HYPOTHESES AS INDUCTIVE ASSUMPTIONS

This chapter reveals the advantages of Bayesian schemes in generating inductive predictions. It discusses Carnap-Hintikka inductive logic, and two ways in which the Bayesian schemes may expand it. Bayesian schemes are then illustrated with two partitions. One partition results in the Carnapian continuum of prediction rules, the other results in predictions typical for hasty generalisation. 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.

The inductive predictions in this chapter are cast in the framework of section 1.2.1, and in particular in the Bayesian scheme introduced in section 1.3. Familiarity with these sections is essential to an understanding of this chapter. The introduction to this chapter is especially useful for those readers who have not read the preceding chapters, but apart from that it also indicates the main line of this chapter.

3.1 Introduction

Inductive predictions. This chapter concerns inductive predictions, taken as the result of inductive inferences. The premisses of an inductive inference include a set of observations and possibly some further assumptions. The conclusions may be predictions or general hypotheses, where predictions concern future observations, and general hypotheses are observational generalisations of some sort or other. For example, from the fact that some internet start-up 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 observation, 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 start-up will have a decreasing stock price on all future days, which is a general hypothesis about the observations.

The Bayesian scheme uses hypotheses to arrive at predictions. The data are first reflected in an opinion about a partition of hypotheses. For example,
from the data on decreasing stock price we first derive an opinion about hypotheses on the state and nature of the internet start-up. The predictions on the internet start-up are subsequently derived from this opinion about hypotheses, together with the data. These predictions and opinions about hypotheses are thus expressed in terms of probability functions. In sum, this chapter concerns probabilistic inductive inferences in which hypotheses are used for making predictions. In conformity with the preceding chapters, I will say that such predictions are based on Bayesian schemes, or alternatively, based on partitions.

Organisation of chapter. The main line of the chapter is the following. First I show that the Bayesian scheme enables us to describe predictions typical for hasty generalisation. The predictions can be generated by choosing a specific partition of hypotheses for the scheme. This example triggers two separate discussions on the function of hypotheses in the Bayesian schemes. The main conclusion of the discussion in this chapter is that hypotheses are tools for making inductive assumptions. They determine which patterns are identified in the data, and subsequently projected onto future observations. Another discussion, which concerns Bayesian schemes in relation to the problem of induction, is presented in chapter 7.

In more detail, the structure of the chapter is as follows. In section 3.2, I introduce the dominant tradition in formalising inductive predictions, called Carnap-Hintikka inductive logic. The Bayesian scheme of this chapter is seen to expand the Carnap-Hintikka tradition. For the Bayesian scheme itself I refer to section 1.3. Section 3.3 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 3.4 the examples are given a further philosophical interpretation. Specifically, the hypotheses are related to inductive assumptions. The conclusion connects these insights to the schemes of the preceding chapters.

3.2 Carnap-Hintikka inductive logic

This section discusses the Carnap-Hintikka tradition of inductive logic. It emphasises two characteristic features of this tradition: its focus on exchangeable prediction rules, and its reluctance, certainly within the Carnapian literature, to employ general or statistical hypotheses. The inductive predictions of this chapter, and more generally of this thesis, extend the Carnap-Hintikka tradition by departing from these two features.
Carnapian prediction rules. Let me briefly rehearse Carnapian predictions for the purpose of this chapter. Recall that predictions are probabilities over future observations based on observations and further assumptions. The observations are encoded in indexed natural numbers $q_i$, and collected in ordered tuples $e_t = (q_1, q_2, \ldots, q_t)$. The further assumptions can be encoded in some collection of parameters $X$. We may then construct a prediction rule $pr_X(q, e_t)$ for the next observation having the result $q$ in terms of a probability function of the preceding observations $e_t$, the next observation $q_{t+1}$ and the collection of parameters $X$. Inductive predictions can be studied by designing and comparing classes of such probability functions, which I call inductive prediction rules.

An exemplary class of probabilistic inductive inference rules for making predictions is given by the $\lambda\gamma$ rules referred to earlier, as developed in Carnap (1950, 1952) and defined fully in Stegmüller (1973):

$$pr_{\lambda\gamma}(q, e_t) = \left(\frac{t}{t+\lambda}\right) \frac{t_q}{t} + \left(\frac{\lambda}{t+\lambda}\right) \gamma_q.$$  \hspace{1cm} (3.1)

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

After Carnap, inductive prediction rules have been studied extensively. Axiomatisations, elaborations and syntheses 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.

