E_k] / p(E_k)$. [6]

Formula (3.1) is the appropriate way to update your prior p in light of new evidence if and only if (1) based on the total evidence, old as well as new, you judge that for each k, the posterior probability $q(E_k)$ should take the value e_k ; and (2) for each E_k, you judge that nothing in the new evidence should disturb the odds between any two states of the world in E_k.[7] Note that Jeffrey conditioning reduces to ordinary conditioning when $\mathbf{E} = \{E\}$. Note also that if p and q are *any* pmfs on the countable set Ω and $\mathrm{Supp}(q) \subseteq \mathrm{Supp}(p)$, then q comes from p by Jeffrey conditioning on the family $\mathbf{E} = \{ \{\omega\} : q(\omega) > 0 \}$.

One would hope that, just as externally Bayesian pooling commutes with ordinary conditioning, such pooling would also commute with

Jeffrey conditioning. Suppose then that n individuals who have assessed priors $p_1,\ldots,p_n$ over Ω subsequently undergo a common experience that prompts each individual i to revise his prior p_i to q_i by Jeffrey conditioning on the family $\mathbf{E} = \{E_k\}$, with $q_i(E_k) = e_k$ for every i. We say that a pooling operator T commutes with Jeffrey conditioning of this sort (CJC_1) iff the following condition holds:

CJC_1 : For all families $\mathbf{E} = \{E_k\}$ of nonempty, pairwise disjoint subsets of Ω, all $(p_1,\ldots,p_n) \in \Delta^{n+}$ such that $p_i(E_k) > 0$ for all i and all k, and all sequences (e_k) of positive real numbers summing to one such that $(q_1,\ldots,q_n) \in \Delta^{n+}$, where

(3.2) $q_i(\omega) : = \sum_k e_k\, p_i(\omega|E_k)$,

it is the case that

(3.3) $T(p_1,\ldots,p_n)(E_k) > 0$ for all k,

and

(3.4) $T(\sum_k e_k\, p_1(\cdot|E_k) ,\ldots, \sum_k e_k\, p_n(\cdot|E_k)) = \sum_k e_k\, T(p_1,\ldots,p_n)(\cdot| E_k)$.

The following simple example shows that EB does not imply CJC_1 : Let $\Omega = \{ \omega_1, \omega_2, \omega_3, \omega_4 \}$, $(p_1(\omega_1),\ldots, p_1(\omega_4)) = (1/7, 4/7, 1/7, 1/7)$, $(p_2(\omega_1),\ldots, p_2(\omega_4)) = (4/13, 1/13, 4/13, 4/13)$, $E_1 = \{ \omega_1, \omega_2 \}$, $E_2 = \{ \omega_3, \omega_4 \}$, $\mathbf{E} = \{ E_1, E_2 \}$, and $e_1 = e_2 = ½$. Let $T(p_1, p_2)(\omega)$ be given by (2.5) above, with n =2 and w(1) = w(2) = ½. As noted earlier, T is externally Bayesian. If q_1 and q_2 denote, respectively, the results of Jeffrey conditioning p_1 and p_2 on $\mathbf{E}$, and $q = T(q_1, q_2)$, then $(q(\omega_1) ,\ldots, q(\omega_4)) = (2/9, 2/9, 5/18, 5/18)$. If r denotes the result of Jeffrey conditioning $T(p_1, p_2)$ on $\mathbf{E}$, then $(r(\omega_1) ,\ldots, r(\omega_4)) =$ $(1/4, 1/4, 1/4, 1/4)$.

