Friday Sep 22 2006 03:29 PM PHOS v72n5 720547 VML

Theory Change and Bayesian Statistical Inference

Jan-Willem Romeyn†

This paper addresses the problem that Bayesian statistical inference cannot accommodate theory change, and proposes a framework for dealing with such changes. It first presents a scheme for generating predictions from observations by means of hypotheses. An example shows how the hypotheses represent the theoretical structure underlying the scheme. This is followed by an example of a change of hypotheses. The paper then presents a general framework for hypotheses change, and proposes the minimization of the distance between hypotheses as a rationality criterion. Finally the paper discusses the import of this for Bayesian statistical inference.

1. Introduction. This paper is concerned with Bayesian statistical inferences. These inferences are here considered in a scheme that generates predictions by means of hypotheses: Bayesian updating is used to adapt a probability over hypotheses to known observations, and this adapted probability is further used to generate predictions over unknown observations. The hypotheses in the scheme represent the theoretical structure that underlies the predictions. However, after we have chosen these hypotheses and a prior probability over them, updating fully determines the probabilities over the hypotheses at any later stage, and thus also the predictions resulting from that. There is no room for any further amendments to the hypotheses or the prior over them after they have been chosen. In Bayesian statistical inference, the theoretical structure is therefore fixed.

The fixity of the theoretical structure in the above schemes is a specific form of a problem for Bayesianism on the whole. In the philosophy of science it has been formulated, among others by Earman (1992, 195–198), as the problem that Bayesianism fails to accommodate theory change. But the fact that Bayesian inference is in this sense dogmatic is at the

†To contact the author, please write to: Department of Philosophy, University of Groningen, UVA/FMG/Psychology, Methodology Unit, Roeterstraat 15, 1018 WB, Amsterdam, Netherlands; e-mail: j.w.romeyn@uva.nl.

Philosophy of Science, 72 (December 2005) pp. 000–000. 0031-8248/2005/7205-0043$10.00

origin of many other criticisms, including the criticism of Dawid (1982) that Bayesian inference is by definition calibrated. Furthermore, as hypotheses can be considered as specific terms in the observation language, changing the hypotheses in the scheme comes down to changing the language with which the predictions are made. The same problem can therefore be seen in light of the fact that Bayesianism fails to accommodate
q1 language change, as noted by Gillies (2000) and discussed elaborately by Williamson (2003).

This paper addresses the above problems with Bayesianism. More in particular, it proposes a way of dealing with theory change within Bayesian statistical inference. The plan of the paper is to introduce the Bayesian scheme for generating predictions from hypotheses, to present an example of such a scheme, then to show in the example how hypotheses can be changed, and finally to give a general framework for it.

2. Hypotheses, Conditioning, and Predictions. This section describes the schemes for making predictions. It defines observations and observational hypotheses in terms of an observational algebra, and it presents degrees of belief as probability assignments over this algebra. This set-theoretical underpinning may seem unnecessary in the context of a short paper. However, as will become apparent in Sections 5 and 6, the underpinning is essential for a correct understanding of hypotheses change.

The predictions range over possible observations K, a set of consecutive natural numbers, say $\{0, 1\}$. At every time t we observe one number $q_t \in K$. We can represent these observations in an observational algebra. Let K^ω be the space of all infinite observation sequences e:

$$e = q_1 q_2 q_3 \cdots. \tag{1}$$

The observational algebra $\mathcal{Q}$, a so-called cylindrical σ-algebra, consists of all possible subsets of the space K^ω. If we denote the tth element in a series e with $e(t)$, we can define an observation Q_t^q as an element of the algebra $\mathcal{Q}$ as follows:

$$Q_t^q = \{e \in K^\omega | e(t) = q\}. \tag{2}$$

Note that there is a distinction between the observations Q_t^q and the values of observations q. The values, represented with small letters, are natural numbers. The observations, denoted with large letters, are elements of the algebra $\mathcal{Q}$.