Exchangeability. 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. These exchangeable rules typically apply to settings in which the events producing the observations are independent. So they 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 (1982), almost al-
ways 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.

*Representation theorem.* The second feature of this tradition that I want to emphasise can only be made explicit after presenting the connection, in both directions, between the exchangeability of observations and the independence of the events that may be supposed to produce these observations. For this I must first elaborate on the notion of independence. A first component of this notion is the assumption that the events producing the observations are in fact part of some underlying process. A second component is that if this underlying process generates the events with constant objective chances, then the chance of an event is independent of events occurring before or after it, so that we can speak of independent events.

The connection reaching 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 in a Bayesian scheme, using hypotheses that are associated with constant chance processes. Section 3.3 deals with this representation theorem in some more detail. The connection reaching from independence to exchangeability, on the other hand, is established by the fact that any Bayesian scheme using hypotheses on constant chance processes must result in an exchangeable prediction rule. This is seen most easily from the fact that updating the probability over the hypotheses for new observations is a commutative operation. 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 3.3 discusses this in more detail. For now it is important to note that the assumption of the independence of the events producing the observations can be equated with the use of exchangeable prediction rules for these observations.

*Against general hypotheses.* I can now make explicit the second characteristic feature of the Carnap-Hintikka tradition. De Finetti interpreted the representation theorem as a reason to omit all 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 dislike of the objective chances featuring in the underlying processes. Rather it is because these chance processes are described by general hypotheses, which cannot be decided with finite data. De Finetti deemed such hypotheses suspect for empiricist reasons.
3.2. CARNAP-HINTIKKA INDUCTIVE LOGIC

In a similar vein, Carnap maintained that all universal hypotheses have measure zero. Now there are large differences between De Finetti and Carnap, but both used the representation theorem to show that it is simply unnecessary to employ chance processes. We can obtain the same results using the exchangeability of the prediction rule, and in this way we stay closer to the empiricist roots of inductive logic.

In line with this, most of the Carnap-Hintikka tradition focuses on the properties of prediction rules, such as exchangeability, and eschews reference to the chance processes that may be underlying these 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. Indeed, in Hintikka (1966) and Tuomela (1966) we find a different attitude towards underlying chance processes, or at least towards the use of hypotheses in inductive logic. More in particular, Hintikka employs universal generalisations on observations in the construction of his $\alpha \lambda$ continuum of inductive prediction rules. Tuomela discusses more complicated hypotheses on ordered universes, and refers to Hintikka for the construction of prediction rules based on these universal statements. Both these authors thus employ hypotheses to inform predictions in a specific way.

Innovations of this chapter. While this already presents a valuable extension, I feel that hypotheses 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 of Hintikka offers little room for varying the kinds of hypotheses used, since the continuum concerns universal generalisations only. It is certainly an advantage that, in the improved versions of Hintikka and Niiniluoto (1976) and Kuipers (1978), the role of these generalisations is not entirely determined by the single parameter $\alpha$. However, as Hintikka himself remarks in (1997), it is eventually much more convenient to be able to express such universal statements in terms of proper premisses, so that other kinds of universal statements can be employed too, and also controlled more naturally. However, many prediction rules in which the use of specific universal statements seems very natural do not employ such statements in their construction. Take, for example, the inductive prediction rules for Markov chains by Kuipers (1988) and Skyrms (1991), and the prediction rules describing analogical reasoning by Niiniluoto (1981) and Kuipers (1984). Here the construction of the rules is based on particular predic-
CHAPTER 3. HYPOTHESES AS INDUCTIVE ASSUMPTIONS

I can now indicate more precisely the innovations that this chapter offers. It extends the Carnap-Hintikka tradition in inductive logic in two ways, connected to the two characteristic features noted above. First, this chapter proposes a prediction rule that is not exchangeable, by adding hypotheses concerning a particular deterministic pattern to an existing partition of constant chance hypotheses. Second, and following up on this, it advocates the explicit use of chance processes, or hypotheses on such processes, in the definition of inductive predictions. In accordance with the Bayesian scheme of chapter 1, and as illustrated in figure 3.1, hypotheses are used to mediate between observations and predictions. The claim following from this is that partitions of hypotheses are a tool in making assumptions about patterns in data, which widens the scope of the Carnap-Hintikka tradition. In connection with the discussions in chapters 1 and 2, it may be said that this chapter presents the main reason for preferring the rather complicated Bayesian scheme over the Carnapian.