However, this example is neither disturbing nor surprising. After all, the property of external Bayesianity is designed to guarantee that a pooling operator with that property will commute with updating based on *identical new learning*. This is hardly the case in the above example, where additional information has prompted the first individual to lower his odds on E_1 against E_2 from 5:2 to 1:1 and the second individual to raise those odds from 5:8 to 1:1.[8] Recall that the posterior probabilities $q_1(E_1) = q_2(E_1)$ and $q_1(E_2) = q_2(E_2)$ are based on the total evidence, old as well as new. What we need as a marker of identical new learning is a numerical representation of what is learned from new evidence alone, with prior probabilities somehow factored out. It is a staple of Bayesianism that ratios of new to old odds furnish the correct such representation (Good 1950, 1983), a view borne out by much subsequent work.[9]

The following terminology and notation will be useful in elaborating the above: If q is a revision of the probability measure p and A and B are events, the *Bayes factor* $\beta_{q,p}(A : B)$ is the ratio

(3.5) $$\beta_{q,p}(A : B) := (q(A)/q(B)) / (p(A)/p(B))$$

of new to old odds, and the *relevance quotient* $\rho_{q,p}(A)$ is the ratio

(3.6) $$\rho_{q,p}(A) := q(A)/p(A)$$

of new to old probabilities. When $q = p(\mathbf{.}|E)$, then (3.5) is simply the *likelihood ratio* $p(E|A)/p(E|B)$. When $A = H$ and $B = H^c$, this likelihood ratio (or its logarithm) is arguably the best measure of the degree to which H is (incrementally) confirmed by E (Fitelson 1999, p. 363, footnote 2). More generally,

(3.7) $$\beta_{q,p}(A : B) = \rho_{q,p}(A) / \rho_{q,p}(B) ,$$

a simple, but useful, identity.

Theorem 3.1. Suppose that q comes from p by Jeffrey conditioning on the family $\mathbf{E} = \{E_k\}$ and let $b_k := \beta_{q,p}(E_k : E_1)$, $k = 1,2,\ldots$. Then, for all ω in Ω,

(3.8) $$q(\omega) = \sum_k b_k \, p(\omega) [\omega \in E_k] \;/\; \sum_k b_k \, p(E_k) .$$

Proof. Dividing (3.1) by $1 = q(\Omega) = \sum_k p(E_k) q(E_k)/p(E_k)$, replacing e_k by $q(E_k)$, dividing the numerator and denominator of the resulting fraction by $q(E_1)/ p(E_1)$, and invoking (3.7) yields (3.8). □

The above parameterization of Jeffrey conditioning allows us to formulate a commutativity property with the proper relation to external

Bayesianity. We say that a pooling operator T commutes with Jeffrey conditioning, so parameterized, (CJC_2) iff the following condition holds:

CJC_2 : For all families $\mathbf{E} = \{E_k\}$ of nonempty, pairwise disjoint subsets of Ω, all $(p_1,\ldots,p_n) \in \Delta^{n+}$ such that $p_i(E_k) > 0$ for all i and all k, and all sequences (b_k) of positive real numbers such that $b_1 = 1$ and

(3.9) $\sum_k b_k\, p_i(E_k) < \infty$, $i= 1,\ldots,n$,

and such that $(q_1,\ldots,q_n) \in \Delta^{n+}$, where

(3.10) $q_i(\omega) := \sum_k b_k\, p_i(\omega)[\omega \in E_k] \;/\; \sum_k b_k\, p_i(E_k)$, [10]

it is the case that

(3.11) $0 < \sum_k b_k\, T(p_1,\ldots, p_n)(E_k) < \infty$,

and

(3.12) $T(\sum_k b_k\, p_1[\,\centerdot \in E_k] / \sum_k b_k\, p_1(E_k), \ldots , \sum_k b_k\, p_n[\,\centerdot \in E_k] / \sum_k b_k\, p_n(E_k))$

$= \sum_k b_k\, T(p_1,\ldots, p_n)[\,\centerdot \in E_k] \;/\; \sum_k b_k\, T(p_1,\ldots, p_n)(E_k)$. [11]

Theorem 3.2. A pooling operator $T\colon \Delta^{n+} \to \Delta$ is externally Bayesian if and only if it commutes with Jeffrey conditioning in the sense of CJC_2 .