In the same way we can define an element in the algebra that refers to

a finite sequence of observations. If we define the ordered sequence $e_t = \langle q_1 q_2 \cdots q_t \rangle$, we can write

$$E_t^{e_t} = \{e \in K^\omega | \forall t' \leq t : e(t') = q_{t'}\}. \tag{3}$$

Again, it must be noted that the small letters e_t refer to a sequence of natural numbers, while the large letters E_t are elements of the algebra and carry a sequence of natural numbers as argument. The observations and sequences of observations are related to each other in the natural way:

$$Q_{t+1}^q \cap E_t = E_{t+1}. \tag{4}$$

As in this equation, I normally refer to sequences of observations with the expression E_t, thereby suppressing the reference to the sequence e_t.

Observational hypotheses can also be seen as elements of the observational algebra. If we say of an observational hypothesis h that its truth can be determined as a function of an infinitely long sequence of observations e, then we can define hypotheses as subsets of K^ω in the following way:

$$H = \{e \in K^\omega | W_h(e) = 1\}. \tag{5}$$

Here $W_h(e) = 1$ if and only if the proposition h is true of e, and $W_h(e) = 0$ otherwise. The hypotheses can thus be an argument of the same probability functions over the observational algebra. A partition of hypotheses is a collection $\mathcal{H} = \{H_0, H_1, \ldots, H_N\}$ defined by the following condition for the indicator functions W_{h_n}:

$$\forall e \in K^\omega : \sum_n W_{h_n}(e) = 1. \tag{6}$$

This means that the hypotheses H_n are mutually exclusive and jointly exhaustive sets in K^ω.

Belief states are represented with probability functions over $\mathcal{Q}$. They take observations Q_t^q, sequences E_t, and hypotheses H_n as arguments. The functions are defined relative to a partition $\mathcal{H}$ and a sequence of known observations e_t: the function $p_{[\mathcal{H},e_t]}$ represents the belief state upon observing E_t under the assumption of a partition $\mathcal{H}$. It can be constructed by conditioning a prior probability function $p_{[\mathcal{H},e_0]}$ on the observations E_t:

$$p_{[\mathcal{H},e_t]}(\cdot) = p_{[\mathcal{H},e_0]}(\cdot | E_t). \tag{7}$$

Because of this, we have $p_{[\mathcal{H},e_t]}(E_t) = 1$. Updating the probability by simple conditioning is known as Bayes' rule. Both the probabilities assigned to observations and those assigned to hypotheses can be updated for new observations in this way. The probability before updating is called the prior probability, and the one after updating the posterior.

To calculate the predictions, we can employ a partition of hypotheses, and apply the law of total probability:

$$p_{[\mathcal{H},e_i]}(Q^q_{t+1}) = \sum_n p_{[\mathcal{H},e_i]}(H_n)p_{[\mathcal{H},e_i]}(Q^q_{t+1}|H_n). \tag{8}$$

The terms $p_{[\mathcal{H},e_i]}(Q^q_{t+1}|H_n)$ are called the posterior likelihoods of Q^q_{t+1} on the hypotheses H_n. The prediction is obtained by weighing these posterior likelihoods with the posterior probability over the hypotheses, $p_{[\mathcal{H},e_i]}(H_n)$.

Both posterior probabilities of equation (8) can be obtained from a Bayesian update of the prior probability $p_{[\mathcal{H},e_0]}$ according to expression (7). In this paper the likelihoods do not change upon conditioning:

$$p_{[\mathcal{H},e_i]}(Q^q_{t+1}|H_n) = p_{[\mathcal{H},e_0]}(Q^q_{t+1}|H_n). \tag{9}$$

That is, the observations influence the predictions only via the probability over the hypotheses. Part of the input probabilities for generating the predictions $p_{[\mathcal{H},e_i]}(Q^q_{t+1})$ are therefore the likelihoods $p_{[\mathcal{H},e_0]}(Q^q_{t+1}|H_n)$.