Remark and disclaimer. In connection with chapter 2, it may be noted that the hypotheses that may be used in a Bayesian scheme belong to the class of frequentist hypotheses. One aspect of this class is in this chapter given a further illustration. Recall that the class also encompasses the hypotheses $H_{w\theta}$ employing a selection function $w$ that sometimes assigns state $w(e_t) = 0$ to sequences $e_t$ that are not given a zero probability by the hypotheses, but that only does so for a finite number of sequences $e_t$. Such hypotheses, it was argued,

\[\text{observation}\]

\[\text{observed facts} \quad \overset{\text{dashed}}{\rightarrow} \quad \text{predictions}\]

\[\text{observational generalisation}\]

\[\text{hypotheses}\]

Figure 3.1: The Bayesian scheme employs hypotheses to mediate between observations and predictions.
3.3. EXAMPLES ON CRASH DATA

This chapter employs hypotheses of exactly this sort, namely the so-called crash hypotheses. Finally, let me disclaim the treatment of some topics that otherwise complicate the discussion too much. First, it can be noted that the prediction rules of this chapter are somewhat similar to those in the paper by Tuomela on ordered universes. Both focus on predictions based on the 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 the representation theorem of De Finetti in full generality, and similarly I will not touch upon the various brands of partial exchangeability. The focus of this chapter is on a particular non-exchangeable prediction rule, generated by hypotheses concerning particular chance processes, and on the moral that derives from the use of such hypotheses. There are excellent discussions of representation theorems on offer.

3.3 Examples on crash data

This section gives two applications of the scheme of section 1.3. The first application employs hypotheses on constant chances for the observations, and results in the Carnapian $\lambda\gamma$ rules. 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 $\lambda\gamma$ continuum, and they are not in general exchangeable.

3.3.1 Bernoulli partition

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

$$e_{35} = 0100100000010110000010000000010000,$$
$$e^*_{35} = 01001011101000000000000000000000000.$$

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 the strings. In particular, $e^*_{35}$ shows what can be called a crash: from some day onwards we only observe decreasing stock price.
**Defining Bernoulli hypotheses.** Now imagine a marketeer who aims to predict stock price movements based on observed price movements over foregoing trading days. Further, assume that she employs the partition $\mathcal{B}$ with a continuum of hypotheses to specify her predictions. To characterise the hypotheses, define

$$f(e) = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} e(i).$$  \hfill (3.2)

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

$$I_{h\theta}(e) = \begin{cases} 
1 & \text{if } f(e) = \theta, \\
0 & \text{otherwise},
\end{cases}$$  \hfill (3.3)

in which $\theta \in [0, 1]$. Further define $I_{h \neg \theta}(e) = 1$ if $f(e)$ is not defined, and $I_{h \neg \theta}(e) = 0$ otherwise. Then $\mathcal{B} = \{H_{\neg \theta}, \{H_{\theta}\}_{\theta \in [0,1]}\}$ is a partition including a continuum of hypotheses on relative frequencies: every $e$ belongs to a unique hypothesis. I call $\mathcal{B}$ the Bernoulli partition, after Johan Bernoulli who first studied chance processes of this form.

Assume that the marketeer employs the following input probabilities:

$$\int_{0}^{1} p_{[e_0]}(H_{\theta}) d\theta = 1,$$

$$p_{[e_0]}(H_{\theta}) = 1,$$  \hfill (3.4)

$$p_{[e_0]}(H_{\neg \theta}) = 0,$$  \hfill (3.5)

$$\forall i > 0 : p_{[e_0]}(Q_{i+1}^q | H_{\theta}) = \begin{cases} 
\theta & \text{if } q = 1, \\
1 - \theta & \text{if } q = 0.
\end{cases}$$  \hfill (3.6)

where again $\theta \in [0,1]$. Equations (3.4) and (3.5) state that the probability distribution over the hypotheses $H_{\theta}$ is uniform. This may be motivated with an appeal to the principle of indifference or some other symmetry principle. Equation (3.6) states that those sequences $e$ in which the frequency of trading days with $e(i) = 1$ has no limit are negligible. 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. Moreover, it is required if we want to illustrate the representation theorem.