Proof. *Necessity*. Let T be externally Bayesian. Suppose that (3.9) holds and that $(q_1,\ldots,q_n) \in \Delta^{n+}$, where q_i is given by (3.10)
Let $\lambda(\omega) = \sum_k b_k [\omega \in E_k]$. Then, for $i = 1,\ldots,n$,

(3.13) $$\sum_{\omega\in\Omega} \lambda(\omega)\, p_i(\omega) = \sum_{\omega\in\Omega} p_i(\omega) \sum_k b_k [\omega \in E_k]$$

$$= \sum_k b_k \sum_{\omega\in\Omega} p_i(\omega) [\omega \in E_k] = \sum_k b_k\, p_i(E_k) ,$$

and so by (3.9) and the fact that each of the terms $b_k\, p_i(E_k)$ is positive, λ is a likelihood for $(p_1,\ldots,p_n)$. By (2.5) and (2.6) it then follows that (3.11) and (3.12) hold.

Sufficiency. Suppose that T satisfies CJC_2 and that λ is a likelihood for $(p_1,\ldots,p_n)$. Let $(\omega_1,\omega_2,\ldots)$ be a list of all those $\omega \in \Omega$ for which $\lambda(\omega) > 0$, and let **E** = $\{ E_1, E_2 ,\ldots\}$, where $E_i := \{ \omega_i \}$. Setting $b_k := \lambda(\omega_k)/ \lambda(\omega_1)$ for k = 1,2,…, we have each $b_k > 0$ and $b_1 = 1$. From the fact that λ is a likelihood for $(p_1,\ldots,p_n)$ it follows that (3.9) holds and that $(q_1,\ldots,q_n) \in \Delta^{n+}$, where $q_i(\omega)$ is given by (3.10). Hence, by CJC_2, (3.11) and (3.12) hold. But (3.11) implies (2.5), and (3.12) implies (2.6). □

Remark 3.1. In Wagner (2002) it is shown that Jeffrey updating on a family **E**, followed by Jeffrey updating on a family **F**, produces the same final result as first updating on **F**, and then on **E**, as long as identical new learning is represented by identical Bayes factors. Indeed, under mild regularity conditions, such a representation is necessary to ensure the commutativity of Jeffrey updates. Theorem 3.2 is a companion to the aforementioned result, and furnishes additional evidence that identical new learning should be reflected in identical Bayes factors. See also Wagner (1997,1999,2001,2003).

4. Relevance Quotients as Possible Rivals of Bayes Factors.

Suppose someone wished to argue that new learning should be represented by relevance quotients, rather than Bayes factors. One rejoinder to this proposal is to note that relevance quotients fail to satisfy an important desideratum for a measure of new learning, namely, that of effacing all traces of the prior.[12] On the other hand, there is a simple variant of the Jeffrey updating formula (3.1) that involves relevance quotients, namely,

$$(4.1) \qquad q(\omega) = \sum_k r_k \, p(\omega) \, [\omega \in E_k] \, ,$$

where $r_k := q(E_k)/ p(E_k)$. If an analogue of Theorem 2.2 held for this parameterization of JC, it would arguably detract from the support that Theorem 2.2 offers for the assertion that Bayes factors are the appropriate measure of what is learned from new evidence alone. So it is worth investigating how things stand on this issue. First, we need an analogue of CJC_2 for the parameterization (4.1). The following may seem like a natural choice:

CJC_3 : For all families **E** = $\{E_k\}$ of nonempty, pairwise disjoint subsets of Ω, all $(p_1,\ldots,p_n) \in \Delta^{n+}$ such that $p_i(E_k) > 0$ for all i and all k, and all sequences (r_k) of positive real numbers such that $(q_1,\ldots,q_n) \in \Delta^{n+}$, where

(4.2) $$q_i(\omega) := \sum_k r_k\, p_i(\omega)[\,\omega \in E_k] ,$$

it is the case that

(4.3) $$\sum_k r_k\, T(p_1,\ldots, p_n)\, [\,\boldsymbol{.} \in E_k] \in \Delta$$

and the following commutativity condition holds:

(4.4) $$T(\sum_k r_k\, p_1 [\,\boldsymbol{.} \in E_k] , \ldots , \sum_k r_k\, p_n [\,\boldsymbol{.} \in E_k]) = \sum_k r_k\, T(p_1,\ldots, p_n)\, [\,\boldsymbol{.} \in E_k] .$$

But CJC_3 neither implies nor is implied by EB, as the following examples show:

