University of Groningen

Hypotheses and Inductive Predictions

ROMEYN, J.-W.

Published in:
Synthese

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

J. W. ROMEYN

HYPOTHESES AND INDUCTIVE PREDICTIONS

Including examples on crash data

ABSTRACT. This paper studies the use of hypotheses schemes in generating inductive predictions. After discussing Carnap–Hintikka inductive logic, hypotheses schemes are defined and illustrated with two partitions. One partition results in the Carnapian continuum of inductive methods, the other results in predictions typical for hasty generalization. Following these examples I argue that choosing a partition comes down to making inductive assumptions on patterns in the data, and that by choosing appropriately any inductive assumption can be made. Further considerations on partitions make clear that they do not suggest any solution to the problem of induction. Hypotheses schemes provide the tools for making inductive assumptions, but they also reveal the need for such assumptions.

1. INTRODUCTION

This paper concerns inductive predictions. It takes these predictions as the result of inductive methods. The input of an inductive method includes a data set, usually consisting of observations, and possibly some further assumptions. The output may consist of predictions or general statements, where predictions concern unobserved singular states of affairs, and general statements, such as empirical generalizations, concern universal states of affairs. For example, from the fact that some internet startup has had decreasing stock price on all days until now, we may derive that the next day it will have decreasing stock price as well. This is a prediction about a single event, namely the decrease of stock price on the next day, based on data of the stock price movements on all days until now. From the same data set we may also derive that the internet startup will have a decreasing stock price on all future days, which is a general statement about events.

The inductive methods in this paper employ general statements to arrive at predictions. In inductive methods of this form, the data are first reflected in an opinion over a specific set of general statements, called a partition of hypotheses. For example, from the data on decreasing stock price we first derive an opinion over some partition of hypotheses on the state and nature of the internet startup. The predictions on the internet startup are subsequently derived from this opinion and the data. Since

predictions and hypotheses have a content that exceeds the content of the data, neither of these can be derived from the data with certainty. As opposed to deductive arguments, inductive methods therefore render the conclusions uncertain. This uncertainty can be expressed in a probability function. In sum, this paper concerns probabilistic inductive methods that employ hypotheses for making predictions. I will say that such predictions are based on hypotheses schemes, or alternatively, based on partitions.

The main line of the paper is the following. I first show that the use of partitions enables us to describe predictions typical for hasty generalizations. These predictions can be generated by choosing a specific partition used in the hypotheses scheme. This example triggers two different discussions, one on the function of partitions in the hypotheses schemes, and one on hypotheses schemes in relation to the problem of induction. The main conclusion of the first is that partitions are tools for making inductive assumptions. They determine which patterns are identified in the data and projected onto future observations. The main conclusion of the second discussion is that hypotheses schemes do not suggest anything towards solving the problem of induction. However, the schemes direct attention to choosing partitions as key element in the inductive method. I argue that they are a first step in a logic of induction that makes explicit the input of both data and projectability assumptions.

The plan of the paper is as follows. In Section 2, the present treatment of inductive predictions is related to a dominant tradition in formalizing inductive predictions, called Carnap–Hintikka inductive logic. The inductive predictions of this paper are seen to expand this tradition. Section 3 deals with the formal details of inductive predictions based on hypotheses schemes, and ends with some further remarks on such schemes. Section 4 considers two prediction rules working on the same data, but based on different partitions, and shows that they result in different predictions. It further elaborates the relation between inductive predictions and the Carnap–Hintikka tradition. In Section 5 and 6 the results are given a further philosophical interpretation, and related to the problem of induction. The conclusion summarizes the results.

2. CARNAP–HINTIKKA INDUCTIVE LOGIC

This section discusses the Carnap–Hintikka tradition of inductive logic. It emphasizes two characteristic features of this tradition: its focus on exchangeable predictions, and its suspicion towards general statements. The inductive predictions of this paper extend the Carnap–Hintikka tradition with respect to these two features.

As indicated, the inductive methods of this paper relate a data set, and possibly some other assumptions, with probabilistic predictions. The data of this paper are records of observations, encoded in natural numbers q_i, indexed with time, and collected in ordered t-tuples $e_t = \langle q_1, q_2, \ldots, q_t \rangle$. At any time t, the probability of the next observation q_{t+1} follows from this data set e_t and some further assumptions. Let us say that these further assumptions can be encoded in some collection of parameters, so that we can define inductive prediction rules as functions expressing the probability of the next observation q_{t+1} in terms of the data e_t and these parameters. This enables us to study inductive methods for making predictions by designing and comparing classes of such functions, which I call inductive prediction rules.

Carnap was the first to study inductive prediction rules at length. An exemplary class of probabilistic inductive inference rules for making predictions is his so-called $\gamma\lambda$ continuum, as elaborated in Carnap (1950, 1952), and Carnap and Stegmüller (1959):

$$C_{\gamma\lambda}(e_t, q) = \left(\frac{t}{t+\lambda} \right) \frac{t_q}{t} + \left(\frac{\lambda}{t+\lambda} \right) \gamma_q. \qquad (1)$$

The function C, the probability for observing q at time $t+1$, is a weighted average of the observed relative frequency t_q/t of instances of q among the ordered set of known observations e_t, and the preconceived or virtual relative frequency of observing q, denoted γ_q. The weights depend on the time t and a learning rate λ. With increasing time, the weighted average moves from the preconceived to the observed relative frequency. The learning rate λ determines the speed of this transition.

After Carnap, inductive prediction rules have been studied extensively. Axiomatizations, elaborations and synthetizations of inductive prediction rules have been developed by Kemeny (1963), Hintikka (1966), Carnap and Jeffrey (1971), Stegmüller (1973), Hintikka and Niiniluoto (1976), Kuipers (1978), Costantini (1979), Festa (1993) and Kuipers (1997). To this research tradition I refer with the names of Carnap and Hintikka.

Most of the work in this tradition concerns exchangeable prediction rules. Exchangeability of a prediction rule means that the predictions do not depend on the order of the incoming observations. As I elaborate below, exchangeable rules typically apply to settings in which the events producing the observations are independent. Exchangeable rules thus have a very wide range of application. Moreover, on the assumption that the prediction rule is exchangeable, it can be proved that the predictions eventually converge to optimal values. That is, if the observations are produced by some process with constant objective chances, the predictions of an

exchangeable rule will, according to Gaifman and Snir (1987), almost always converge on these chances, whatever the further initial assumptions. Both for their range of applicability and for this convergence property, exchangeable rules are a main focus in the Carnap–Hintikka tradition.

The second feature that I want to emphasize can only be made explicit if we establish the link, in both directions, between exchangeability of observations and the independence of the events that are supposed to produce these observations. The first component of this link is the assumption that the events producing the observations are part of some underlying process. The second component is the fact that if this underlying process generates the events with constant objective chances, then the chance of an event is independent from events occurring before or after it, so that the events can be called independent.

The link from exchangeability to independence is then established by the representation theorem of De Finetti, as discussed in (1964). This theorem shows that any exchangeable prediction rule can be represented uniquely as a Bayesian update over the partition of hypotheses that concern processes with constant chances. Section 4 deals with this representation theorem in some more detail. The link from independence to exchangeability, on the other hand, is established by the fact that any Bayesian update over a partition of hypotheses on constant chance processes results in an exchangeable prediction rule. This is seen most easily from the fact that the influence of observations on the probability over the hypotheses are commutative operations. The order of such updates is therefore inessential to the resulting probability assignment over the hypotheses, and thus inessential to the predictions resulting from this assignment. Again Section 4 shows this in more detail. In sum, assuming the independence of the events producing the observations can be equated with the use of exchangeable prediction rules.

The above leads up to the second characteristic feature of the Carnap–Hintikka tradition, which is connected to its empiricist roots. De Finetti interpreted the representation theorem as a reason to leave out the reference to underlying processes, and to concentrate on exchangeable prediction rules instead. As Hintikka (1970) argues, this is not so much because of a subjectivist suspicion towards objective chances, but mainly because these chance processes are described with universal statements, which cannot be decided with finite data. De Finetti deemed such universal statements suspect for empiricist reasons. For similar reasons, Carnap maintained that universal statements have measure zero. For both De Finetti and Carnap, the representation theorem showed that it is simply unnecessary to employ chance processes: we can obtain the same results using the exchangeability

of the prediction rule. In line with this, most of the Carnap–Hintikka tradition focuses on the properties of prediction rules, such as exchangeability or partial exchangeability, and avoids reference to the chance processes underlying these prediction rules.

Prediction rules with this feature I call Carnapian. This terminology signals that this second feature is not fully applicable to the Hintikka part of the Carnap–Hintikka tradition. In Hintikka (1966) and Tuomela (1966) we find a different attitude towards underlying chance processes, or at least towards the use of universal statements in inductive logic. More in particular, Hintikka employs universal generalizations on observations in the construction of his $\alpha\lambda$ continuum of inductive prediction rules. Tuomela discusses universal statements on ordered universes, and refers to Hintikka for the construction of prediction rules based on these universal statements. Both these authors thus employ universal statements to inform predictions in a specific way.

While this already presents a valuable extension, I feel that universal statements have not been employed with full force in the Carnap–Hintikka tradition. Perhaps some empiricist feelings have remained, which have curbed the further development of Hintikka systems. The $\alpha\lambda$ continuum offers little room for varying the kind of universal statements: the continuum concerns universal generalizations only, and the role of these generalizations is controlled completely by the value of the parameter α. As Hintikka himself remarks in (1997), it would be more convenient if the universal statements can simply be expressed by premises, so that other kinds of universal statements can be employed too, and also controlled more naturally. Related to this, many prediction rules in which the use of specific universal statements seems very natural do not employ such statements in their construction. For example, in the inductive prediction rules for Markov chains by Kuipers (1988) and Skyrms (1991), and the prediction rules describing analogy reasoning by Niiniluoto (1981) and Kuipers (1984), the construction of prediction rules is based on particular properties of the prediction rules. Underlying chance processes are not really used in the construction.