Equation (3.7) can be motivated with the restriction on likelihoods as discussed in section 2.4. We may further assume that at every $i$ the Bayesian agent
3.3. EXAMPLES ON CRASH DATA

Figure 3.2: Predictions $p_{[e_{i}]}(Q_{t+1}^{q})$ against time $t$, based on the partition $\mathcal{B}$. The dashed line shows the predictions for the normal data $e_{35}$, the unbroken line shows the predictions for the crash data $e_{35}^{*}$.

updates the likelihood that is used in the next prediction to

$$p_{[e_{i}]}(Q_{t+1}^{q}|H_{\theta}) = p_{[e_{0}]}(Q_{t+1}^{q}|H_{\theta}),$$

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

Resulting predictions. With the prior density $p_{[e_{0}]}(H_{\theta})$ and the likelihoods $p_{[e_{i}]}(Q_{t+1}^{q}|H_{\theta})$, we have specified all probabilities that are needed. We can compute the predictions on next observations $p_{[e_{i}]}(Q_{t+1}^{Q_{t+1}})$ that the marketeer makes when confronted with the observations $e_{35}$ and $e_{35}^{*}$ respectively. I have calculated these predictions and depicted them in figure 3.2. In the remainder of this subsection I make some remarks on these predictions, and on the Bayesian scheme with the Bernoulli partition in general.

Note that after 32 days, the predictions in figure 3.2 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 $p_{[e_{i}]}(Q_{t+1}^{Q_{t+1}})$ converge to 1. The speed of convergence, however, decreases with the addition of further instances of $e_{t}(i) = 0$. More precisely, the second derivative to time of the predictions, taken as a function over time, is negative. Thus the
predictions do not accommodate the fact that the data \( e_{35}^* \) may be the result of a crash.

The Bayesian scheme using \( \mathcal{B} \) illustrates the representation theorem of De Finetti. The hypotheses \( H_\theta \) are the hypotheses on processes with constant chances that I alluded to in sections 3.2. The representation theorem is that any exchangeable prediction rule \( pr_X(q, e_t) \) can be represented in a Bayesian scheme with the partition \( \mathcal{B} \). Different exchangeable prediction rules may be defined by choosing different priors \( p[q_0](H_\theta) \). For example, choosing a Dirichlet density for \( p[q_0](H_\theta) \) results in a \( \lambda \gamma \) prediction rule. As described in Festa (1993), the parameters of the Dirichlet density fix the values of \( \gamma \) and \( \lambda \) in this rule. Specifically, choosing the uniform prior of equation (3.5) results in the so-called straight rule, which has the parameters \( \lambda = 2 \) and \( \gamma_q = \frac{1}{2} \). Note however that the range of the representation theorem is much wider than the equivalence of the \( \lambda \gamma \) rules and the Dirichlet distributions over \( \mathcal{B} \).

As section 3.2 indicated, the representation theorem was welcomed as a way to replace Bayesian schemes using \( \mathcal{B} \) for exchangeable prediction rules. One reason for the replacement was that the Bayesian schemes committed to the assumption of underlying chance processes and the assignment of probability to universal hypotheses. Another, more immediate reason for not using the Bayesian schemes may be that they are unnecessarily roundabout: in the end they generate the same predictions as the Carnapian scheme. In the foregoing and in the following, however, I explicitly use the Bayesian schemes to design and study predictions. In section 3.4 I argue that there are independent reasons for doing so.

### 3.3.2 Crash hypotheses

**Alternative hypotheses.** Figure 3.2 shows the inductive predictions of a marketeer who is not sensitive to the possibility of a crash. Below I alter the partition in such a way that this sensitivity is modelled. This is done by adding hypotheses to the Bernoulli partition, thus implicitly altering the resulting prediction rule. In particular, I add the hypotheses \( g^2_{\gamma, \lambda} \) to the Bernoulli partition, the meaning of which can be phrased as follows: until trading day \( \tau \), stock price behaves like the \( \lambda \gamma \) rule says, but from trading day \( \tau \) onwards, all stock price movements are \( q \).