(i) Let $\Omega = \{\omega_1,\omega_2\}$. In this case *any* pooling operator $T: \Delta^{n+} \to \Delta$ that preserves unanimity satisfies CJC_3. For here we must have either $\mathbf{E} = \{\{\omega_1\}\}$, $\mathbf{E} = \{\{\omega_2\}\}$, $\mathbf{E} = \{\{\omega_1\},\{\omega_2\}\}$, or $\mathbf{E} = \{\Omega\}$. In the first two cases, and in the third case when $r_1 \neq r_2$, if $(p_1,\ldots,p_n) \in \Delta^{n+}$ and (r_k) are such that $(q_1,\ldots,q_n) \in \Delta^{n+}$(indeed, if it it merely true that $(q_1,\ldots,q_n) \in \Delta^n$), then it must be the case that $p_1=\ldots= p_n$ and so $q_1=\ldots= q_n$. Hence (4.3) and (4.4) hold in virtue of unanimity preservation and (4.2). In the third case when $r_1 = r_2$ (which implies that their common value is 1) and in the fourth case, if $(p_1,\ldots,p_n) \in \Delta^{n+}$ and (r_k) are such that $(q_1,\ldots,q_n) \in \Delta^n$, then $p_i = q_i$ for i=1,…,n. Hence, (4.3) and (4.4) actually hold for *every* pooling operator. But there are clearly many unanimity preserving pooling operators (for example, non-dictatorial weighted arithmetic means) that fail to be externally Bayesian.

(ii) Let $\Omega = \{\omega_1,\omega_2,\omega_3\}$, $(p_1(\omega_i)) = (1/3,1/3,1/3)$, and $(p_2(\omega_i)) = (1/3,1/2,1/6)$. Update p_1 to q_1 and p_2 to q_2 by Jeffrey conditioning on $\mathbf{E} = \{ \{\omega_1\},\{\omega_2\},\{\omega_3\} \}$, parameterized as in (4.2), with $(r_i) = (2,1/2,1/2)$, whence $(q_1(\omega_i)) = (2/3,1/6,1/6)$ and $(q_2(\omega_i)) = (2/3,1/4,1/12)$. Let T be the externally Bayesian pooling operator defined by a normalized geometric mean. Then $(T(q_1,q_2)(\omega_i)) = (8/s, \sqrt{6}/s, \sqrt{2}/s)$, where $s = 8 + \sqrt{6} + \sqrt{2}$. Suppose, on the other hand, that we first pool p_1 and p_2, so that $(T(p_1, p_2)(\omega_i)) =$

$(2/s^*, \sqrt{6}/s^*, \sqrt{2}/s^*)$, where $s^* = 2 + \sqrt{6} + \sqrt{2}$. However, applying formula (4.2), with $(r_i) = (2,1/2,1/2)$, to update $T(p_1, p_2)$ fails to produce a pmf (since $4/s^* + \sqrt{6}/2s^* + \sqrt{2}/2s^* > 1$) and thus violates (4.3).

Remark 4.1. Note that normalizing the quantities $4/s^*$, $\sqrt{6}/2s^*$, and $\sqrt{2}/2s^*$ so that they sum to one yields a pmf identical to $T(q_1,q_2)$. So perhaps CJC_3 is too stringent . Perhaps the appropriate commutativity condition for Jeffrey conditioning parameterized in terms of relevance quotients is

CJC_3^* : For all families $\mathbf{E} = \{E_k\}$ of nonempty, pairwise disjoint subsets of Ω, all $(p_1,\ldots,p_n) \in \Delta^{n+}$ such that $p_i(E_k) > 0$ for all i and all k, and all sequences (r_k) of positive real numbers such that $(q_1,\ldots,q_n) \in \Delta^{n+}$, where

(4.5) $$q_i(\omega) := \sum_k r_k\, p_i(\omega)[\,\omega \in E_k\,],$$

it is the case that

(4.6) $$0 < S := \sum_k r_k\, T(p_1,\ldots, p_n)(E_k)$$

$$(= \sum_{\omega\in\Omega} \sum_k r_k\, T(p_1,\ldots, p_n)(\omega)\,[\,\omega\in E_k]\,) < \infty, \text{ and}$$

(4.7) $$T(\sum_k r_k\, p_1 [\,\cdot\in E_k], \ldots, \sum_k r_k\, p_n [\,\cdot\in E_k]) =$$

$$= \sum_k r_k\, T(p_1,\ldots, p_n)\,[\,\cdot\in E_k] / S.$$

Theorem 4.1. EB implies CJC_3^*.