With this introduction, I can make precise the innovations that this paper offers. It extends the Carnap–Hintikka tradition in inductive logic in two ways, connected to the two characteristic features noted above. First, it advocates an almost unrestricted use of chance processes. Below I explicitly employ hypotheses, associated with chance processes, to define prediction rules. Second, this paper proposes a prediction rule that is not exchangeable, by adding hypotheses concerning a particular deterministic pattern to an existing partition of constant chance hypotheses. The claims

deriving from this are that partitions are a tool in making assumptions on patterns in data, and furthermore, that the use of this tool does not suggest specific preferred rules, nor restrict the class of prediction rules in any way. On the contrary, the tool widens the range of interesting inductive prediction rules.

The last two paragraphs of this section disclaim some topics that otherwise complicate the discussion too much. First, it can be noted that the prediction rules of this paper are somewhat similar to those of the paper by Tuomela on ordered universes. Both focus on predictions based on these specific patterns in the data. But for lack of space, I will not elaborate on this similarity in the following. Second, I will not discuss representation theorems like De Finetti's in full generality, and similarly I will not touch upon the various brands of partial exchangeability. The focus of this paper is on a particular non-exchangeable prediction rule, generated by a partition of hypotheses concerning particular chance processes, and on the moral that derives from the use of such partitions.

Finally, I do not discuss issues on the use and interpretation of probability. Inductive inferences concern degrees of belief, and are usually associated with epistemic or subjective probability. If, for instance, on the basis of data I assign an epistemic probability of 0.9 to the event that some internet startup has decreasing stock price on the next day, this means that I consider it likely that the stock price decreases the next day, and not necessarily that there is a tendency in the startup itself to have decreasing stock price. However, the present paper also involves explicit reference to chance processes, in which the probabilities are objective. For example, I may assign an epistemic probability of 0.9 to the hypothesis that the objective probability for any internet startup to have increasing stock price is smaller than 0.5. In the following, I assume that both objective and epistemic probability can be given an unproblematic interpretation in such a setting (Cf. Jeffrey (1977)). In the present context, I cannot resolve the tensions that may result from the simultaneous usage of these interpretations.

3. PREDICTIONS USING HYPOTHESES SCHEMES

This section introduces Bayesian hypotheses schemes. Bayesianism, as presented by Jeffrey (1984), Howson and Urbach (1989), and Earman (1992), is a dominant position in the philosophical discussion concerning probabilistic inferences. While the Carnapian prediction rules directly relate the data to a prediction, Bayesian updating can also relate the data with general statements or hypotheses. If the data are observations of trading days of an internet startup, a hypothesis can for instance be that all

trading days result in decreasing stock price, or that the portion of trading days with decreasing stock price has some value θ. The following elaborates the use of Bayesian updating over hypotheses in generating inductive predictions.

3.1. *Hypotheses Schemes*

In this subsection I define, in that order, set theoretical notions of observation and observational hypotheses, belief states over observations and hypotheses, Bayesian updating as a way of adapting belief states to new observations, partitions as specific sets of hypotheses, and predictions based on Bayesian updating over such partitions. The construction in which a partition of hypotheses is used for predictions is called the hypotheses scheme.

Let us define a space of possible observations K. As the simplest example of this, take $K = \{0, 1\}$, so that observations can be represented as bits $q_i \in K$. We can think of such observations as simple facts about some system, e.g. that a particular internet startup has decreasing or increasing stock price on day i. As in the above, finite sequences of observations can be written as $e_t = \langle q_1, q_2, \ldots, q_t \rangle$, so that $e_t \in K^t$, the t-th Cartesian power of K. The result of a particular observation i in such a finite sequence is denoted $e_t(i) = q_i$. In all of the following, the lower case letters e and q refer to such single numbers and finite sequences of numbers respectively. They constitute a finite observational language.

Now consider the space of infinite sequences of possible observations. Analogous to the above, we can define infinite sequences of results $e \in K^\omega$ as infinite ordered sequences $e = q_1 q_2 \ldots$ We can again define the separate observations $e(i) = q_i$. But we can also define the finite sequences e_t in terms of these infinite sequences of observations. Define the set $E_t \subset K^\omega$ as the set of all infinite sequences e that start with the observations E_t and diverge after that:

$$E_{t,\langle q_1, q_2, \ldots, q_t \rangle} = \{ e \in K^\omega \mid \forall i \le t : e(i) = q_i \}. \tag{2}$$

For later purposes we also define $Q_{t+1} \subset K^\omega$, the set of strings e that start with any $E_t \in K^t$, have the same observation q at time $t + 1$, and diverge again after that:

$$Q_{t+1,q} = \{ e \in K^\omega \mid e(t + 1) = q \}. \tag{3}$$

The sets E_t and Q_{t+1} are determined by the sequence of numbers $\langle q_1, q_2, \ldots, q_t \rangle$ and the single number q respectively. However, for sake of brevity I do not normally attach these numbers when I refer to the sets.

The upper case letters E_t and Q_{t+1} thus refer to specific sets in K^ω as well. Together these sets constitute a so-called cylindrical algebra, which I call the observation field $\mathcal{Q}_0$. It may be taken as the set theoretical equivalent of the finite observational language.

We can also define hypotheses as sets of infinite sequences of observations in K^ω. It is because they are defined on K^ω that we can call these hypotheses observational. Let $I_h : K^\omega \mapsto \{0, 1\}$ be the characteristic function of some hypothesis h applying to infinite sequences e. That is, $I_h(e) = 1$ if and only if the proposition h is true of the infinite sequence e, and $I_h(e) = 0$ otherwise. We can then define:

$$H = \{e \in K^\omega \mid I_h(e) = 1\}. \tag{4}$$

In the context of this paper, I consider hypotheses that are defined by reference to the infinite sequences e only. That is, I consider hypotheses that are verifiable only on the basis of an infinite sequence of observations e. Such hypotheses are, as Kelly (1997) calls them, gradually refutable or gradually verifiable. Following Billingsley (1995), they can also be called tail events in the observation field $\mathcal{Q}$, which contains all observations Q_i, and further all infinite intersections $e = Q_1 \cap Q_2 \cap \ldots$ This extended observation field $\mathcal{Q}$ is also called the σ field generated by $\mathcal{Q}_0$, denoted $\sigma(\mathcal{Q}_0)$. It is the set theoretical equivalent of an infinite observational language.

To represent belief states over these observations and hypotheses, define a collection of probability functions $p_{[e_t]}$ on the extended observation field $\mathcal{Q}$:

$$p_{[e_t]} : \mathcal{Q} \mapsto [0, 1]. \tag{5}$$

These probability functions are defined over all sets in $\mathcal{Q}$, and obey the standard Kolmogorov axioms of probability. They represent the belief states of an observer concerning both observations and hypotheses, who is given a knowledge base, or data set, e_t. It must be noted that every different knowledge base is connected to a unique belief state. Note finally that these probabilities are defined over sets in $\mathcal{Q}$. Therefore, when referring to the probability of hypotheses or observations at $t + 1$, I refer not to the propositions h or the numerical values q_{t+1}, but rather to the sets H and Q_{t+1}.

Bayes' rule can be defined as a way of relating these separate belief states, or more precisely, as a recursive relation over these probability functions:

$$p_{[e_{i+1}]}(\cdot) = p_{[e_i]}(\cdot \mid Q_{i+1}). \tag{6}$$

This rule applies to all sets in $\mathcal{Q}$, and thus to observations and hypotheses equally. Bayes' rule dictates in what manner the probability of a set must

be updated when some new observation is added to the data set. In particular, the new probability is equal to the old probability conditional on this observation. In this paper, the main use of Bayes' rule is in adapting the probability assignments of the hypotheses for incoming observations. Taking hypothesis H as the argument, the conditional probability on the right hand side of Equation (6) is defined as

$$p_{[e_i]}(H|Q_{i+1}) = p_{[e_i]}(H)\frac{p_{[e_i]}(Q_{i+1}|H)}{p_{[e_i]}(Q_{i+1})} \tag{7}$$

So to compute the probability for some H relative to a data set e_{i+1}, we need the preceding probability assignment $p_{[e_i]}(H)$, the observation q_{i+1} by means of which we can pick out the set Q_{i+1}, and the probabilities $p_{[e_i]}(Q_{i+1}|H)$ and $p_{[e_i]}(Q_{i+1})$ belonging to that set. The probabilities of observations conditional on a hypothesis, $p_{[e_i]}(Q_{i+1}|H)$, are called the likelihoods of the observations. The function from i to $p_{[e_i]}(Q_{i+1}|H)$ is called the likelihood function of H.