The predictions are further determined by the probability assignment over the hypotheses, $p_{[\mathcal{H},e_i]}(H_n)$. This probability can be determined by means of the relation

$$p_{[\mathcal{H},e_i]}(H_n) = p_{[\mathcal{H},e_{i-1}]}(H_n)\frac{p_{[\mathcal{H},e_{i-1}]}(Q^q_i|H_n)}{p_{[\mathcal{H},e_{i-1}]}(Q^q_i)}, \tag{10}$$

where q equals the last number in the sequence e_i. Note that the denominator $p_{[\mathcal{H},e_{i-1}]}(Q^q_i)$ can be rewritten with equation (8), substituting $t = i - 1$. Recall further that the likelihoods $p_{[\mathcal{H},e_{i-1}]}(Q^q_i|H_n)$ are in this paper equal for all sequences e_{i-1}, as expressed in equation (9). The posterior probability $p_{[\mathcal{H},e_i]}(H_n)$ can therefore be determined recursively by the prior probability $p_{[\mathcal{H},e_0]}(H_n)$ for all n, and the likelihoods $p_{[\mathcal{H},e_0]}(Q^q_i|H_n)$ for all n and $i \leq t$. These are the other input probabilities for generating the predictions.

In sum, predictions can be generated if we assume hypotheses, their likelihoods, and a prior probability over them. The prior and the likelihoods are first used to determine the posterior probability over the partition. The likelihoods are then used together with this probability over the partition for generating the prediction itself. The whole construction that uses hypotheses to generate predictions is called the hypotheses scheme.

3. Contaminated Cows. This section gives an example of a hypotheses scheme. Needless to say, the case presented falls short of actual scientific cases in many respects. The focus here is on the conceptual issues rather than on actual applications.

Consider a veterinary investigating a herd of cows during an epidemic,

classifying them into contaminated and uncontaminated. The farmer claims that the herd has been treated with a drug that reduces the risk of contamination. It is an accepted fact about the epidemic that the average incidence rate among untreated cows is 0.4, as more than half of the cows shows a natural resistance against contamination from other cows. The incidence rate among treated cows is 0.2 on average, because the drug is not always effective. The aim of the investigation is to decide whether the cows have been treated with the drug, and further to predict the incidence rate of the contamination in the herd.

The observations of the veterinary consist in test results concerning a number of cows. The result of testing cow t can be that it is contaminated, $q_t = 1$, or that it is not, $q_t = 0$. The test results can then be framed in the observational algebra. The veterinary may set up a scheme using a partition $\mathcal{D}$ of two hypotheses on treatment with the drug, in which D_1 means that the cows are in fact treated and D_0 means that they are not. It must be noted that these hypotheses are not linked to observations directly, since the observations only concern contaminations of cows. The relation that treatment bears to the observations is given by the incidence rates, and this relation is purely statistical. For the observational content of the hypothesis on treatment D_1 we may take

$$W_{d_1}(e) = \begin{cases} 1 & \text{if } f(e) = 0.2 \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

where $f(e)$ is the relative frequency of results $q_t = 1$ in the infinite sequence e. The hypothesis D_0 may be defined in a similar way using $f(e) = 0.4$. A more precise definition is that the hypotheses comprise all so-called Von Mises collectives for the given incidence rates, but for present purposes the loose definition suffices.

Being sets in the observational algebra, the hypotheses can also appear as arguments in the probability functions $p_{[\mathcal{D},e_t]}$. The fact that the veterinary is undecided on whether the farmer has treated his cows can be reflected in

$$p_{[\mathcal{D},e_0]}(D_0) = p_{[\mathcal{D},e_0]}(D_1) = 0.5. \tag{12}$$

Hypotheses on other relative frequencies, which are strictly speaking part of the partition, are thus given a zero probability. The likelihoods of cow t being contaminated on the hypotheses that it has or has not been treated are

$$p_{[\mathcal{D},e_0]}(Q_t^1|D_1) = 0.2, \tag{13}$$

$$p_{[\mathcal{D},e_0]}(Q_t^1|D_0) = 0.4. \tag{14}$$

These likelihoods are determined by the hypotheses. I further assume that

TABLE 1.

	Number of Tests t					
	0	1	2	3	4	5
$p_{[\mathcal{D},e_5]}(D_1)$	.50	.33	.20	.11	.06	.03
$p_{[\mathcal{D},e_5]}(Q^1_{t+1})$	.30	.33	.36	.38	.39	.39

the estimated incidence rates are not affected by the running investigations, so that equation (9) holds.

