A Bayesian Analysis of Some Forms of Inductive Reasoning

Evan Heit

University of Warwick

In Rational Models of Cognition, M. Oaksford & N. Chater (Eds.), Oxford University Press, 248-274, 1998.

Please address correspondence to:

Evan Heit
Department of Psychology
University of Warwick
Coventry CV4 7AL, United Kingdom
Phone: (024) 7652 3183
Email: E.Heit@warwick.ac.uk

One of our most important cognitive goals is prediction (Anderson, 1990, 1991; Billman & Heit, 1988; Heit, 1992; Ross & Murphy, 1996), and category-level information enables a rich set of predictions. For example, you might not be able to predict much about Peter until you are told that Peter is a goldfish, in which case you could predict that he will swim and he will eat fish food. Prediction is a basic element of a wide range of everyday tasks from problem solving to social interaction to motor control. This chapter, however, will focus on a narrower range of prediction phenomena, concerning how people evaluate inductive “syllogisms” or arguments such as the following example:

```
Goldfish thrive in sunlight
---------------------------
Tunas thrive in sunlight.
```

(The information above the line is taken as a premise which is assumed to be true, then the task is to evaluate the likelihood of the conclusion, below the line.) Despite the apparent simplicity of this task, there are a variety of interesting phenomena that are associated with inductive arguments. Taken together, these phenomena reveal a great deal about what people know about categories and their properties, and about how people use their general knowledge of the world for reasoning.

The goal of this chapter is to take initial steps towards applying an account from Bayesian statistics to the task of evaluating inductive arguments. A Bayesian model may be considered an optimal account of induction that, ideally, would make predictions that bear some resemblance to what people actually do in inductive reasoning tasks.

Assumptions for Rational Analysis

The Bayesian model presented here is meant to be a computational-level account (Marr, 1982), in that it is a description of the task that is performed in evaluating inductive arguments, rather than a detailed process-level account. In this way, the Bayesian account fulfils the first step of Anderson’s (1990) scheme for rational analyses, specifying the goals of the system during a particular task. However, this account does not contain other elements of a rational analysis, such as a description of the environment. For inductive reasoning, the environment might be something as large as all properties of all objects, or all beliefs about properties of objects, and it is not clear how a description of the environment would be undertaken. The Bayesian model for inductive reasoning does have one other element in common with Anderson's models, though: the use of Bayesian statistics. Anderson has shown for several tasks the value of considering people's prior hypotheses along with the new information that is available. Bayesian statistics, with its methods for deriving posterior distributions by putting together prior distributions with new evidence, is especially appropriate for such analyses. For inductive reasoning, too, it seems appropriate to develop a model that considers people's prior knowledge as well as the new information contained in the premises of an inductive argument.

The Bayesian analysis of induction depends on three assumptions, which represent a new way of conceptualising inductive reasoning at the computational level. The following assumptions refer to a canonical inductive argument of the form:

Category A_1 has Property P
Category A_2 has Property P
...
Category A_n has Property P

Category B has Property P.

Assumption 1. Evaluating an inductive argument is conceived of as estimating the range of Property P. The range of Property P refers to which categories have P and which do not. In its simplest form, the range of P can be thought of as a function that maps a set of categories onto the values {True, False}. (However, more elaborate schemes would be possible, for example adding a Don't Know value.) Thus, at its core, induction is conceived of as a kind of learning about properties, in which people try to form a better estimate of the range of a property. This conception may seem novel; on the surface the goal of induction might appear to be drawing an inference about Category B rather than learning about Property P.

Assumption 2. When learning about novel properties, people rely on a number of prior hypotheses about the possible ranges for properties. Most notably, people initially assume that novel properties will be distributed like already-known properties. For example, if someone learns a new fact that dogs have substance P in their blood, a number of hypotheses may be raised, derived from a number of known properties. One hypothesis could be that the property "has substance P" is a property of all animals, analogous to the known property "is a living thing." Another hypothesis is that "has substance P" is true only of mammals, like the known property "has hair or fur." Still another hypothesis is that the novel property is true for dogs and false for all other animals, like the known property "is considered man's best friend."