Note that the other probability of the observation Q_{i+1} in (7) is the prediction $p_{[e_i]}(Q_{i+1})$, the probability of the next observation Q_{i+1} given the data E_i. In an update procedure aimed at making predictions, this cannot be part of the required input. Fortunately we can avoid this by employing a partition of hypotheses. From now on, I focus on such partitions instead of on separate hypotheses H. A partition $\mathcal{P}$ is a set of hypotheses $\{H_0, H_1, \ldots, H_N\}$, all of them defined as in Equation (4), such that

$$\forall e \in K^\omega : \sum_{j=0}^{N} I_{h_j}(e) = 1. \tag{8}$$

Recall that $I_{h_j} : K^\omega \mapsto \{0, 1\}$, so that with definition (8), every e is included in precisely one of the $H_j \in \mathcal{P}$. The hypotheses in a partition are thus mutually exclusive and jointly exhaustive. We can write for the prediction:

$$p_{[e_i]}(Q_{i+1}) = \sum_{j=0}^{N} p_{[e_i]}(H_j)p_{[e_i]}(Q_{i+1}|H_j). \tag{9}$$

At every time i, the predictions are written as a function of the probability assignments over the partition, $p_{[e_i]}(H_j)$ for all j, and the likelihood functions for all the hypotheses, that is, the probabilities of the observation Q_{i+1} conditional on these hypotheses, $p_{[e_i]}(Q_{i+1}|H_j)$.

We are now in a position to define predictions based on a hypotheses scheme. First we compute the probability assignments over the

H_j in the partition $\mathscr{P}$ in terms of the priors, the likelihoods and the data $e_t = \langle q_1, q_2, \ldots, q_t \rangle$ by repeated application of (6), (7) and (9):

$$p_{[e_t]}(H_j) = p_{[e_0]}(H_j) \prod_{i=0}^{t-1} \frac{p_{[e_i]}(Q_{i+1}|H_j)}{\sum_{j=0}^{N} p_{[e_i]}(H_j) p_{[e_i]}(Q_{i+1}|H_j)}. \quad (10)$$

Having thus obtained $p_{[e_t]}(H_j)$, we can compute the predictions $p_{[e_t]}(Q_{t+1})$ by yet another application of Equation (9). The prediction of Q_{t+1} then takes as input the probabilities $p_{[e_0]}(H_j)$ for every $j \leq N$, called the priors, and the likelihoods $p_{[e_i]}(Q_{i+1}|H_j)$ for every $j \leq N$ and $0 < i \leq t$. The belief attached to the hypotheses thus functions as an intermediate state in determining the predictions.

3.2. *Elaborating the Schemes*

The definition of predictions using hypotheses schemes is now complete. In the remainder of this section, I further elaborate these schemes. I consider the relation between the partition on the one hand, and the priors and likelihoods for the hypotheses in the partition on the other. The definition of the hypotheses suggests natural restrictions on the likelihoods, which are provided in the next two paragraphs. After that, I briefly discuss the relation between the Carnap–Hintikka tradition and the hypotheses schemes. Finally, I deal with the possibility of a continuum of hypotheses. The above defines Bayesian updating over finite partitions, but the extension to a continuum of hypotheses is relatively easy. This extension prepares for Section 4.

First, let us consider how the partition restricts priors and likelihoods. For priors, the normalization of $p_{[e_0]}$ over the hypotheses is the only restriction:

$$\sum_{j=0}^{N} p_{[e_0]}(H_j) = 1, \quad (11)$$

Much more can be said on the relation between likelihoods and hypotheses. Clearly the likelihoods are restricted by the axioms of probability and the definition of conditional probability:

$$p_{[e_i]}(Q_{i+1}|H_j) = \frac{p_{[e_i]}(Q_{i+1} \cap H_j)}{p_{[e_i]}(H_j)}. \quad (12)$$

Whenever a hypothesis h_j deems some observation q_{i+1} to be either impossible or positively certain, this carries over to the likelihoods:

(13) $Q_{i+1} \cap H_j = \emptyset$ $\Rightarrow$ $p_{[e_i]}(Q_{i+1}|H_j) = 0,$

(14) $Q_{i+1} \cap H_j = H_j$ $\Rightarrow$ $p_{[e_i]}(Q_{i+1}|H_j) = 1.$

These restrictions, which follow from deterministic aspects of the hypotheses, are indisputable.

In many cases, however, the hypotheses are not deterministic, and in those cases the link between the sets $H_j \in \mathcal{Q}$ and the probabilities of observations Q_{t+1} within these H_j is less straightforward. In the example on the internet startup, a hypothesis H_j may be that the relative frequency of trading days that have increasing stock price has some particular value $\theta_j \in [0, 1]$. One way of connecting this hypothesis with probability assignments for the observations is to adopt a frequentist interpretation of probability. The hypothesis is then identical to the statistical hypothesis that on any day the probability for an increasing stock price has this particular value $\theta_j \in [0, 1]$, so that

$$p_{[e_i]}(Q_{i+1,1}|H_j) = \theta_j. \tag{15}$$

Note that the set Q_{i+1} here has the further argument $q_{i+1} = 1$ to signify the set of infinite sequences e in which the $(i+1)$-th trading day has increasing stock price.

In this paper, I cannot deal with the relation between hypotheses and likelihoods any further than this. I am aware that the use of the frequentist interpretation is rather unusual in a Bayesian framework, which is typically associated with the subjective interpretation of probability. A complete picture of inductive Bayesian logic must include a further elaboration and explanation of the restriction (15). In the following, I simply employ the frequentist interpretation of probabilities to connect hypotheses and likelihoods, and I use the resulting restriction (15) uncritically.

A further remark concerns the relation between Carnapian prediction rules and the predictions based on hypotheses. Within the framework of Section 3.1, Carnapian prediction rules may be defined as a direct computation of predictions from data,

$$p_{[e_t]}(Q_{t+1,q}) = P(e_t, q) \tag{16}$$

in which P is a function of e_t and q ranging over the interval $[0, 1]$, as in expression (1). For these rules, the probabilities for next observations, $p_{[e_t]}(Q_{t+1})$, are determined directly, without using an intermediate probability assignment over hypotheses. We can now use Bayes' rule to trace back these assignments to a restriction on the prior probability over the field $\mathcal{Q}$:

$$p_{[e_0]}(E_t \cap Q_{t+1,q}) = P(e_t, q)\, p_{[e_0]}(E_t), \tag{17}$$

for all e_t and q_{t+1} with $t > 0$, where the sets E_t and Q_{t+1} are obviously associated with the actual observational results e_t and q. This restriction

amounts to a full specification of the prior probability assignment over the observation field $\mathcal{Q}$. A Carnapian prediction rule is thus a direct choice of a probability function over $\mathcal{Q}$.

A similar choice of $p_{[e_0]}$ is effected by choosing a partition, along with its prior and likelihoods, in a hypotheses scheme. That is, any hypotheses scheme comes down to some prediction rule, and any Carnapian prediction rule corresponds to some hypotheses scheme, most trivially to a scheme with only one hypothesis. It can further be noted that different hypotheses schemes may entail the same restriction over $\mathcal{Q}$, and thus the same prediction rule. However, it is very difficult to make general claims on the relation between hypotheses schemes and prediction rules. De Finetti's representation theorem is one of the exceptions. It states that prediction rules $P(e_t, q)$ that are exchangeable can always be replicated in a hypotheses scheme with hypotheses on constant chances. However, the salient point here is that any hypotheses scheme can be replicated with a Carnapian prediction rule, and vice versa.

Another remark concerns the number of hypotheses in a partition. In the above this number has been finite, but there is no reason to exclude partitions with an infinite number, or even a continuum, of hypotheses. Consider the continuous partition $\mathcal{P} = \{H_\theta\}_{\theta \in [0,1]}$, in which the index j is replaced by a variable θ. The probability assignments over the hypotheses $p_{[e_t]}(H_j)$ then turn into probability densities $p_{[e_t]}(H_\theta)d\theta$. Bayesian updating becomes an operation which transforms this probability density:

$$p_{[e_{i+1}]}(H_\theta)d\theta = \frac{p_{[e_i]}(Q_{i+1}|H_\theta)}{p_{[e_i]}(Q_{i+1})} p_{[e_i]}(H_\theta)d\theta. \tag{18}$$

Further, in the definition of a partition (8), the predictions (9) and the normalization condition (11), the summation must be replaced by an integration:

$$(19) \quad \forall e \in K^\omega : \int_0^1 I_{h_\theta}(e)\, d\theta = 1,$$

$$(20) \qquad\qquad p_{[e_i]}(Q_{i+1}) = \int_0^1 p_{[e_i]}(H_\theta) p_{[e_i]}(Q_{i+1}|H_\theta)\, d\theta,$$

$$(21) \qquad\qquad \int_0^1 p_{[e_0]}(H_\theta)\, d\theta = 1.$$

In all other expressions, the index j must be replaced by the variable θ. Apart from that, there are no further changes to the update machinery, and the remarks on the hypotheses schemes remain equally valid.

Some remarks complete this section. First, note that until now I have only considered the likelihoods of Q_{i+1} at time i, that is, the assignments

$p_{[e_i]}(Q_{i+1}|H_j)$. It is more in line with the interpretation of $p_{[e_i]}$ as the belief state at time i to fix only the likelihoods $p_{[e_0]}(Q_{i+1}|H_j)$, which express the opinion at the start of the update, and to update the likelihoods according to

$$p_{[e_{i'}]}(Q_{i+1}|H_j) = p_{[e_0]}(Q_{i+1}|H_j \cap E_{i'}) \tag{22}$$

for all $i' \leq i$. For hypotheses on constant chances we can choose that

$$p_{[e_0]}(Q_{i+1}|H_j \cap E_{i'}) = p_{[e_0]}(Q_{i+1}|H_j), \tag{23}$$

but depending on the hypothesis, the update operation may also change the likelihoods.