Let us denote the partition consisting of the Bernoulli hypotheses \( h_\theta \) and the crash hypotheses \( g^2_{\gamma, \lambda} \) with \( \mathcal{C} \). The crash hypotheses can be associated with
sets $G^q_{\gamma \lambda \tau}$ in $Q$ using a characteristic function that selects for crashes:

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

(3.9)

$$G^q_{\gamma \lambda \tau} = \{ e : I_{G^q_{\gamma \lambda \tau}}(e) = 1 \}. $$

(3.10)

Note that the parameters $\gamma$ and $\lambda$ 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 $\tau$.

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

$$p[e_0](Q^q_{i+1} | G^q_{\gamma \lambda \tau} \cap E_i) = \begin{cases} 
\frac{i_q + \lambda q}{i + \lambda} & \text{if } t < \tau, \\
1 & \text{if } i = \tau - 1, q \neq q', \text{ or } i \leq \tau, q = q', \text{ or } i \leq \tau, q \neq q', \\
0 & \text{if } i = \tau - 1, q = q', \text{ or } i \leq \tau, q \neq q',
\end{cases}$$

(3.11)

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 the definition of the sets $G^q_{\gamma \lambda \tau}$. However, as there is no restriction on the first $\tau - 1$ observations in these sets, there is no restriction motivating the first clause. The likelihoods before $\tau$ may be chosen in accordance with the predictions generated by the partition $B$, 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. Note that the likelihoods of $G^q_{\gamma \lambda \tau}$ thus depend on the actual data $e_i$. This means that the likelihoods change with the update of every observation before $\tau$.

Choosing a prior. The hypotheses $G^q_{\gamma \lambda \tau}$ may be given prior probabilities of the following form:

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

(3.12)

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

(3.13)

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. The factor $\alpha$ is the total probability that is assigned to all the crash hypotheses. From the definition of the discount factor, we have $\alpha = \sum_{\tau=0}^{\infty} p[e_0](G^0_{\gamma \lambda \tau})$, so that we must choose $0 < \alpha < 1$. Note that because of equation (3.13), booming markets, in which from some time onwards prices only go up, are not considered.
The probability $1 - \alpha$ can be divided over the remaining hypotheses from $\mathcal{B}$ according to

$$
\int_0^1 p_{[\alpha]}(H_\theta) d\theta = 1 - \alpha, \quad (3.14)
$$

$$
p_{[\alpha]}(H_\theta) = 1 - \alpha, \quad (3.15)
$$

where in this case $\theta \in (0, 1]$. The likelihoods of the crash hypotheses can be made to accord with this prior by setting $\lambda = 2$ and $\gamma = \frac{1}{2}$. Note from the domain of $\theta$ that the hypothesis $H_0$ is excluded from the subpartition $\mathcal{B}$. This is because all $e \in G^0_{\gamma \lambda \tau}$ have the relative frequency $f(e) = 0$, so that for each $\tau$ we have $G^0_{\gamma \lambda \tau} \subset H_0$. However, according to the original likelihoods of $H_0$ the hypotheses $G^0_{\gamma \lambda \tau}$ must have zero probability within $H_0$, because any observation of $q = 1$ is given zero probability 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 Bayesian scheme with $\mathcal{B}$ anyway, banning it from the combined partition $\mathcal{C}$ does not affect the prediction rule that was initially derived for $\mathcal{B}$.

In sum, we have created a new partition $\mathcal{C}$, including both $H_\theta$ 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 $\alpha$ and $\delta$ determines to what extent this sensitivity influences the predictions. Admittedly, the partition $\mathcal{C}$ involves considerable idealisations, for example that a crash lasts forever and that the prior probability for a crash slowly diminishes. These idealisations are not compulsory: the Bayesian scheme offers space for further elaborations in these respects. In the following, however, I want to focus on the fundamental possibilities that the freedom in choosing partitions presents. The idealisations of $\mathcal{C}$, and the ways to avoid them, are not discussed here.