Proof. Let $T: \Delta^{n+} \rightarrow \Delta$ be externally Bayesian, and set $\lambda(\omega) = \sum_k r_k\,[\omega \in E_k]$. Then for i = 1,…,n,

(4.8) $$\sum_{\omega\in\Omega} \lambda(\omega)\, p_i(\omega) = \sum_{\omega\in\Omega} \sum_k r_k\, p_i(\omega)\,[\omega \in E_k]$$

$$= \sum_{\omega\in\Omega} q_i(\omega) = 1.$$

Hence λ is a likelihood for $(p_1,\ldots, p_n)$ and (4.6) and (4.7) follow from (2.5) and (2.6). □

However, CJC_3^* does not imply EB, and for the very same reason that CJC_3 fails to imply EB: the premisses under which CJC_3^* certifies that (4.6) and (4.7) will hold so restrict the domain of application of the pooling operator T that any unanimity-preserving pooling operator (most of which clearly fail to satisfy EB) does the trick.[13] Suppose then that we further modify these premisses as follows:

CJC_3^{**} : For all families $\mathbf{E} = \{E_k\}$ of nonempty, pairwise disjoint subsets of Ω, all $(p_1,\ldots,p_n) \in \Delta^{n+}$ such that $p_i(E_k) > 0$ for all i and all k, and all sequences (r_k) of positive real numbers such that

(4.9) $$S_i := \sum_k r_k\, p_i(E_k) < \infty \quad, i = 1,\ldots,n,$$

and $(q_1,\ldots,q_n) \in \Delta^{n+}$, where

(4.10) $$q_i(\omega) := \sum_k r_k\, p_i(\omega)[\,\omega \in E_k] \,/\, S_i \,,$$

it is the case that

(4.11) $$0 < S := \sum_k r_k\, T(p_1,\ldots, p_n)\,(E_k)$$

$$(= \sum_{\omega\in\Omega} \sum_k r_k\, T(p_1,\ldots, p_n)(\omega)\,[\,\omega \in E_k]\,) < \infty \,, \text{ and}$$

(4.12) $$T(\sum_k r_k\, p_1 [\,\cdot \in E_k] /S_1 \,, \ldots , \sum_k r_k\, p_n [\,\cdot \in E_k] /S_n) =$$

$$= \sum_k r_k\, T(p_1,\ldots, p_n)\, [\,\cdot \in E_k] \,/\, S.$$

It is easy to see that CJC_3^{**} *is* equivalent to EB, the proof being virtually the same as the proof of the equivalence of CJC_2 and EB. But this equivalence has been purchased at too high a price. After all, the whole point of initiating the exercise in this section was to demonstrate that identical relevance quotients were, at least from the standpoint of commutativity theorems, not inferior to Bayes factors in the representation of identical new learning. But unlike the quantities b_k uniformly occurring in (3.10) and (3.12), which represent Bayes factors (see notes 10 and 11 *infra*), the quantities r_k uniformly occurring in (4.10) and (4.12) no longer represent relevance quotients. Indeed, (4.13) $\rho_{q_i,p_i}(E_k) = r_k / S_i$, $i = 1,\ldots,n$

and

(4.14) $\rho_{q,p}(E_k) = r_k / S$,

where $p := T(p_1,\ldots, p_n)$ and $q := \sum_k r_k\, T(p_1,\ldots, p_n)\, [\,\cdot \in E_k] / S$.
In general the normalizing quantities S_i will differ from each other, as well as from the normalizing quantity S. So if identical relevance quotients are supposed to represent identical new learning, (4.13) and (4.14) show that, contrary to assumption, individuals in the group are *not* Jeffrey updating based on identical new learning, either when they update their priors or when the group "surrogate" updates $T(p_1,\ldots, p_n)$.