Further, many arguments have been proposed to the effect that Bayesian updating, or updating by strict conditioning, is the only consistent way of updating probabilities. This position is criticised for being too restrictive, as for example in Bacchus, Kyburg and Thalos (1990). However, the present paper works completely within the Bayesian framework, and argues that the restrictive character of strict conditioning is overrated. Related to this, it must be noted that the hypotheses schemes allow for a wide variety of possible partitions $\mathscr{P}$, none of which is excluded in principle. In the next sections, the freedom in choosing partitions will turn out that because of this freedom, the predictions that can be generated by Bayesian updating over hypotheses are completely unrestricted.

4. EXAMPLES ON CRASH DATA

The above introduces a general framework for using hypotheses and Bayesian updating in making predictions. This section gives two applications of the framework. The first application employs hypotheses on constant chances for the observations, resulting in the prediction rules of Carnap's $\gamma\lambda$ continuum. This also serves as an illustration of the representation theorem of de Finetti. The second application provides an extension of the Carnap–Hintikka tradition. Apart from the hypotheses on constant chances, it employs hypotheses concerning a particular pattern in the data. The resulting predictions are not covered by the $\gamma\lambda$ continuum, and they are not in general exchangeable.

4.1. *Statistical Partition*

The example concerns stock price movements. Consider the following strings of data, representing stock prices for $t = 35$ days. In the data, q_i equals 0 if the stock price decreased over day i, and q_i equals 1 if the

stock price increased or remained unchanged over that day. Here are two possible histories of the prices of a stock of some internet startup:

$$e_{35} = 0100010000001011000001000000010000,$$
$$e_{35}^* = 0100101011101000000000000000000000.$$

Note that e_{35} and e_{35}^* have an equal number of trading days i with $q_i = 1$, but that the order of increase and decrease is different for both strings. In particular, e_{35}^* shows what can be called a crash: from some day onwards we only observe decreasing stock price.

Now imagine a marketeer who aims to predict stock price movements based on observed price movements on foregoing trading days. Further, assume that she employs some partition $\mathcal{C}$ with a continuum of hypotheses to specify her predictions. To characterize the hypotheses, define

$$\text{frq}(e) = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} e(i). \tag{24}$$

For any infinitely long sequence of days e, the function $\text{frq}(e)$ gives the ratio of trading days i for which $e(i) = 1$. Note that $\text{frq}(e)$ is undefined for some of the $e \in K^\omega$. Now define I_{h_θ} as follows:

$$(25) \quad I_{h_\theta}(e) = \begin{cases} 1 & \text{if } \text{frq}(e) = \theta, \\ 0 & \text{otherwise,} \end{cases}$$

in which $\theta \in [0, 1]$. Further define $I_{h_{\neg\theta}} = 1$ if $\text{frq}(e)$ is not defined, and $I_{h_{\neg\theta}} = 0$ otherwise. It then follows that

$$\forall e : I_{h_{\neg\theta}}(e) + \int_0^1 I_{h_\theta}(e) \, d\theta = 1, \tag{26}$$

so that $\mathcal{C} = \{H_{\neg\theta}\} \cup \{H_\theta | \theta \in [0, 1]\}$ is a partition including a continuum of hypotheses on relative frequencies. I call $\mathcal{C}$ the Carnapian partition.

Assume that the marketeer employs the following input probabilities:

$$(27) \qquad\qquad p_{[e_0]}(H_\theta) d\theta = 1,$$
$$(28) \qquad\qquad p_{[e_0]}(H_{\neg\theta}) = 0,$$
$$(29) \quad \forall i > 0 : \quad p_{[e_0]}(Q_{i+1,Q} | H_\theta) = \begin{cases} \theta & \text{if } Q = 1, \\ 1 - \theta & \text{if } Q = 0. \end{cases}$$

where again $\theta \in [0, 1]$. Equation (27) states that the probability density over the hypotheses H_θ is uniform. This may be motivated with an appeal to the principle of indifference or some other symmetry principle. Equation
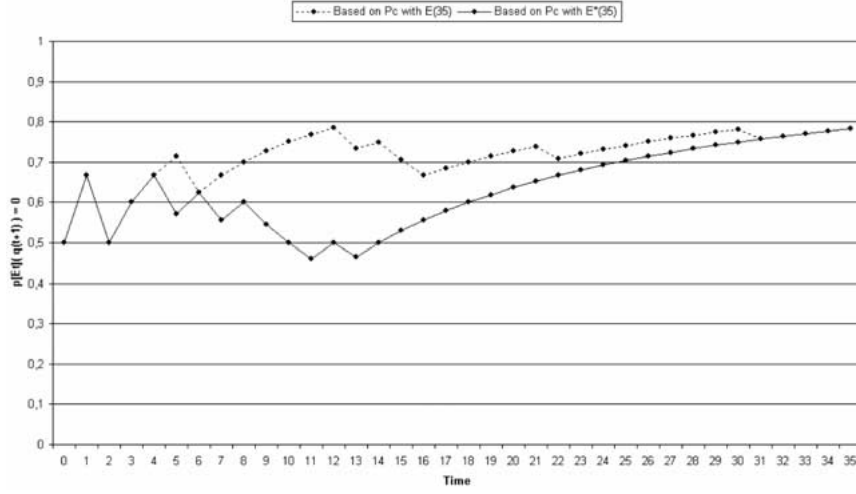
Figure 1. Predictions $p_{[e_t]}(Q_{t+1})$ against time t, based on the partition $\mathcal{C}$. The dotted line shows the predictions for the normal data e_{35}, the unbroken line shows the predictions for the crash data e_{35}^*.

(28) states that those sequences e in which the frequency of trading days with $e(i) = 1$ has no limit are almost impossible. This assumption is not compulsory. It is here made for computational simplicity, as it allows us to ignore hypothesis $H_{\neg\theta}$ in further calculations. Also, it is required if we want to illustrate the representation theorem.

Equation (29) can be motivated with the restriction on likelihoods as given in expression (15). Assume that at every i, the Bayesian agent updates the likelihood that is used in the next prediction and subsequent update to

$$p_{[e_i]}(Q_{i+1}|H_\theta) = p_{[e_0]}(Q_{i+1}|H_\theta), \tag{30}$$

so that, in conformity with the definition of the hypotheses H_θ, the accumulation of data e_i does not change the original likelihoods. The hypotheses H_θ on relative frequencies then have constant likelihoods $p_{[e_i]}(Q_{i+1}|H_\theta) = \theta$.

We have specified all probabilities that are needed for generating predictions. Using the values of the priors $p_{[e_0]}(H_\theta)d\theta$ for all θ, and of the likelihoods $p_{[e_i]}(Q_{i+1}|H_\theta)$ for all θ and $i \geq 0$, we can compute the predictions on next observations $p_{[e_t]}(Q_{t+1})$ that the Bayesian agent will make when confronted with the observations e_{35} and e_{35}^* respectively. I have calculated these predictions on a computer, and depicted them in figure 1. In the remainder of this subsection I make some remarks on these predictions, and on the hypotheses scheme with the statistical partition in general.

First of all, the above hypotheses scheme illustrates the representation theorem of de Finetti. The hypotheses H_θ are exactly the hypotheses on processes with constant chances, which I alluded to in sections 2 and 3. The representation theorem is that any exchangeable prediction rule $P(e_t, Q)$ can be represented as a Bayesian update over the partition $\mathcal{C}$, given the above restrictions (28) and (29). Different exchangeable prediction rules are defined by choosing different priors $p_{[e_0]}$ over the hypotheses H_θ. For example, choosing a member from a particular family of so-called Dirichlet distributions for $p_{[e_0]}(H_\theta)d\theta$ results in a prediction rule from the $\gamma\lambda$ continuum of Carnap, as given in expression (1). As described in Festa (1993), the parameters of the Dirichlet distribution thereby fix the values of γ and λ. More in particular, choosing the uniform prior of Equation (27) results in rule (1) with parameters $\lambda = 2$ and $\gamma_0 = \gamma_1 = 1/2$. Note however that the range of the representation theorem is much wider than the specific equivalence of the $\gamma\lambda$ continuum and the Dirichlet distributions over $\mathcal{C}$.

Section 2 indicated that the representation theorem was taken as an opportunity to dispose of hypotheses schemes using $\mathcal{C}$, in favour of exchangeable prediction rules. One reason was that the hypotheses schemes committed to the assumption of underlying chance processes and the assignment of probability to universal statements. Another, more immediate reason for not using the hypotheses schemes may have been that they are unnecessarily roundabout. In the above and in the following, however, I explicitly use the hypotheses schemes to design and study predictions. Even while the predictions based on $\mathcal{C}$ may always be generated with a simpler exchangeable prediction rule, I explicitly employ hypotheses in the construction. In Section 5 I argue that there are independent reasons for doing so.

Finally, it can be noted that after 32 days the predictions are the same for both strings of data. This shows the exchangeability of the above Bayesian update procedure. Probability assignments after any e_t are invariant under the permutation of results $e_t(i)$ within that e_t, and as said, e_{35} and e_{35}^* have the same number of 1's. For both e_{35} and e_{35}^* it is further notable that the predictions $p_{[e_i]}(Q_{i+1,0})$ converge to 1. The speed of convergence, however, decreases with the addition of further instances of $e_t(i) = 0$, or more precisely, the second derivative to time of the predictions, taken as a function over time, is negative. For this reason, the predictions depicted in figure 1 do not accommodate the fact that the data e_{35}^* may be the result of a crash.

4.2. *Crash Hypotheses*

Figure 1 shows the inductive predictions of a marketeer who is not sensitive to the possibility of a crash. In the example below, I alter the hypotheses scheme in such a way that this sensitivity is modelled. This is done by adding hypotheses to the statistical partition, thereby implicitly altering the resulting prediction rule. In particular, I add the hypotheses $g^q_{\gamma\lambda\tau}$ to the statistical partition, the meaning of which can be phrased as: until trading day τ, stock price behaves like the Carnapian $\gamma\lambda$ rule says, but from trading day τ onwards, all stock price movements are q.