**Resulting predictions.** We can calculate the predictions $p_{[\alpha]}(Q_{t+1}^q)$ using the partition $\mathcal{C}$. Figure 3.3 shows a comparison of two marketeers confronted with the crash data $e_{35}^*$. The diamond curve shows the predictions based on the use of the partition $\mathcal{C}$, and the bullet curve shows the predictions of the Bernoulli partition $\mathcal{B}$. The hypotheses $G^0_{\gamma \lambda \tau}$ of this particular update have $\alpha = \frac{1}{4}$ and $\delta = \frac{1}{5}$. Note that the predictions based on $\mathcal{C}$ deviate from the predictions based on $\mathcal{B}$. As the unbroken string of $q_i = 0$ grows, the marketeer using $\mathcal{C}$ picks up on the crash regularity, and in subsequent days gives higher probability to the prediction that next days will show the result $q = 0$ 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 $\mathcal{C}$. This is because the probability
3.4 The use of partitions

This section discusses the use of partitions in the Bayesian scheme. After some discussing some immediate insights, I develop the idea that partitions function as inductive assumptions, or projectability assumptions, in the inductive arguments. In the last subsection I discuss how this presents an advantage for the Bayesian scheme.
3.4.1 Immediate insights

*Pattern recognition.* Several insights may be drawn from the example that uses partition $\mathcal{C}$. Firstly, the above example shows that inductive predictions based on hypotheses can be adapted to model pattern recognition, and in this particular case, hasty generalisation. 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 is already a useful extension of that tradition. Moreover, and as also discussed above, the modelling of hasty generalisation may convince those who consider updating on generalisations impossible due to the negligible measure of these generalisations in the observation field.

Secondly, and related to this, the example may be taken to qualify 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 Bayesian schemes. But the above reasoning can nevertheless be called ampliative on the level of predictions: hasty generalisation is a typically ampliative inferential move. Thus, even though Bayesian updating is itself not ampliative, the predictions resulting from a Bayesian update can in a sense model ampliative reasoning. Note that the ampliativeness is implicit in the choice of the partition $\mathcal{C}$, and not in the inference rule of the inductive scheme.

*The partition as a pair of glasses.* In choosing a different partition, I implicitly alter the resulting prediction rule: the straight rule, generated by the partition $\mathcal{B}$ with uniform prior, is replaced with some other prediction rule $p_{\alpha\delta}(q, e_t)$. Put differently, the probability assignment $p_{e_0}$ over the field $Q$, initially determined by the partition $\mathcal{B}$ and some prior probability assignment over it, now encodes a different prediction rule, determined by the partition $\mathcal{C}$ and its prior. The probability over the added hypotheses of $\mathcal{C}$ depends on a crash pattern in the data, and the resulting predictions will therefore not be exchangeable. Thus we have defined a different prediction rule by choosing a different partition in the Bayesian scheme.

In the examples, the influence of the observations is really determined by the partition. We first choose a partition and, define a prior probability assignment over it, and via the observations determine a posterior probability assignment. The predictions can then be derived from this posterior probability and the likelihoods, which are given with the choice of partition. So the pos-
terior probability over the partition is the only term in the predictions which
depends on the observations. 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.

3.4.2 Partitions as inductive assumptions

In this subsection, the function of choosing a partition is subject to further
scrutiny. I shall characterise how partitions limit the view of an observer on
observations, and how this connects to inductive assumptions.

Sufficient statistics. Consider the Bernoulli partition \( B \). The posterior proba-
bility 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 specific characteristics of the obser-
vations: for all \( q \) we must know the number of times that it occurred within
the data \( e_t \). These numbers were denoted by \( t_q \) in the above. They are the
so-called sufficient statistics for computing the probability over \( B \) at time \( t \), and
thus for generating the predictions based on \( B \). The statistics \( t_q \) express those
characteristics of the observations which are taken to be relevant for the pre-
dictions. Note that the exchangeability of the predictions based on \( B \) follows
from the fact that the sufficient statistics are independent of the order of the
observations.

The partition with crash hypotheses \( C \) limits the view on the observations
in a different way. As with the Bernoulli 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 \( C \) at time \( t \), and therefore in the generation of the predictions
based on \( C \). 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 exposition shows 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 discussion 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 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.

**Partitions as projectable patterns.** 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 given 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 competitive struggle 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 Bayesian scheme.

These considerations show 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 patterns are at the onset trusted with the predictive task, but the observations eventually determine which patterns perform this task best on the actual data. The predictions are subsequently derived by means of a weighing factor, the probability assignment
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 amounts 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 charac-
teristics of the partitions, and link them with the sufficient statistics alluded to
above. Recall once again the examples of section 3.3. Choosing the Bernoulli
partition $B$ means that we limit the possible probabilistic patterns to those for
which the observations occur with specific relative frequencies. The projectabil-
ity 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 hypothes-
oses 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 $t - \tau$.