Because people already know about a number of familiar properties, it is assumed that in general, people entertain a number of hypotheses about the possible range of a new property. For example, people might simultaneously consider the hypotheses that "has substance P" is an all-animal property, a mammal-only property, and a dog-only property.

Assumption 3. The premises in an inductive argument (Category A_1 has Property P, etc.) serve as new evidence, and they are used to strengthen or weaken the hypotheses about Property P using Bayes's Theorem. Again, the goal of induction is to come up with a better estimate of the range of Property P, and this process may be described in terms of standard Bayesian revision mechanisms applied to the hypotheses already under consideration. To give a simple example, if a person learns that dogs have substance P in their blood, and that sheep also have substance P in their blood, then the hypothesis that this property is true only for dogs must be ruled out. The next section shows in more detail how these calculations would be carried out.

The Bayesian Model

It is helpful to present the model in the context of a simple example. Let's say that a person is evaluating inductive arguments concerning just two categories of animals, cows and horses. People know quite a few properties of animals, but these known properties must fall into four types: properties that are true of cows and horses, properties that are true of cows but not horses, properties that are true of horses but not cows, and properties that are not true of either cows or horses. These four types of known properties can serve as four hypotheses when reasoning about novel properties, because any new property must also be one of these four types. (The model also has been extended to deal with properties that are true for proportions of cases, such as a property that is true for 10% of all cows, but this issue is not considered directly in this chapter.) These four types of properties are listed in Table 1, with cases where the property is true for a category shown in boldface for emphasis.

Table 1

Hypothesis No	Range	Degree of Prior Belief
1	**Cow -> True** **Horse -> True**	.70
2	**Cow -> True** Horse -> False	.05
3	Cow -> False **Horse -> True**	.05
4	Cow -> False Horse -> False	.20

As shown in Table 1, a person would have prior beliefs about these hypotheses. For example, the value of .70 for hypothesis 1 represents the belief that there is a 70% chance that a new property would be true of both cows and horses. This high value could reflect the high degree of similarity between cows and horses, and that people know many other animal properties that are true of both cows and horses. (The particular numbers are used only for illustration at this point.) However, the person might see a 5% chance that a new property would be true of cows and not horses, a 5% chance that a new property would be true of horses and not cows, and a 20% chance that the property is true of neither category. Note that because the four hypotheses are exhaustive and mutually exclusive, their corresponding prior beliefs add up to 1.

This table describes prior beliefs not only about the four hypotheses but also about the two categories. By combining hypotheses 1 and 2, it appears that the person believes there is a 75% chance that cows would have the new property, and likewise by combining hypotheses 1 and 3, the person believes there is a 75% chance that horses have the new property.

At this point, only prior beliefs have been described, embodying Assumptions 1 and 2 above. The next step is to combine these prior beliefs with new evidence, using Bayes's Theorem according to Assumption 3. To continue with this simple example, consider the following inductive argument:

```
Cows have Property P
-----------------------
Horses have Property P.
```

The premise, "Cows have Property P," is used to update beliefs about the four hypotheses, so that a better evaluation of the conclusion, "Horses have Property P," may be achieved. In applying

Bayes's Theorem (Equation 1), the premise is treated as the data, D. The prior degree of belief in each hypothesis i is indicated by $P(H_i)$. (Note that there are four hypotheses, so n = 4 here.) The task is to estimate $P(H_i|D)$, that is the posterior degree of belief in each hypothesis given the data.

$$P(H_i|D) = \frac{P(H_i)P(D|H_i)}{\sum_{j=1}^{n} P(H_j)P(D|H_j)} \quad (1)$$

In Table 2, the calculations are shown for all four hypotheses, given the data that Cows have Property P. The prior beliefs, $P(H_i)$, are simply copied from Table 1. The calculation of $P(D|H_i)$ is quite easy. Under hypotheses 1 and 2, cows have the property in question, so obtaining the data (that cows have the property) has a probability of 1. But under hypotheses 3 and 4, cows do not have the property, so the probability of obtaining the data must be 0 under these hypotheses. The final column, indicating the posterior beliefs in the four types of properties, has been calculated using Equation 1. Notably, hypothesis 1, that cows and horses have the property, and hypothesis 2, that just cows have the property, have been strengthened. The two remaining hypotheses have been eliminated from contention because they are inconsistent with the data or premise that cows have the property.