Let us denote the partition consisting of the statistical hypotheses h_θ and the crash hypotheses $g^q_{\gamma\lambda\tau}$ with $\mathcal{G}$. The crash hypotheses can be associated with sets $G^q_{\gamma\lambda\tau}$ in $\mathcal{Q}$ using a characteristic function to select for crashes:

$$(31) \quad I_{g^q_{\gamma\lambda\tau}}(e) = \begin{cases} 1 & \text{if } e(\tau) \neq q \wedge \forall i > \tau : e(i) = q, \\ 0 & \text{otherwise}, \end{cases}$$

$$(32) \quad G^q_{\gamma\lambda\tau} = \{e \in K^\omega \mid I_{g^q_{\gamma\lambda\tau}}(e) = 1\}.$$

Note that the γ and λ do not occur in the definition of the sets $G^q_{\gamma\lambda\tau}$. The sets can be defined solely on the basis of the crash starting at time τ.

The hypotheses $G^q_{\gamma\lambda\tau}$ can be given likelihoods that reflect the above meaning:

$$p_{[e_0]}(Q_{i+1,q'}|G^q_{\gamma\lambda\tau} \cap E_i) = \begin{cases} \frac{i_{q'}+\lambda\gamma_{q'}}{i+\lambda} & \text{if } t < \tau, \\ 1 & \text{if } i = \tau, q' \neq q, \text{ or } i > \tau, q' = q, \\ 0 & \text{if } i = \tau, q' = q, \text{ or } i > \tau, q' \neq q, \end{cases}$$

(33)

where i_q denotes the number of results q' in the observations e_i. The last two clauses of the likelihood definition are motivated with definition (31) and restriction (13). However, as there is no restriction on the first $\tau - 1$ observations in the sets $G^q_{\gamma\lambda\tau}$, there is no restriction motivating the first clause. The likelihoods before τ may be chosen in accordance with the predictions generated by the partition $\mathcal{C}$, so that, when the hypotheses $G^q_{\gamma\lambda\tau}$ are added to that partition, they only distort the predictions insofar as there is a crash pattern in the data. Because of this choice, the first clause in the likelihood definition depends on the actual data e_i. This means that with every new observation before τ, updating the likelihoods according to (22) changes these likelihoods.

Finally, the hypotheses $G^q_{\gamma\lambda\tau}$ are given prior probabilities of the following form:

(34) $p_{[e_0]}(G^0_{\gamma\lambda\tau}) = \alpha\,(1-\delta)\,\delta^\tau,$

(35) $p_{[e_0]}(G^1_{\gamma\lambda\tau}) = 0,$

where $\tau > 0$ and $0 < \delta < 1$ so that $(1-\delta)\delta^\tau$ is a discount factor, which describes how a trader slowly grows less suspicious for crashes, and where $0 < \alpha < 1$. The factor α is the total probability that is assigned to all the crash hypotheses, that is, $\alpha = \sum_{\tau=1}^{\infty} p_{[e_0]}(G^0_{\gamma\lambda\tau})$. Note that because of (35), booming markets, in which from some time onwards prices only go up, are not reckoned with.

The probability mass $1 - \alpha$ can be divided over the remaining hypotheses from $\mathcal{C}$ according to

$$p_{[e_0]}(H_\theta)d\theta = 1 - \alpha, \qquad (36)$$

where in this case $\theta \in (0, 1]$. The likelihoods (33) of the crash hypotheses can be made to accord with this prior by setting $\gamma_q = 1/2$ for both values of q, and $\lambda = 2$. Note further that the hypothesis H_0 is now excluded from the subpartition $\mathcal{C}$. This is because for all $e \in G^0_{\gamma\lambda\tau}$, the relative frequency frq(e) is 0, so that $G^0_{\gamma\lambda\tau} \subset H_0$ for all τ. On the other hand, according to the original likelihoods in (29), the hypotheses $G^0_{\gamma\lambda\tau}$ have zero probability within H_0, because any observation of $q = 1$ is deemed impossible within it. The simplest solution to all this is to exclude the hypothesis H_0 from the partition altogether. Since hypothesis H_0 had a negligible measure in the original hypotheses scheme with $\mathcal{C}$, leaving it out of the combined partition $\mathcal{G}$ does not affect the predictions generated by the new update.

In sum, we have created a new partition $\mathcal{G}$, including both H_θ and $G^0_{\gamma\lambda\tau}$. As will be seen, updating over this partition generates predictions which express a sensitivity for crashes. Choosing values for α and δ determines to what extent this sensitivity influences the predictions. Admittedly, the partition $\mathcal{G}$ involves considerable idealizations, for example that crashes are taken to last forever and that the prior probability for a crash slowly diminishes. These idealizations are not compulsory: the hypotheses schemes offer space for further specifications and elaborations in these respects. In the following, however, I want to focus on the fundamental possibilities that the freedom in choosing partitions presents. The idealizations of $\mathcal{G}$, and the ways to avoid them, are not discussed in this paper.

Like for figure 1, we can calculate the predictions $p_{[e_t]}(Q_{t+1})$ using equations (6), (7) and (9). Figure 2 shows a comparison of two market-eers confronted with the crash data e^*_{35}. The diamond curve shows the
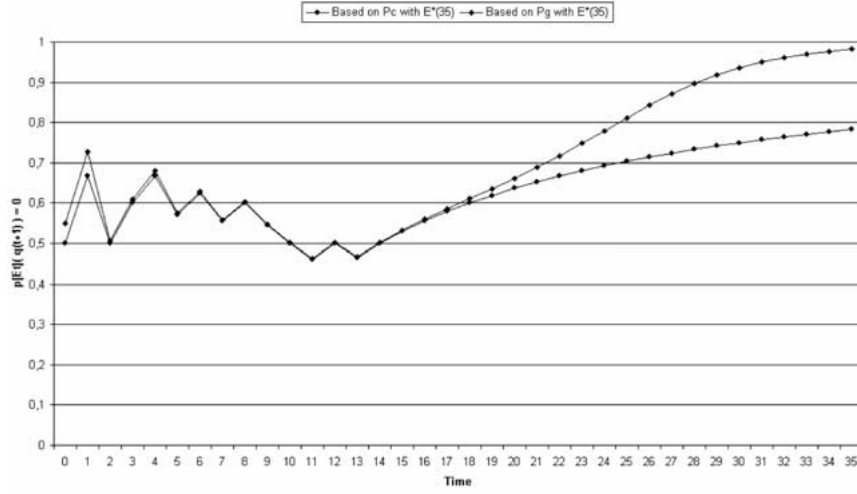
Figure 2. Predictions $p_{[e_t]}(Q_{t+1})$ against time t for the crash data e_{35}^*. The bulleted curve is based on the partition $\mathcal{C}$, the diamond curve is based on the partition $\mathcal{G}$.

predictions based on the use of the partition $\mathcal{G}$, and the bullet curve shows the predictions of the statistical partition $\mathcal{C}$. The hypotheses $G_{\gamma\lambda\tau}^0$ of this particular update have $\alpha = 0.5$ and $\delta = 0.8$. Note that the predictions based on $\mathcal{G}$ deviate from the predictions based on $\mathcal{C}$. As the unbroken string of $q_i = 0$ grows, the marketeer using $\mathcal{G}$ picks up on the apparent regularity and in subsequent days gives higher probability to the prediction that next days will show the result $q_{t+1} = 0$ as well. Further, note that the exchangeability of the observations within the data e_{35}^* is indeed violated with the use of the alternative partition $\mathcal{G}$. This is because the probability assignments depend directly on whether the data e_t show an unbroken string of 0's up until t. The partition $\mathcal{G}$ thus introduces a sensitivity for the occurrence of a crash pattern in the data, next to the usual attention that is given to the relative frequencies of 0's and 1's.

It must be stressed that using the partition $\mathcal{G}$ in no way violates the Bayesian framework developed in Section 3. First, strict conditioning on new observations, as expressed in Equation (6), invariably holds. In the above example, the probabilities $p_{[e_t]}(G_{\gamma\lambda\tau}^0)$ are adapted according to it just as well, causing them to be turned to zero at every time $i > \tau$ for which $q_i = 1$, or immediately if $i = \tau$ and $q_i = 0$. Second, it is not problematic to assign nonzero priors to hypotheses in the alternative partition which had negligible or zero probability in the original partition. Assigning nonzero probabilities to hypotheses on specific patterns has been proposed in a similar way by Jeffreys (1939). Hintikka systems (1966) also use nonzero probabilities for general hypotheses. More in general, as reviewed

in Howson (1973), many have argued against the contention that giving nonzero priors to generalizations is inconsistent. Finally, note that partitions cannot be altered during an update. Such a move is in disagreement with Bayesian updating.

Before drawing the morals of the above examples, let me briefly pause over the alteration itself in light of the preceding remarks on hypotheses schemes and prediction rules. It is clear that in altering a partition in the above way, I change more than just a prior probability assignment over hypotheses. Introducing new hypotheses into a partition changes the predictions. In particular, the Carnapian $C_{\langle 1/2, 1/2 \rangle 2}(e_t, q)$, which was generated by the partition $\mathcal{C}$ with uniform prior, becomes some other prediction rule $G_{\alpha\delta}(e_t, q)$. Put differently, the probability assignment $p_{[e_0]}$ over the field $\mathcal{Q}$, initially determined by the partition $\mathcal{C}$ and some prior over it, now encodes a different prediction rule, determined by the partition $\mathcal{G}$, an a prior. The added hypotheses of $\mathcal{G}$ focus on a crash pattern in the data, and the resulting predictions will therefore not in general be exchangeable. They depend on the occurrence of consecutive observations of $q = 0$. This also means that the convergence results alluded to in Section 2 are not applicable anymore. In sum, I have defined a different prediction rule by choosing a different partition in the hypotheses scheme.