With these values in place, the veterinary can start to predict the incidence rate in the herd, and decide over the treatment efforts by the farmer. Imagine that the first five test results are positive,

$$e_5 = 11111. \tag{15}$$

Subsequent updating on these test results yields the probabilities and predictions shown in table 1. The probability that the farmer has treated his cows diminishes, and the probability that the next test result is positive tends to 0.4.

The conclusions expressed in the above values are that the farmer very probably did not treat his cows, and that a random cow from the herd has a probability close to 0.4 of being contaminated. It must be stressed, however, that these conclusions follow from the test results only if they are combined with the hypotheses scheme $\mathcal{D}$. The scheme offers two possible hypotheses, and the observations are used to divide the probability between them. It is only relative to the partition $\mathcal{D}$ that most probability settles on D_0 after e_5, so that the predictions are equal to the likelihoods that D_0 prescribes for the test results.

This example thus illustrates that the hypotheses in the scheme determine a range of probabilistic patterns, from which the observations may select the best fitting one. The hypotheses partition functions as an assumption on what patterns can be picked up in the observations. The partition may therefore be called an inductive assumption.

Finally, it can be noted that the partition of hypotheses is associated with the theory underlying the scheme. In this case it concerns a classification of a state of the cows into treated and not treated. Both these concepts come with specific observational contents, which define the relevant patterns in the observations. There is no conceptual space within the hypotheses scheme, at least not as it is set up in the above, to conclude anything other than that the cows are treated or not treated. In order to create this conceptual space, we must add hypotheses to the scheme.

4. Careless Vaccination. This section shows how the hypotheses employed

in the above scheme can be changed. I describe this change, and illustrate that it allows us to derive different conclusions and predictions.

Imagine that the veterinary becomes suspicious of the test results. After all, more than half of the cows are normally immune. The sequence of test results must therefore be a rather unusual stochastic fluctuation on the average relative frequency of 0.4. The veterinary therefore decides to reconsider the inductive assumptions that underly the scheme, and to run a number of additional tests with an adapted scheme. In particular, she investigates the drug that the farmer claims to have used, and finds that it is a vaccinate with a rather unstable quality. In most cases it works very well, even reducing the risk of contamination to 0.025, but careless use turns the vaccinate into a substance that causes a portion of 0.9 cows to be, or at least to appear, contaminated after treatment. The hypotheses that the veterinary wants to add to the scheme are that the drug has been used either carefully or carelessly.

The additional hypotheses may be collected in a separate partition $\mathcal{C}$, with C_1 for careful and C_0 for careless treatment. Both hypotheses only apply to the case in which the cows have actually been treated, D_1. The combined partition is $\mathcal{B} = \{B_0,\ B_{10},\ B_{11}\}$ in which $B_0 = D_0$, $B_{10} = D_1 \cdot C_0$, and $B_{11} = D_1 \cdot C_1$. Hypothesis B_0 is again defined with the relative frequency of 0.4, and the new hypotheses B_{10} and B_{11} can be defined with 0.9 and 0.025 respectively. These three relative frequencies define the new partition.

It is notable that the hypotheses B_{10} and B_{11} cannot be viewed as intersections $D_1 \cap C_0$ and $D_1 \cap C_1$: judged from the definition using relative frequencies, the original set D_1 and both sets B_{10} and B_{11} are disjoint. The relation between the old and the new hypotheses is a rather different one. We must imagine that within every infinite sequence $e \in D_1$, that is, within every possible world in which all cows are treated, we make a further selection of the observations q_t into those concerning cows that have been vaccinated with care, and those concerning cows that have been vaccinated carelessly. So B_{10} and B_{11} can be distilled from the old one by breaking up every $e \in D_1$, for which $f(e) = 0.2$, into two subrows e_0 and e_1 by means of a place selection, taking care that the relative frequencies of the two subrows are 0.9 and 0.025 respectively, and by grouping these subrows into B_{10} and B_{11}. Because $0.025 < 0.2 < 0.9$, such place selections can always be constructed.

The likelihoods of the hypotheses may again be equated to the relative frequencies that define the hypotheses:

$$p_{[\mathcal{B},e_0]}(Q_t^1|B_{10}) = 0.9, \tag{16}$$

$$p_{[\mathcal{B},e_0]}(Q_t^1|B_{11}) = 0.025. \tag{17}$$

TABLE 2.