Table 2

Hypothesis No	Range	Degree of Prior Belief $P(H_i)$	$P(D\|H_i)$	Posterior Belief $P(H_i\|D)$
1	**Cow -> True** **Horse -> True**	.70	1	.93
2	**Cow -> True** Horse -> False	.05	1	.07
3	Cow -> False **Horse -> True**	.05	0	.00
4	Cow -> False Horse -> False	.20	0	.00

Finally, the values in Table 2 may be used to evaluate the conclusion, that horses have Property P. The degree of belief in this conclusion is simply the sum of the posterior beliefs for hypotheses 1 and 3, or .93. Recall that before the introduction of evidence that cows have the property, the prior belief that horses have the property was only .75. Thus, the premise that cows have the property led to an increase in the belief that horses have the property.

Some initial considerations. This example was meant to be a simple illustration of how Bayes's Theorem can be applied to evaluating an inductive argument. At least two issues are raised immediately by this example. The first issue is what else can the Bayesian analysis accomplish beyond this simple example? The following sections of this chapter contain a number of further applications of the model to interesting, well-documented, and more complex phenomena in inductive reasoning. The benefit of the Bayesian analysis is that a fairly simple conceptualisation of inductive reasoning facilitates the explanation of a number of experimental results.

The second issue concerns the prior beliefs, such as the numbers in the third columns of Tables 1 and 2. How might these priors be derived? Are the exact values of the priors important? These questions are fundamental issues for Bayesian statistics (Box & Tiao, 1973; Raiffa & Schlaifer, 1961), and no easy answers will be provided here. For initial purposes of exposition, it will be assumed that the priors are determined by the number of known properties of each type that are brought to mind in the context of evaluating the inductive argument. It might be said that the prior beliefs for new properties are estimated using something like an availability heuristic (Tversky & Kahneman, 1973) based on known properties. The basic idea is that when reasoning about novel animal properties, people would retrieve a set of familiar animal properties from memory. Then they would count up how many known properties are consistent with each of the four properties, e.g., how many known properties of animals are true of both cows and horses. The priors in Tables 1 and 2, for example, are consistent with the idea that 20 known properties are brought to mind: 14 of type 1, 1 of type 2, 1 of type 3, and 4 of type 4.

Furthermore, it will be shown a number of times in this chapter that the exact values for the prior beliefs are not critical. For instance, in the present example, the initial degree of belief in hypothesis type 4, that neither cows nor horses have the property, was not at all important. The posterior belief in hypothesis 1, $P(H_1|D)$, can be calculated simply from the prior beliefs in hypotheses 1 and 2, $P(H_1|D) = P(H_1) / (P(H_1) + P(H_2))$, or $.93 = .70 / (.70 + .05)$. The posterior belief in hypothesis 1 would be the same regardless of the value of $P(H_4)$, as long as $P(H_1)$ and $P(H_2)$ maintain the same ratio to each other. The issue of derivation of priors will be returned to at several points in this chapter.

Basic Phenomena in Inductive Reasoning

This section will address three well-documented results in inductive reasoning: similarity, typicality, and diversity effects. What these results have in common is that they depend on the categories in the inductive arguments and not so much on the property being projected. These three phenomena were identified by Rips (1975) and Osherson, Smith, Wilkie, Lopez, & Shafir (1990; see Osherson, Stern, Wilkie, Stob, & Smith, 1991, for further evidence), and they are so close to the core of inductive reasoning that any complete model of induction would need to address them.

Similarity Effects

The most widespread and robust finding in inductive reasoning is a basic similarity effect. For example, “Cows have Property P, therefore Sheep have Property P” is a stronger argument than “Cows have Property P, therefore Ferrets have Property P.” (Osherson et al., 1990; Rips, 1975). People are more willing to project a property of cows to a similar category (sheep) than to a less similar category (ferrets). In reasoning about these three categories, there are eight possible hypotheses about Property P, as shown in Table 3.