5. PARTITIONS AS INDUCTIVE ASSUMPTIONS

Several conclusions may be drawn from the example with the partition $\mathcal{G}$. First, the example shows that inductive predictions based on hypotheses can be adapted to model pattern recognition, and in this particular case, hasty generalization. This can be done by adding hypotheses that pertain to the relevant kind of pattern. Following Putnam's critical remarks on the Carnap–Hintikka tradition in (1963a) and (1963b), this may already be a useful extension of that tradition. Secondly, and as I also discussed above, the modelling of hasty generalization may convince those who consider updating on generalizations impossible due to the negligible measure of these generalizations in the observation field.

Thirdly, the example may be taken to nuance the fact that Bayesian updating is not suitable for modelling ampliative reasoning, as is argued by van Fraassen (1989). It is true that Bayesian updating cannot capture reasoning that decides between hypotheses with the same observational content, which therefore have the same likelihoods in the hypotheses schemes. But the above reasoning can nevertheless be called ampliative on the level of predictions: hasty generalization is a typically ampliative inferential move. Note that the ampliativeness is then implicit in the choice

of the partition $\mathcal{G}$. Thus, even though Bayesian updating is itself not ampliative, the predictions resulting from a Bayesian update can in a sense model ampliative reasoning.

The main conclusion concerns the use of partitions in both examples, and the function of choosing a partition in general. Note that in the above examples, the influence of the observations is encoded entirely in the partition. We first determine a posterior probability over the partition, using the prior probability and the observations. The predictions are subsequently derived from this posterior probability and the likelihoods. These likelihoods are determined prior to the accumulation of observations, so the posterior probability over the partition is the only term in the predictions which depends on the observations. The partition mediates the influence that the observations exert on the predictions. As Niiniluoto (1976) puts it, a partition defines a closed question, which has a limited set of possible answers, for the observations to decide over. So partitions do not provide an impartial or completely general view on the observations. Rather they are a pair of glasses for looking at the observations in a particular way.

Let me characterize how partitions limit the view on observations. Recall the statistical partition $\mathcal{C}$. The posterior probability over this partition can be computed from the prior and the observations. However, we do not need to know all the details of the observations for this computation. In fact, it suffices to know a specific characteristic of the observations: for all q we must know the number of times that it occurred within the data e_t. These numbers were denoted as t_q in the above. They can be called the sufficient statistics for computing the probability over $\mathcal{C}$ at time t, and thus for generating the predictions based on $\mathcal{C}$. The statistics t_q express those characteristics of the observations which are taken to be relevant for the predictions. Note that the exchangeability of the predictions based on $\mathcal{C}$ follows from the fact that the sufficient statistics are independent of the order of observations. This can be seen easily from the equations determining the probability over the partition: none of the terms in the product of Equation (10) depends on the order of the q_i in e_t.

The partition with crash hypotheses $\mathcal{G}$ limits the view on the observations in a different way. As with the statistical partition, we can identify a set of sufficient statistics for it. This set includes not just the numbers t_q, but also the length of the time interval $[\tau, t]$ within which all results are 0. The numbers t_q and the number $t - \tau$ are employed together in a full determination of the probability over $\mathcal{G}$ at time t, and therefore in the generation of the predictions based on $\mathcal{G}$. It is notable that, because the value of $t - \tau$ depends on the order of the observations e_t, the resulting predictions are not exchangeable.

The above examples suggest how partitions limit the view on observations: partitions determine a set of sufficient statistics, and these statistics represent the characteristics of the observations which are taken to be relevant for further predictions. Put differently, by choosing a partition we focus on a particular set of patterns in the data, and by making predictions based on the partition we deem these patterns relevant to future observations. However, from the above it is not clear what the exact function of this limitation is, or more specifically, what the nature of this relevance is. As Skyrms suggests in (1996), the answer to this is that sufficient statistics determine the so-called projectable characteristics of data. The function of partitions then is that they determine the projectable characteristics of the observations. They are a tool in controlling the projectability assumptions that are used in inductive predictions.

Now let me explicate in general terms how the use of a partition relates to the assumption of a projectable pattern in the observations. Recall that the hypotheses in a partition are all associated with a likelihood function. These likelihood functions may be in accordance with the actual observations to differing degrees: hypotheses that have high overall likelihoods for the observations are said to fit the data better than those with low overall average likelihoods. An update over a partition can thus be viewed as a competition among the hypotheses in the partition, in which hypotheses that fit the observations best acquire most probability. Note further that the likelihood functions associated with the hypotheses describe probabilistic patterns in the observations. An update over a partition is thus also a competition between probabilistic patterns in the observations. Choosing a particular partition thus limits the range of possible patterns that are allowed to compete in the update.

Furthermore, if we go on to employ the results of such a competition for the generation of predictions, we implicitly assume that those probabilistic patterns that fitted the observations better in the past are more likely to perform better in the future as well. This is because predictions of future observations are mainly based on the hypotheses which, relative to the chosen partition, were most successful in predicting the past observations: those hypotheses gain more probability in the update. This is exactly where the assumption on the uniformity of nature, with respect to a specific set of probabilistic patterns, is introduced into the hypotheses scheme.

The above shows in what way the partitions are assumptions on the projectability of patterns in the observations: a partition determines a collection of probabilistic patterns, all of them patterns which may be employed for successful predictions, or projectable patterns for short. A prior probability over the hypotheses expresses how much these respective pat-

terns are trusted with the predictive task at the onset, but the observations eventually determine which patterns perform this task best on the actual data. The predictions are subsequently derived with a weighing factor, the probability over the partition, which favours the patterns that perform best. However, it must be stressed that the projectability assumption concerns not just these best performing patterns, but the partition as a whole, because the patterns perform better or worse only relative to a collection of patterns. The projectability assumptions are therefore implicit in the common features of the hypotheses involved. Limiting the collection of patterns to a collection with some general feature comes down to the assumption that the observations themselves exhibit this general feature, and that this general feature can therefore be projected onto future observations.

Finally, let me illustrate the projectability assumptions as general characteristics of the partitions, and link them with the sufficient statistics alluded to above. Consider once again the examples of Section 4. Choosing the statistical partition $\mathcal{C}$ means that we limit the possible probabilistic patterns to those for which the observations occur with specific relative frequencies. The projectability assumption is therefore exactly that this characteristic of the observations, namely the relative frequencies, are in fact exhibited in the observations. This is quite naturally related to the sufficient statistics for this partition, which are the observed relative frequencies t_q. Similarly, choosing to include hypotheses on crashes means that we include this particular set of crash patterns in the set of possible patterns. The projectability assumption is therefore exactly that this characteristic of a crash may be exhibited in the observations too. This additional focus of the partition is reflected in the additional statistic $t - \tau$.

The main conclusion is that choosing a partition functions as a projectability assumption, by focusing on a set of sufficient statistics and by specifying how these statistics are used in the predictions. In the remainder of this section, I draw two further conclusions which derive from this main one. The first concerns the difference between predictions based on hypotheses schemes on the one hand, and Carnapian prediction rules or Carnap–Hintikka logic on the other. The upshot of this is that the former provides much better access to the projectability assumptions underlying the predictions. The second concerns the range of the predictions based on hypotheses schemes. I argue that there is no restriction on possible partitions, and that, when it comes to induction, the restrictive character of updating by conditioning is overrated.

As shown in Section 4, any update over the statistical partition $\mathcal{C}$ results in exchangeable predictions. This section further indicates that updates over this partition that start with a Dirichlet distribution as prior probab-

ility result in predictions which are equivalent to those produced in the Carnapian $\gamma\lambda$ continuum. As discussed in both sections, these results have sometimes been taken as a reason to refrain from using underlying chance processes or hypotheses, and to use the simpler prediction rules instead. But in both sections I also claimed that there are good reasons for adhering to the complicated hypotheses schemes after all. Recall further that even while Hintikka systems did employ universal statements in the construction of inductive prediction rules, I claimed that these systems did not make full use of the possibilities that universal statements seem to offer. I can now make explicit the reasons for adhering to the hypotheses schemes as opposed to Carnapian prediction rules, and also the way in which these schemes offer more room for using universal statements than the Hintikka systems.

First, note that any inductive method must be based on some kind of projectability assumption. This can be concluded from the abundant literature on the Humean problem of induction, and the further literature on projectability, as in Stalker (1996). In this context inductive means that the method allows for learning from past observations: dogmatic prediction rules, which are completely insensitive to incoming data and predict constant chances, are not inductive. So any inductive method must assume that past observations are somehow indicative of future observations, and this comes down to a projectability assumption. Inductive Carnapian prediction rules employ such projectability assumptions just as well as hypotheses schemes. A first advantage of hypotheses schemes over Carnapian rules then is that they provide direct insight into these projectability assumptions, as is argued in the above. But the advantage of hypotheses schemes is not just that they provide insight into the projectability assumptions. It may be argued that the Carnapian $\gamma\lambda$ continuum provides this insight just as well, because the prediction rules of this continuum also depend on the data e_t only through the corresponding sufficient statistics t_q. The more discriminative advantage is that hypotheses schemes provide better control over the projectability assumptions.