	Number of Tests t					
	5	7	9	11	13	15
$p_{[\mathcal{B},e_{15}]}(B_{10})$	.01	.03	.14	.49	.80	.95
$p_{[\mathcal{B},e_{15}]}(Q^1_{t+1})$	.39	.42	.47	.62	.80	.88

In order to arrive at the overall incidence rate of 0.2 for treated cows, the veterinary may further assume that a portion of 0.2 of all farmers do not treat the vaccinate with the necessary care, as $0.2 \times 0.9 + (1 - 0.2) \times 0.025 = 0.2$. I come back to this choice in Section 6. Finally, using the probability assignment after five tests, the combined probability of treatment with the drug and the lack of care is

$$p_{[\mathcal{B},e_5]}(B_{10}) = 0.03 \times 0.2 = 0.006 \tag{18}$$

It must be noted that with the employment of $\mathcal{B}$, the probability over the observational algebra really undergoes an external shock: instead of allocating 0.030 probability on the set D_1, we now allocate 0.006 on B_{10} and 0.024 on B_{11}.

With these new hypotheses and the associated inductive assumptions, the veterinary can run a number of additional tests. Let us say that the next ten test results are all positive too,

$$e_{15} = 111111111111111. \tag{19}$$

Subsequent updating on these test results yields the probabilities and predictions shown in table 2. Now the probability for B_{10} approaches 1, while the predictions for a cow in the herd to be contaminated tend to 0.9. Clearly these values differ from those that were to be expected on the basis of $\mathcal{D}$.

The conclusions expressed in these values are that the farmer did treat his cows with the drug, but that he did not apply it with the necessary care. The further conclusion is that the incidence rate of the epidemic in his herd is 0.9. Again, these conclusions are drawn from the test results in combination with the inductive assumptions of partition $\mathcal{B}$. It is only when compared to the other members of the partition that the hypothesis B_{10}, which prescribes an incidence rate of 0.9, fits the test results best. For present purposes, however, it is most notable that these conclusions differ from those derivable from $\mathcal{D}$.

5. A Framework for Changing Partitions. The above illustrates how we can change a partition of hypotheses during an update procedure. This section gives a general framework for such changes, and draws attention to the need for new criteria of rationality to guide them.

On the change of partition itself I can be relatively brief. Let us say that the old partition $\mathcal{H} = \{H_0, H_1, \ldots, H_N\}$ consists of hypotheses H_n with likelihoods

$$p_{[\mathcal{H},e_t]}(Q^q_{t+1}|H_n) = \theta^q_n. \tag{20}$$

The addition of a partition $\mathcal{F} = \{F_0, F_1, \ldots, F_M\}$ to this partition generates a combined partition $\mathcal{G} = \mathcal{H} \times \mathcal{F}$, which consists of $N \times M$ hypotheses $G_{nm} = H_n \cdot F_m$. Each of these hypotheses may be associated with a relative frequency of the observation q, denoted γ^q_{nm}, so that

$$p_{[\mathcal{G},e_t]}(Q^q_{t+1}|G_{nm}) = \gamma^q_{nm}. \tag{21}$$

The details of the partition change may be such that for some of the H_n we have that $\gamma^q_{nm} = \theta^q_n$ for all q and m. We can then collect the hypotheses G_{nm} under the single index number n, as for example B_0 above. More in general, if two hypotheses G_{nm} and $G_{n'm'}$ are such that $\gamma^q_{nm} = \gamma^q_{n'm'}$ for all q, we can merge them into a single hypothesis. In the extreme case in which for all q the γ^q_{nm} vary only with m, the change of partition comes down to a replacement of $\mathcal{H}$ by $\mathcal{F}$.