3.4.3 Advantages of the Bayesian scheme

The main conclusion of the foregoing 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 shall draw two further conclusions which derive from this main
one.

Access to projectability assumptions. Within the inductive schemes presented in
this thesis, any inductive argument must be based on some kind of projectabil-
ity assumption. This can be concluded from the abundant literature on the
Humean problem of induction, and the further literature on projectability, as
collected in, for instance, Stalker (1996). So an inductive argument is a method
that is sensitive to past observations. Prediction rules that completely ignore
data and predict the same irrespectively of these data are not inductive. But
in that case, any inductive argument must assume that past observations are
somehow indicative of future observations, which comes down to a projectability assumption. Inductive prediction rules in a Carnapian scheme therefore employ projectability assumptions just as well as the Bayesian scheme does. For Carnap himself, the projectability assumptions are part and parcel of the choice of language. But the fact that the language provides a basis for the projectability assumptions in the Carnapian scheme must not distract from the fact that there are such assumptions in the first place.

I can now make explicit the reasons for adhering to the Bayesian schemes as opposed to Carnapian prediction rules. Recall that any update over the Bernoulli partition $B$ results in exchangeable predictions. Further, the use of a Dirichlet density as prior probability assignment over this partition results in predictions that are identical to those produced by the $\lambda\gamma$ rule. As indicated, these results have been interpreted as a reason to refrain from using underlying chance processes or hypotheses, and to use the simpler prediction rules instead. However, the foregoing claims that there are good reasons for adhering to the complicated Bayesian schemes after all: these schemes provide direct insight into the projectability assumptions, as represented in the statistical hypotheses. Moreover, even though Hintikka systems did employ universal hypotheses in the construction of inductive prediction rules, we saw that these systems did not make full use of the possibilities that hypotheses offer. In short, the Bayesian scheme has an advantage over the Carnapian scheme because it provides immediate access to the projectability assumption.

Control over projectability assumptions. The advantage of Bayesian schemes is not just that they provide insight into the projectability assumptions. It may be argued that the $\lambda\gamma$ rules, for example, provide this insight just as well, because these prediction rules depend on the data $e_t$ only via the sufficient statistics $t_q$. The further advantage, which perhaps discriminates more clearly between the Bayesian and the Carnapian scheme, is that Bayesian schemes provide better control of the projectability assumptions.

Let me illustrate the control over inductive assumptions with the crash example. Imagine that we already model a focus on relative frequencies, using a $\lambda\gamma$ rule, and that we want to model an additional focus on a crash pattern in the observations. Then we must somehow incorporate the statistic $t - \tau$ into the $\lambda\gamma$ rule we are using. But it is unclear how exactly to incorporate it, because we do not have insight in the projectability assumptions implicit to the form of the computation that we choose. The same problem appears for Hintikka systems, as there is no room for hypotheses on specific patterns, other than universal hy-
3.5. CONCLUSION

Potheses. In sharp contrast with this, modelling an additional focus on a crash pattern with Bayesian schemes is straightforward: just add the hypotheses that pertain to the patterns of interest to the partition. Therefore, the Bayesian scheme may be more complicated, but in return it offers a better control of the projectability assumptions which are implicit in the predictions.

In view of the preceding chapter, it is not surprising that the Bayesian scheme improves the access to and control over the projectability assumptions. The Bayesian scheme employs an extended observational algebra, or an extended observation language, and it is only natural that this extended language offers us more expressive power.

**Freedom in choosing projectable patterns.** A final remark concerns the freedom in choosing partitions. Note that the choice of a partition is entirely under the control of the inductive reasoner. The only possible restriction lies in the fact that we may decide to employ frequentist hypotheses only, but this is not mandatory. Bayesian inductive logic itself provides no directions or restrictions as to what hypotheses to choose. Just as 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 Bayesian schemes are in this sense a very general tool: any inductive prediction rule, as long as it is based on the assumption of some projectable pattern, can be captured in predictions generated with a Bayesian scheme. This shows that Bayesianism is not a particular position on inductive predictions, but rather an impartial tool for modelling predictions.

### 3.5 Conclusion

**Summary.** Sections 3.1 and 3.2 introduced inductive predictions and the tradition of Carnap-Hintikka inductive logic. The examples of section 3.3 illustrated how partitions determine the resulting predictions. In section 3.4 I argued that a partition expresses inductive assumptions concerning the projectability of particular characteristics of the observations. Partitions came out as a useful tool in defining the predictions. The main conclusion of this chapter is therefore that inductive predictions can be determined by choosing a partition in a Bayesian scheme, and that a partition expresses inductive assumptions on the projectability of particular characteristics of observations.