Let me illustrate the control over inductive assumptions with the example of Section 4. Imagine that we already model a focus on relative frequencies, and that we want to model an additional focus on a crash pattern in the observations. Now if we employ prediction rules for the original model, we must incorporate the statistic $t - \tau$ into the current rule of the $\gamma\lambda$ continuum. But it is unclear how exactly to incorporate it, because we do not have insight in the projectability assumptions implicit to the form of that computation. This problem appears for Hintikka systems as well, because there is no room for alternative universal statements next

to the generalizations on the number of possible observations. By contrast, modelling an additional focus on a crash pattern with hypotheses schemes is straightforward: just add the hypotheses that pertain to the patterns of interest. Therefore, the hypotheses schemes may be more complicated, but in return they offer a better control over the projectability assumptions which are implicit in the predictions.

The final conclusion concerns the freedom in choosing partitions. It is related to the philosophical status of Bayesianism. As argued above, a partition determines the projectability assumptions that are implicit to the predictions. Furthermore, the partition is entirely under the control of the inductive agent, and in particular, Bayesianism gives no directions as to what partition to choose. So there is no restriction on projectability assumptions that stems from Bayesianism. Just like we can choose a partition which focuses on relative frequencies and crash patterns, we can choose a partition that expresses the gambler's fallacy, so that with the piling up of 0's in the crash the observation of 1 is predicted with growing confidence. The hypotheses schemes are in this sense a very general tool: any inductive prediction rule, as long as it is based on the assumption of projectable patterns, can be captured in predictions generated with a hypotheses scheme. This suggests that Bayesianism is not a particular position on inductive predictions at all, but rather an impartial tool for modelling predictions.

6. THE PROBLEM OF INDUCTION

It is instructive to confront the predictions based on hypotheses schemes with the Humean problem of induction, according to which we cannot justify any kind of prediction, certain or probable, on the basis of observations only. The following shows that hypotheses schemes do not suggest anything towards solving the problem of induction, but rather that they reveal the need for inductive assumptions.

The predictions based on hypotheses schemes, if considered as an attempt at solving the problem of induction, employ the following strategy: use past observations to determine the probability over a partition of hypotheses, and then use the probability over these hypotheses in a computation of probabilities for future observations. It is tempting to say that, in choosing a partition as part of such a strategy, we do not make any substantial assumptions, because the hypotheses in a partition exhaust the space of logical possibilities and therefore constitute a tautology. Moreover, in the case of the statistical partition, it can be argued that the prior over the partition does not present an assumption either: according to the aforementioned convergence results of Gaifman and Snir (1982), any

prior eventually leads to the same predictions. In this way the hypotheses schemes can be seen as a fruitful cooperation of past observations with a completely innocent partition of hypotheses.

However, as I have argued in the above, the predictions based on partitions are made at the cost of similar inductive assumptions. Whereas for deductive purposes the partitions are indeed innocent, for inductive purposes they introduce an assumption of projectability. Therefore the predictions using hypotheses schemes do not solve the problem of induction simpliciter. Rather they reveal the assumptions needed to justify inductive predictions: assumptions of projectable characteristics, as expressed in a partition. These projectability assumptions are stronger than the unqualified assumption of the uniformity of nature. Assuming the general uniformity of nature leaves unspecified the kind of pattern with respect to which nature is uniform, while the use of a particular partition comes down to the assumption of uniformity with respect to a specific set of patterns. The predictions based on partitions bring to the fore the specific, stronger assumptions of inductive predictions. This neatly ties in with Goodman (1955, pp. 59–81). The present paper can in fact be seen as a formal expression of the fact that employing probabilistically valid inference rules reduces the Humean problem of induction to that of Goodman.

Let me consider some ways of justifying or at least weakening the projectability assumptions. The first concerns the possibility of an impartial partition, which does not preselect any kind of pattern in advance. If such a partition is possible, it can be argued that the predictions based on this partition assume just the overall uniformity of nature, or even perhaps no uniformity at all. The second concerns the possibility of considering all projectable patterns simultaneously.

To assess the first possibility, consider again the example of the statistical partition $\mathcal{C}$. Intuitively this may have seemed a modest partition, one in which there is no assumption. But as can be seen from the fact that there are sufficient statistics for this partition, predictions based on it focus on some pattern in the observations, and thus assume the projectability related to this pattern. Now we can easily generalize this way of identifying projectability assumptions: any partition that has sufficient statistics, as long as they do not at all times coincide with the complete ordered sequence of observations e_t, focuses on some pattern in the observations, and therefore must employ some kind of projectability assumption. There is only one partition for which the sufficient statistics in fact coincide with the complete sequence of observations. In this partition, here denoted with $\mathcal{E}$, the hypotheses H_e consist of the singletons $\{e\}$. With this partition I deal

below. The conclusion thus far is that, apart from the limiting case $\mathcal{E}$, there is no partition that is impartial with respect to projectability assumptions.

Another way of weakening the strong uniformity assumption is to generalize it, by simultaneously using all possible partitions. An ambitious Bayesian may argue that a partition must encompass all hypotheses that can be formulated in the current language, because as long as nothing is known of the observations, none of the possible patterns can be excluded. Such a partition focuses on all possible patterns, corresponding to the general uniformity assumption that there is some unspecified pattern in the observations. However, given the observation field $\mathcal{Q}$, we can always find some observation q_i which tells apart any two infinite sequences e and e'. Therefore, it seems that the partition which encompasses all hypotheses that can be formulated in the given language is again the limiting case mentioned in the preceding paragraph, the singleton partition $\mathcal{E}$. I now fully concentrate on this limiting case.

Note first that the predictions resulting from $\mathcal{E}$ are determined entirely by the prior over the partition, $p_{[e_0]}(H_e)$. This is because the likelihoods of all the separate singleton hypotheses is either 0 or 1. With every new observation q_i, conditioning over $\mathcal{E}$ therefore means that all singleton hypotheses H_e for which $e(i) \neq q_i$ are discarded, and all other singleton hypotheses stay in. Because of this, conditioning over the hypotheses in $\mathcal{E}$ does not in itself give any comprehensible information on future observations, so that the singleton partition does indeed not carry any projectability assumption. However, it can also be noted that the task of determining the prior over the singleton partition is very similar to the task of deciding over the prior by means of a Carnapian prediction rule, as in expressions (16) and (17). The singleton partition also requires us to specify the prior $p_{[e_0]}$ directly and as a single function over the whole of $\mathcal{Q}$. But now it seems that we are back where we started. As I remarked in the above, inductive Carnapian prediction rules employ projectability assumptions just as well as hypotheses schemes do. In the limiting case $\mathcal{E}$, it seems that these assumptions remain completely implicit to the probability over $\mathcal{E}$. In an attempt at finding some partition which expresses an impartial or generalized projectability assumption, we have pushed this assumption out of sight.

The above suggests that the tools provided in the hypotheses schemes do not offer any help in solving the problem of induction. The schemes express the assumptions needed for predictions, but they do not suggest any natural or minimal assumption. However, we may be able to find independent reasons for certain inductive assumptions. Central to finding such independent reasons is the question what it means to assume the

projectability of certain patterns in the observations. For a realist, such assumptions will indicate the fixed properties of an underlying process that generates the observations. They may be based on assuming natural kinds or essences in reality. In the case of the statistical partition $\mathcal{C}$ and the stock market, a realist may suppose that some bidding process with fixed stochastic properties underlies the formation of stock price. For an empiricist, by contrast, the inductive assumption implicit in the use of $\mathcal{C}$ is probably no more than the empirical generalization on the constancy of relative frequencies itself, and can perhaps be based further on a natural axiomatization of the observation language. I do not exclude that there are good and independent, realist or empiricist, reasons for adhering to certain assumptions. For now it is important to note that these reasons are not implicit in the tools which are offered by the hypotheses schemes. That is, conditioning over partitions provides useful insight into the problem of induction, but we cannot solve the problem with an appeal to the formal aspects of partitions.

The above conclusions are very much in line with what I like to call the logical solution to the problem of induction. This solution has recently been proposed by Howson in (2000), but it has its roots already in Ramsey and De Finetti. The same solution is in fact implicit in many papers arguing for local as opposed to global induction in Bogdan (1976) and, in a sense, in Norton (2003). The negative part of this solution is that, taken on itself, the problem of induction cannot be solved. Predictions must be based on inductive assumptions, and there is no way of deciding over these assumptions by formal or other aprioristic means. Rather more prosaically, we cannot build a house just by buying nice tools, because we also need bricks, planks and mortar. The positive part of the logical solution is that once the inductive assumptions are made, a Bayesian logician can tell how to deal with the observations. Bayesian updating functions as a consistency constraint, and generates predictions from the assumptions and observations together. Note that it is implicit to this that there is nothing inductive about Bayesian updating itself. It merely links inductive assumptions with observations to render the consistent inductive predictions.

As for the role of the present paper, it shows how partitions provide access to inductive assumptions in a Bayesian framework. It can therefore be seen as a further elaboration of the logical solution to the problem of induction. Moreover, in this guise it is a starting point for dealing with a host of other philosophical problems. For example, ordinary life and science show that humans and other animals can be quite skillful in making inductive predictions. The suggestion of Peirce, that we guess efficiently, is deeply unsatisfactory as an explanation of this. The above suggests that

in a logical picture of these skills, choosing a partition is the essential component. That is, the logical picture brings the puzzle of predictive skills down to the puzzle of how we choose partitions. Admittedly, this is only a small step further. As is nicely illustrated by Chihara (1987), the complexity of actual inductive practice leaves us with little hope for a unified theory of choosing partitions. But nevertheless, it may be a comfort that the hypotheses schemes allow us to isolate the overall validity of reasoning from the specific truth or falsity of the assumptions of that reasoning.