Notes

1. Many treatments of the problem of combining probability distributions deal with pooling operators defined, slightly more generally, on Δ^n. For reasons that will soon be obvious, we are excluding consideration of cases where there is no state of the world ω to which all individuals assign positive probability. This is, however, considerably less stringent than the usual restriction posited in treatments of external Bayesianity, namely, that for all pmfs p_i under consideration, $\mathrm{Supp}(p_i) = \Omega$.

2. This is a rather weak unanimity condition. One often sees in the literature a much stronger condition of this type postulated, namely, that for each $c \in [0,1]$, if $p_1(\omega) = \ldots = p_n(\omega) = c$, then $T(p_1,\ldots,p_n)(\omega) = c$. It isn't clear that pooling should be restricted in this way in any of the scenarios (i)-(v).

3. See Jeffrey [1992,pp.117-119].

4. Specifically, $[\omega \in E] = 1$ if $\omega \in E$, and 0 if $\omega \in E^c$. This notation is a special case of the wonderful notational device introduced by Kenneth Iverson [1962] in his programming language APL, in which propositional functions are denoted simply by enclosing the propositions in question in square brackets. See Knuth [1992] for an

elaboration of how this notation simplifies the representation of various restricted sums.

5. Note that $\sum_{\omega \in \Omega} \Pi_{1\leq i \leq n} p_i(\omega)^{w(i)}$ is strictly positive since $(p_1,\ldots,p_n) \in \Delta^{n+}$, and finite since $\Pi_{1\leq i \leq n} p_i(\omega)^{w(i)} \leq \sum_{1\leq i \leq n} w(i)p_i(\omega)$ by the generalized arithmetic-geometric mean inequality. See Royden (1963, p. 94).

6. In the standard exposition of Jeffrey conditioning (also called *probability kinematics*) the family **E** = $\{E_k\}$ is taken to be a *partition* of Ω, so that, in addition to pairwise disjointness of the events E_k, one has $U_k E_k = \Omega$. Standardly, (e_k) is a sequence of *nonnegative* real numbers summing to one, and it is assumed that zeros are not raised, i.e., that $p(E_k) = 0$ implies that e_k (= $q(E_k)$) = 0. Finally, it is stipulated that $0 \times p(\omega|E_k) = 0$ if $p(E_k) = 0$, so that $e_k\, p(\omega|E_k)$ is well-defined even if $p(\omega|E_k)$ isn't. Given the standard formulation, our family **E** simply comprises those E_k in the partition for which $e_k > 0$. Conversely, our format yields the standard one by associating to our family **E** = $\{E_1, E_2,\ldots\}$, if it fails to be a partition, the partition **E** = $\{E_0, E_1, E_2,\ldots\}$, where $E_0 := \Omega \setminus \cup_k E_k$, and setting $e_0 = 0$.

7. Condition 2) is known as the *rigidity condition*, and is equivalent to the condition $q(A|E_k) = p(A|E_k)$ for all k and all subsets A of Ω. See Jeffrey [1992, pp. 124-125]. This condition is also called the *sufficiency condition*.

8. It is important to note here that individuals can have the same *isolated* sensory, or other, experience without undergoing what I have chosen to call identical *new learning*. Suppose, in the example under consideration, that E_1 = rain, E_2 = clear, p_1 is your prior, and p_2 is mine. We then both look at the sky at the same time, with the result that you lower your subjective probability of rain from 5/7 to ½, and I raise mine from 2/7 to ½. We both saw the same sky, but it must have been lighter than you expected and darker than I expected, given our prior probabilities. So our identical sensory experience has produced different new learning. An equivalent way of putting this is that in looking at the sky we have had different *holistic experiences* (experiences considered in light of ambient memory and prior probabilistic assessment). This example was suggested by a similar

illustration in Osherson (2002). See also Lange (2000) for a detailed elaboration of this claim.

9. See, e.g., Fitelson(1998),Heckerman(1988), Hawthorne (2004), Jeffrey (2004), Kemeny and Oppenheim (1952), Schum(1994), and Wagner (1997,1999, 2000, 2002, 2003)

10. Note that, given (3.9) and (3.10), we have $\beta_{q_i,p_i}(E_k : E_1) = b_k$ for $i = 1,\ldots,n$, so these conditions are capturing precisely the situation in which individuals 1,…,n are Jeffrey conditioning their priors based on identical new learning.