With the introduction of new hypotheses, the probability over the observational algebra undergoes an external shock. First, the probability over the hypotheses themselves changes. But since the new hypotheses have different likelihoods, the probability over most other elements of the algebra changes as well. It is in this paper assumed that at the time of change τ, the new probability assignment over the hypotheses observes the following restriction:

$$\sum_m p_{[\mathcal{G},e_\tau]}(G_{nm}) = p_{[\mathcal{H},e_\tau]}(H_n). \tag{22}$$

That is, the probability assignment arrived at by updating over $\mathcal{H}$ is taken over into the new partition $\mathcal{G}$. This restriction serves to link every collection $\cup_m G_{nm}$ to the original hypotheses H_n, but it can be dropped if further details of the partition change permit it. Finally, within the limits set by this restriction, the probabilities of the hypotheses G_{nm} can vary freely.

It can be noted that the change in probability due to partition change is not one that can be represented as Bayesian conditioning. Conditioning determines how to adapt probability assignments if for some observation Q^q_t or E_t the probability is externally fixed to 1. It is quite different to set the probability of a number of hypotheses H_n to zero, and to redistribute this probability over new hypotheses G_{nm}. A partition change is therefore an external shock to the probability assignment to which we cannot apply Bayesian updating. Now there are many arguments to the effect that Bayesian updating is the only rational way to adapt a probability assignment to new information, but these arguments do not apply in this

case. It seems that the possibility of partition change necessitates new criteria of rationality, and the definition of an associated update operation.

6. Distance between Partitions. This section answers the need for a rationality criterion and an associated update operation. In particular, it elaborates on a distance function between the old and the new partition, and shows how to minimize this distance during the partition change.

Williamson (2003) argues that changes in the assignment must be conservative, that is, as small as possible, and further that such conservatism can be explicated by a minimization of the cross-entropy distance function between the old probability p_0 and the new probability p, under the restrictions imposed by the external shock. The distance function is defined by

$$\Delta(p, p_0) = \sum_U p(U) \log \frac{p(U)}{p_0(U)}, \tag{23}$$

where the index U runs over all sets in the finite algebra over which p_0 and p are defined. As elaborated in Kullback (1959) and Paris (1994, 120–126), minimizing this distance under the external restrictions effectively minimizes the information change that is induced in the probability assignment by the external shock. Interestingly, the operation of minimizing cross-entropy coincides with the operation of a Bayesian update in the case that some probability $p_{[\mathcal{H},e_l]}(Q_t^q)$ is restricted to 1. It therefore accords with Bayesian statistical inference to adopt the minimization of cross-entropy as the update operation in cases of partition change.

We are not yet done with the update operation for partition change. For one thing, the above distance function blows up if the algebra contains an infinite number of elements, as is the case for the algebra $\mathcal{Q}$. We need to select a finite collection of elements of the algebra, for which we may then minimize the distance between the old and the new probability assignment.

As already indicated in the example, it is rather intuitive to choose a minimization of the distance between the likelihoods of H_n and of the associated collection $\cup_m G_{nm}$: the likelihoods fully express the hypotheses, and the distance between the likelihoods is therefore an intuitive measure for the closeness of the two partitions.

A further reason for choosing this collection can be found in the relation between the old and the new hypotheses. Recall that the likelihoods of observations Q_t^q in H_n are determined by the relative frequencies of the observations $q \in K$ within the possible worlds for which H_n is true. With the change of hypotheses, we effectively make a further division of these possible worlds into the hypotheses G_{nm}: each infinite sequence of obser-

vations $e \in H_n$, having a relative frequency θ_n^q, must be split into M infinite subsequences e_m, having relative frequencies γ_{nm}^q, and these subsequences can then be incorporated into separate hypotheses, $e_m \in G_{nm}$. Because the hypotheses G_{nm} are derived from the original hypotheses H_n in this way, we may expect the relative frequency associated with the aggregate $\cup_m G_{nm}$ to be the same as, or at least close to, the original relative frequency associated with H_n.