7. CONCLUSION

Sections 1 to 3 have introduced inductive predictions, and the use of partitions in defining such predictions. The examples of Section 4 illustrate how partitions determine the resulting predictions. In Section 5 I argue that a partition in fact expresses inductive assumptions concerning the projectability of particular characteristics of the observations. Partitions thus come out as a useful tool in defining the predictions. Section 6 further shows that the partitions themselves do not offer any directions for solving the problem of induction. Finally, this is seen to be in line with the so-called logical solution to the problem of induction.

This paper has three main conclusions. The first is that inductive predictions can be determined by choosing a partition in a hypotheses scheme. The second, more general conclusion is that a partition expresses inductive assumptions on the projectability of particular characteristics of observations. A third conclusion is that we are entirely free in choosing these projectability assumptions, and that no such assumption is naturally suggested by the hypotheses schemes themselves.

Further conclusions were seen to follow from these main ones. One specific conclusion concerns the range of prediction rules covered by hypotheses schemes. The example shows that the schemes enable us to model predictions typical for hasty generalization. But because there is no restriction on choosing partitions, it seems that any prediction rule can be formulated in a hypotheses scheme. The use of Bayesian updating is therefore not restrictive over possible prediction rules. Hypotheses schemes simply provide a way of modelling predictions, not of restricting them to a specific class. Another specific conclusion is that the hypotheses schemes offer a better control over inductive predictions than the prediction rules from the Carnap–Hintikka tradition. In part, this tradition has focused on the properties of prediction rules only, and in part it has not fully exploited the use of underlying chance processes. The above shows how partitions provide control over projectability assumptions. This suggests that in the

construction of prediction rules, there are good reasons for employing hypotheses on chance processes after all.

A different set of conclusions concerns the hypotheses schemes in relation to the Humean problem of induction. As said, hypotheses schemes bring out the inductive assumptions underlying predictions. In the hypotheses schemes, the assumptions show up in a more specific form than the general uniformity of nature, as it was suggested by Hume. This latter uniformity concerns any possible pattern in data, while partitions limit the uniformity of nature to a particular set of probabilistic patterns. Moreover, the hypotheses schemes do not suggest a partition representing some minimal uniformity assumption. There is no impartial or preferred partition which flows naturally from the hypotheses schemes themselves. We are thus forced to choose the set of patterns that deserves focus solely by our own lights.

The general tendency in all this is to view inductive logic as a proper logic: any prediction must be based on inductive assumptions, or premisses, and given these assumptions, the predictions follow from the observations by conditioning, which functions as the only inference rule. The work of induction is not done by an inference rule that implicitly contains uniformity assumptions, but by partitioning logical space and fixing the likelihoods and priors on the basis of that. In making inductive predictions, choosing partitions thus occupies a central place.

REFERENCES

Bacchus, F., H. E. Kyburg Jr., and M. Thalos: 1990, 'Against Conditionalization', *Synthese* **85**, 475–506.

Billingsley, P.: 1995, *Probability and Measure*, John Wiley & Sons, New York.

Bogdan, R.: 1976, *Local Induction*, D. Reidel, Dordrecht.

Carnap, R.: 1950, *The Logical Foundations of Probability*, University of Chicago Press, Chicago.

Carnap, R.: 1952, *The Continuum of Inductive Methods*, University of Chicago Press, Chicago.

Carnap, R. and R. Jeffrey (eds.): 1971, *Studies in Inductive Logic and Probability*, Vol. 1, University of California Press, Berkeley.

Carnap, R. and W. Stegmüller: 1959, *Inductive Logik und Wahrscheinlichkeit*, Springer Verlag, Wien.

Chihara, M.: 1987, 'Some Problems for Bayesian Confirmation Theory', *British Journal for the Philosophy of Science* **38**, 551–560.

Costantini, D.: 1979, 'The Relevance Quotient', *Erkenntnis* **14**, 149–157.

De Finetti, B.: 1964, 'Foresight: its Logical Laws, its Subjective Sources', in H. Kyburg and H. Smokler (eds.), *Studies in Subjective Probability*, John Wiley & Sons, New York, pp. 97–158.

De Finetti, B.: 1972, *Probability, Induction and Statistics*, John Wiley & Sons, New York.

Earman, J.: 1992, *Bayes or Bust*, MIT Press, Cambridge, MA.

Festa, R.: 1993, *Optimum Inductive Methods*, Kluwer Academic Publishers, Dordrecht.

Gaifman, H. and M. Snir: 1982, 'Probabilities over Rich Languages', *Journal of Symbolic Logic* **47**, 495–548.

Goodman, N.: 1955, *Fact, Fiction, and Forecast*, Harvard University Press, Cambridge, MA.

Hintikka, J.: 1966, 'A Two-Dimensional Continuum of Inductive Methods', in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*, North Holland Publishing Company, Amsterdam.

Hintikka, J.: 1970, 'Unknown Probabilities, Bayesianism, and De Finetti's Representation Theorem', in R. C. Buck and R. S. Cohen (eds.), *Boston Studies in the Philosophy of Science*, Vol. VIII, D. Reidel, Dordrecht.

Hintikka, J.: 1997, 'Comment on Theo Kuipers', in M. Sintonen (ed.), *Knowledge and Inquiry*, Rodopi, Amsterdam.

Hintikka, J. and I. Niiniluoto: 1976, 'An Axiomatic Foundation of the Logic of Inductive Generalization', in M. Przelecki, K. Szaniawski, and R. Wójcicki (eds.), *Formal Methods in the Methodology of the Empirical Sciences*, Synthese Library 103, Dordrecht.

Howson, C.: 1973, 'Must the Logical Probability of Laws be Zero?', *British Journal of Philosophy Science* **24**, 153–182.

Howson, C. and P. Urbach: 1989, *Scientific Reasoning, The Bayesian Approach*, Open Court Publishing Company, La Salle, IL.

Howson, C.: 2000, *Hume's Problem*, Clarendon Press, Oxford.

Jeffrey, R.: 1977, 'Mises Redux', in R. E. Butts and J. Hintikka (eds.), *Problems in Methodology and Linguistics*, Kluwer Academic Publishers, Dordrecht.

Jeffrey, R.: 1984, *Probability and the Art of Judgement*, Cambridge University Press, Cambridge.

Jeffreys, H.: 1939, *Scientific Inference*, Cambridge University Press, Cambridge.

Kelly, K.: 1997, 'Learning Theory and the Philosophy of Science', *Philosophy of Science* **64**, 245–267.

Kemeny, J.: 1963, 'Carnap's Theory of Probability and Induction', in P. A. Schilpp (ed.), *The Philosophy of Rudolph Carnap*, Open Court Publishers, La Salle, IL.

Kuipers, T. A. F.: 1978, *Studies in Inductive Probability and Rational Expectation*, D. Reidel, Dordrecht.

Kuipers, T. A. F.: 1984, 'Two Types of Inductive Analogy by Similarity', *Erkenntnis* **21**, 63–87.

Kuipers, T. A. F.: 1988, 'Inductive Analogy by Similarity and Proximity', in D. H. Helman (ed.), *Analogical Reasoning*, Kluwer Academic Publishers, Dordrecht, pp. 299–313,.

Kuipers, T. A. F.: 1997, 'The Carnap–Hintikka Programme in Inductive Logic', in M. Sintonen (ed.), *Knowledge and Inquiry*, Rodopi, Amsterdam.

Niiniluoto, I.: 1976, 'Inquiries, Problems, and Questions: Remarks on Local Induction', in R. Bogdan (ed.), *Local Induction*, D. Reidel, Dordrecht.

Niiniluoto, I.: 1981, 'Analogy and Inductive Logic', *Erkenntnis* **16**, 1–34.

Norton, J. D.: 2003, 'A Material Theory of Induction', *Philosophy of Science* **70** (October), 647–670.

Peirce, C.: 1934, *Collected Papers*, Vol. 5, Harvard University Press, Cambridge, MA.

Putnam, H.: 1963, 'Degree of Confirmation and Inductive Logic', in P. A. Schilpp (ed.) *The Philosophy of Rudolf Carnap*, Open Court Publishing Company, La Salle, IL, reprinted in H. Putnam: 1975, *Mathematics, Matter and Method*, Cambridge University Press, Cambridge, pp. 270–292.

Putnam, H.: 1963, 'Probability and Confirmation', *The Voice of America* **10**, reprinted in *Mathematics, Matter and Method*, 1975, Cambridge University Press, Cambridge, pp. 293–304.

Skyrms, B.: 1991 'Carnapian Inductive Logic for Markov Chains', *Erkenntnis* **35**, 439–460.

Skyrms, B.: 1996, 'Bayesian Projectability', in D. Stalker (ed.), *Grue!*, Open Court Publishers, La Salle, IL.

Stalker, D.: 1996, *Grue!*, Open Court Publishers, La Salle, IL.

Stegmüller, W.: 1973, *Carnap II: Normatieve Theorie des Inductiven Räsonierens*, Part C, Volume IV, Springer Verlag, Berlin.

Tuomela, R.: 1966, 'Induction in Ordered Universes', in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*, North Holland Publishers, Amsterdam.

Van Fraassen, B. C.: 1989, *Laws and Symmetry*, Clarendon Press, Oxford.

Filosofisch Instituut
A-Weg 30, 9718 CW
Groningen
The Netherlands
E-mail: j.w.romeyn@philos.rug.nl