11. Note that $\beta_{q,p}(E_k : E_1) = b_k$, where $p := T(p_1,\ldots, p_n)$ and $q := (\sum_k b_k T(p_1,\ldots, p_n) [\,\cdot \in E_k] / \sum_k b_k T(p_1,\ldots, p_n) (E_k)$.

12. For example, if $q(E)/p(E) = 2$, we know that $p(E) \leq ½$. On the other hand, if $\mathbf{E} = \{ E_1, \ldots, E_m \}$ and $(b_1,\ldots, b_m)$ is *any* sequence of positive real numbers with $b_1 = 1$, then *every* prior p with $p(E_i) > 0$ for all i admits of a revision by JC to a probability q such that $\beta_{q,p} (E_i : E_1) = b_i$ for all i. See Wagner (2002, p.275, footnote 7).

13. The premisses at issue are of course the demand that (4.2) and (4.5) define pmfs, *with no normalization*. As remarked in note 12 *supra*, specifying the value of a relevance quotient puts severe constraints on a prior, and all the more so when multiple relevance quotients and multiple priors are involved.

References

Bacharach, Michael (1972), “Scientific Disagreement”, Unpublished manuscript.
Fitelson, Branden (1998), “Independent Inductive Support and Measures of Confirmation”, unpublished manuscript.
______ (1999), “The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity”, *Philosophy of Science* 66 (Proceedings), S362-S378.
Genest, Christian, Kevin J. McConway, and Mark J. Schervish (1986), “Characterization of Externally Bayesian Pooling

Operators", *The Annals of Statistics* 14: 487-501.
Good, Irving J. (1950), *Probability and the Weighing of Evidence*. London: Charles Griffin.
______(1983), *Good Thinking*, Minneapolis: University of Minnesota Press.
Hawthorne, James (2004), "Three Models of Sequential Belief Updating on Uncertain Evidence", *Journal of Philosophical Logic* 33: 89-123.
Heckerman, D. (1988), "An Axiomatic Framework for Belief Updates", in L. Kanal and J. Lemmer (eds,), *Uncertainty in Artificial Intelligence 2*, New York: Elsevier Science Publishers, 11-22.
Iverson, Kenneth E. (1962), *A Programming Language*, New York: Wiley.
Jeffrey, Richard (1992), *Probability and the Art of Judgment*, Cambridge: Cambridge University Press.
______(2004), *Subjective Probability: The Real Thing*, Cambridge: Cambridge University Press.
Kemeny, John, and Paul Oppenheim (1952), "Degrees of Factual Support," *Philosophy of Science* 19: 307-24.
Knuth, Donald E. (1992), "Two Notes on Notation", *The American Mathematical Monthly* 99: 403-422.
Lange, Marc (2000), "Is Jeffrey Conditionalization Defective in Virtue of Being Non-Commutative? Remarks on the Sameness of Sensory Experience", *Synthese* 123: 393-403.
Madansky, Albert (1964), "Externally Bayesian Groups", Technical Report RM-4141-PR, RAND Corporation.
______(1978), "Externally Bayesian Groups", Unpublished manuscript, University of Chicago.
Osherson, Daniel (2002), "Order Dependence and Jeffrey Conditionalization", unpublished paper available at: http://www.princeton.edu/~osherson/papers/jeff3.pdf .
Raiffa, Howard (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Reading, MA: Addison-Wesley.
Royden, H. L. (1963), *Real Analysis*, New York: The Macmillan Company.
Schum, David (1994), *The Evidential Foundations of Probabilistic Reasoning*. New York: John Wiley and Sons.
Wagner, Carl G. (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”, Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association, *Philosophy of Science* 68: S165-S175.
______(2002), “Probability Kinematics and Commutativity”, *Philosophy of Science* 69: 266-278.
______(2003), “Commuting Probability Revisions: The Uniformity Rule”, *Erkenntnis* 59: 349-364.