Further conclusions were seen to follow from this main one. One specific conclusion concerned the range of prediction rules covered by Bayesian schemes.
The example shows that the schemes enable us to model predictions typical for hasty generalisation. Now if we adopt the view of chapter 2, hypotheses must be chosen from the frequentist class, and it is not clear that just any prediction rule can be formulated in a Bayesian scheme. Note, however, that this restriction is not inherent to Bayesian logic, but rather to the frequentist add-on. Another specific conclusion was that the Bayesian scheme offers better insight in, and control over, inductive predictions than the prediction rules from the Carnap-Hintikka tradition. This tradition has focused primarily on the properties of prediction rules. It has not fully exploited the use of general hypotheses. The present chapter argues that in the construction of prediction rules, there are good reasons for employing these hypotheses after all.

*Inductive logic.* The general tendency in all this is in line with the main point of chapter 1. It 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 must follow from the observations by probability axioms and Bayesian updating, which function as inference rules. So the work of induction is not done by an inference rule that implicitly contains uniformity assumptions, but by partitioning the space of possible worlds, fixing the likelihoods on the basis of that, and then choosing prior probabilities. As further discussed in chapter 7, this view on inductive predictions has consequences for the way we deal with the central problem of this thesis, the problem of induction. But the logical picture itself does not suggest any solution to the problem: the choice of a partition is not informed by the logical scheme.

Let me elaborate this latter point a bit. One of the aims specific for the Carnapian tradition is to provide a predictive scheme based solely on symmetries in the prediction rules, such as exchangeability. These symmetries are given independent justification in the notion of logical probability: the gap between past and future observations is bridged with logical means. The schemes considered here do not aim for such a justification. For the purpose of this thesis, it is enough to provide a scheme in which inductive assumptions can be expressed clearly, and in which the arguments from assumptions and observations to probabilistic predictions are valid. Certainly, the quest for plausible assumptions is an important and interesting task, but I take this task to fall outside of the logical analysis of inductive inference, and more naturally situated in epistemology. For some brief considerations on this, I refer to the conclusion of this thesis.

*Analogy and independence.* While the motivation of certain inductive assumptions is not included in the task of inductive logic, it may be considered part of its
task to provide the translation of specific pre-formal considerations into formal premises. As an example, if there is reason to make inductive assumptions based on simplicity, there is still the task of making this simplicity formally precise. This may be done with a Bayesian or Akaike information criterion, as discussed in Akaike (1978), Sober (1998), Kieseppä (1997) and Bandyopadhyay and Boik (1999), or by means of minimal description length, as in Rissanen (1982). It is to the task of making certain considerations formally precise that I turn in the second part. It concerns translations of specific extra-logical considerations or insights into a form suitable for Bayesian inductive logic.

In particular, the second part shows that partitions may be employed to design prediction rules that incorporate analogy effects and independence assumptions. This also illustrates that the possibilities of the Bayesian scheme have not been employed fully in defining such predictions. Analogical prediction rules from the Carnap-Hintikka tradition may be combined using a Bayesian scheme with a partition that differs from $\mathcal{B}$, and interesting variations on these rules can be constructed by suitable transformations between partitions of hypotheses. Chapters 4 and 5 are concerned with these analogical predictions. Chapter 6 discusses inductive inference for Bayesian networks by means of the Bayesian scheme. As it turns out, the formal framework for these networks is exactly the same as the framework for analogy reasoning.

*Philosophy of science.* The third part concerns the relation between inductive assumptions in the Bayesian scheme and some main themes in the philosophy of science. It discusses the use of suppositions of underlying structure in chapters 7 and 9, and the control over changes in the assumptions within the Bayesian scheme in chapter 8. This latter research follows up on the debate over conceptual enrichment in Niiniluoto and Tuomela (1973), and also Gillies (2001), who argues that changes in the conceptual framework are a problem for the Bayesian theory. Recall that choosing a partition fixes the basic concepts that are used in the update. For example, with the hypotheses on crashes we include the phenomenon of a crash in the conceptual framework of the marketeer. We can therefore model a change in the focus on a projectable pattern by changing the partition.