Any hypothesis prescribes the likelihoods for infinitely many observations $Q_{\tau+t}^q$, associated with different times $t \geq 0$. However, these likelihoods are in this paper constant, and it seems natural to define the distance between the partitions as the distance between the likelihoods at a single time $\tau + t$. For p_0 we can use the old likelihoods $p_{[\mathcal{H},e_\tau]}(Q_{\tau+t}^q|H_n)$. For p we use the aggregated likelihoods, given by

$$\begin{aligned}\gamma_n^q &= p_{[\mathcal{G},e_\tau]}(Q_{\tau+t}^q|\cup_m G_{nm}) \\ &= \sum_m \frac{p_{[\mathcal{G},e_\tau]}(G_{nm})}{\sum_m p_{[\mathcal{G},e_\tau]}(G_{nm})} p_{[\mathcal{G},e_\tau]}(Q_{\tau+t}^q|G_{nm}) \qquad (24) \\ &= \sum_m \rho_{nm}\gamma_{nm}^q. \qquad (25)\end{aligned}$$

Here the ρ_{nm} are defined by the fraction in equation (24), so that $\sum_m \rho_{nm} = 1$. The γ_n^q are a function of these ρ_{nm}.

We can now use the distance function to find the aggregated likelihoods $p_{[\mathcal{G},e_\tau]}(Q_{\tau+t}^q|\cup_m G_{nm})$ that are closest to the likelihoods $p_{[\mathcal{H},e_\tau]}(Q_{\tau+t}^q|H_n)$, for any time t. These distances are defined for each hypothesis H_n separately:

$$\Delta_n(\rho_{nm}) = \sum_q \gamma_n^q \log\frac{\gamma_n^q}{\theta_n^q}. \qquad (26)$$

The distance is a function of the fractions ρ_{nm}, which determine how the probability of H_n is distributed over the G_{nm}. The update operation after a hypotheses change is to find, for every H_n separately, the values of ρ_{nm} that minimize the distance function Δ_n.

This can be employed to provide a further underpinning for the choice of the probabilities $p_{[\mathcal{B},e_5]}(B_{10})$ and $p_{[\mathcal{B},e_5]}(B_{11})$ in the example. It was stated there that the veterinary chooses these probabilities in order to arrive at the overall incidence rate of 0.2. Note that the distance between the likelihoods of $\mathcal{H}$ and the aggregated likelihoods of $\mathcal{G}$ is zero and therefore minimal if we find values for ρ_{nm} so that $\gamma_n^q = \sum_m \rho_{nm}\gamma_{nm} = \theta_n^q$. In the case of the partitions $\mathcal{D}$ and $\mathcal{B}$, the equation simply becomes $0.9 \times \rho_{10} + 0.025 \times (1 - \rho_{10}) = 0.2$, for which $\rho_{10} = 0.2$ is the solution.

It must be stressed that the above is not the full story on partition change. There are many cases of partition change that are not covered

by the above framework, but that can in principle be treated in a similar way. One such case deserves separate attention here. The above example presents a probability assignment that is not open-minded: almost all hypotheses on relative frequencies are given a zero probability. This may cause the impression that the framework for partition change can only be applied if the old probability assignment is not open-minded. It may be hard to see what other hypotheses can be added if, for instance, the prior probability already includes all possible hypotheses on relative frequencies. However, the above framework can also be used to change a partition of all hypotheses on relative frequencies into a partition of hypotheses that concern Markov processes. The application of the framework for partition change is thus not limited to cases in which the prior is not open-minded.

7. Concluding Remarks. The above shows how we can frame a partition change, and provides a procedure to render this change rational, employing a distance function between the partitions. I complete the paper with a summary and some remarks on the proposed framework in the context of Bayesian statistical inference.

The proposed framework enables us to adapt the hypotheses that function in a scheme for making predictions. By writing down the predictions in terms of such hypotheses schemes, I locate the theoretical structure underlying the predictions inside the probability assignment. Theoretical developments can therefore be framed as external shocks to the probability assignment representing the opinions, just as new observations. I then argue that the operation that updates the assignment for the external shock is a generalized version of Bayesian conditioning, namely cross-entropy minimization. The framework is therefore a natural extension of Bayesian statistical inference. On the whole, the paper proposes an answer to the problem that Bayesian statistical inference cannot accommodate theory change.

The paper may also fulfill a role in an older discussion between inductivists and Popperians: the above basically shows how we can encompass a notion of conjecture within an inductivist setting. It is a typical feature of Carnapian inductive logic that there is no room for an explicit formulation of inductive assumptions, as such assumptions are part and parcel of the choice of language. Conjectures can therefore not be captured within a Carnapian logic. However, the above framework locates the premisses in the hypotheses schemes, and further allows us to change them. It provides a truly nonmonotonic probabilistic inductive logic, in the sense that the inductive assumptions may be altered underway. It is hoped that this paper is a first step in freeing inductive logic from its dependence on language.

REFERENCES

Dawid, A. P. (1982), "The Well-Calibrated Bayesian", *Journal of the American Statistical Association* 77: 605–613.
Earman, J. (1992), *Bayes or Bust*. Cambridge, MA: MIT Press.
Gillies, D. (2001), "Bayesianism and the Fixity of the Theoretical Framework", in D. Corfield and J. Williamson (eds.), *Foundations of Bayesianism*. Dordrecht: Kluwer, 363–379.
q2 Howson, C., and P. Urbach (1989), *Scientific Reasoning: The Bayesian Approach*. LaSalle, IL: Open Court.
Kullback, S. (1959), *Information Theory and Statistics*. New York: Wiley.
Paris, J. (1994), *The Uncertain Reasoner's Companion*. Cambridge: Cambridge University Press.
Williamson, J. (2003), "Bayesianism and Language Change", *Journal of Logic, Language and Information* 12: 53–97.

QUERIES TO THE AUTHOR

1 Gillies is dated 2001 in the Refs. Which date is correct?

2 Au: Not cited in text. Either cite or delete from the References.

Philosophy of Science

Department of Philosophy
University of South Carolina
Columbia, SC 29208

Reprint Order Form

Return this form with your proofs ONLY if you wish to order reprints

REPRINT ORDER MUST BE RECEIVED PRIOR TO PRINTING OF JOURNAL ISSUE. Reprints are shipped 2-4 weeks after publication of the journal.

TO BE COMPLETED BY AUTHOR:

Philosophy of Science Vol _____ No _____ Month _____

Author(s): _____ No of pages in article _____

Title of Article: _____

REPRINT CHARGES (please compute)

_____ Quantity	$ _____
Covers	$ _____
Subtotal	$ _____
GST (7% for Canadian destinations only)	$ _____
Non-U.S. and non-Canada shipping	
(Non-U.S. orders add 45% to subtotal)	$ _____
TOTAL DUE (US $)	**$** _____

Reprint rate chart on page 2

Prices include shipping for U.S. and Canadian orders. Non-U.S and non-Canadian orders are shipped via Airmail at an additional cost of 45% of the total printing charge.

SHIPPING INSTRUCTIONS

Name _____

_____

_____

Street_____

City _____ State_____ Zip _____

Country _____

BILLING INSTRUCTIONS (Institutional Orders Only)

Institution _____

_____

City _____ State _____ Zip _____

Country _____

*Phone _____

* Please include a phone number in case we need to contact you about your order

MAKE CHECKS AND PURCHASE ORDERS PAYABLE TO: The University of Chicago Press. Orders need to be received prior to the printing of the journal issue. All orders must be accompanied by one of the three payment options: purchase order, check/money order, or Visa/Mastercard in U.S. currency via an American bank. Terms are net 30 days. Reprints are shipped 2-4 weeks after publication of the journal.

1) ☐ Check or Money Order for total charges is attached **OR** 2) Please charge to: ☐ VISA ☐ MASTERCARD

Cardmember name as it appears on card (please print clearly) _____

Card Number _____Expiration Date _____

Signature _____ Phone _____

3) Institutional Purchase Order No. _____ Purchase Order attached ☐ to come ☐

REPRINT

PRICE LIST

Number of pages

Quantity	1-4	5-8	9-12	13-16	17-20	21-24	add'l 4 pages	COVERS
25	$26	$30	$35	$41	$46	$50	$6	$14.00
50	$34	$40	$46	$53	$59	$66	$7	$18.00
100	$52	$65	$77	$91	$103	$116	$13	$35.00
150	$67	$86	$104	$125	$143	$163	$19	$53.00
200	$87	$111	$135	$162	$174	$202	$25	$70.00
250	$116	$147	$177	$211	$241	$276	$31	$88.00
300	$138	$176	$212	$253	$289	$330	$37	$105.00
350	$161	$204	$243	$294	$336	$373	$43	$123.00
400	$183	$233	$329	$337	$385	$440	$49	$140.00