Table 3

Hypothesis No	Range	Degree of Prior Belief $P(H_i)$	$P(D\|H_i)$	Posterior Belief $P(H_i\|D)$
1	**Cow -> True** **Sheep -> True** **Ferret -> True**	.40	1	.67
2	**Cow -> True** Sheep -> False Ferret -> False	.05	1	.08
3	Cow -> False **Sheep -> True** Ferret -> False	.05	0	.00
4	Cow -> False Sheep -> False **Ferret -> True**	.25	0	.00

5	**Cow -> True** **Sheep -> True** Ferret -> False	.15	1	.25
6	**Cow -> True** Sheep -> False **Ferret -> True**	.0	1	.00
7	Cow -> False **Sheep -> True** **Ferret -> True**	.0	0	.00
8	Cow -> False Sheep -> False Ferret -> False	.10	0	.00

The prior beliefs in the third column were chosen to illustrate prior knowledge about cows, sheep and ferrets. Because cows, sheep, and ferrets are all mammals, and thus would share many anatomical properties, a fairly high prior belief for hypothesis 1 (.40) was chosen. Because cows and sheep are similar in other ways that differ from ferrets, a relatively high value for hypothesis 5 (.15) was chosen. And since cows and ferrets do not have many (or any?) properties in common which are not shared by sheep as well, a value of 0 was assigned to hypothesis 6. The other values in column 3 are not critical to this illustration of the similarity effect.

What happens when the premise is given, that cows have Property P? As shown in column 4, some of the hypotheses are inconsistent with this evidence. For example, hypothesis 3, that the new property is a sheep-only property, can be ruled out from the premise that cows have the property. It would be impossible to observe the data that cows have the property under hypothesis 3. After applying Bayes's Theorem, the posterior beliefs shown in column 5 are obtained. Only three hypotheses are left standing: It is a property of all three animals (hypothesis 1), it is a cow-only property (hypothesis 2), or it is a cow-and-sheep property (hypothesis 5).

From these values in column 5, estimates can be calculated for the two conclusions under consideration. The probability that sheep have Property P is .67 + .25 = .92, and the probability that ferrets have the property is just .67. Hence a similarity effect is obtained: Asserting the premise that cows have some property leads to a stronger conclusion for a similar category, sheep , than for a less similar category, ferrets.

Does the similarity effect depend on the particular values of prior beliefs shown in column 3 of Table 3? In fact, a similarity effect would be predicted for a wide range of values. Table 4 shows the same information as in Table 3, presented symbolically rather than numerically. Only the prior beliefs that are critical to the calculations are shown.

Table 4

Hypothesis No	Range	Degree of Prior Belief $P(H_i)$	$P(D \mid H_i)$	Posterior Belief $P(H_i \mid D)$
1	**Cow -> True** **Sheep -> True** **Ferret -> True**	a	1	a/(a+b+c+d)
2	**Cow -> True** Sheep -> False Ferret -> False	b	1	b/(a+b+c+d)
3	Cow -> False **Sheep -> True** Ferret -> False			
4	Cow -> False Sheep -> False **Ferret -> True**			
5	**Cow -> True** **Sheep -> True** Ferret -> False	c	1	c/(a+b+c+d)
6	**Cow -> True** Sheep -> False **Ferret -> True**	d	1	d/(a+b+c+d)
7	Cow -> False **Sheep -> True** **Ferret -> True**			
8	Cow -> False Sheep -> False Ferret -> False			

The posterior belief that sheep have the property is (a+c) / (a+b+c+d), whereas the posterior belief that ferrets have the property is (a+d) / (a+b+c+d). Clearly, people will make a stronger inference about sheep than ferrets whenever c is greater than d. In other words, people will be more willing to project a novel property from cows to sheep than from cows to ferrets whenever people already know more common properties for cows and sheep than for cows and ferrets. This assumption seems rather plausible and it suggests that similarity effects will be predicted for a very broad range of prior beliefs.

Evaluation. Of course, the Bayesian model's predictions extend beyond cows, sheep, and ferrets. For any three categories X, Y, and Z, the model predicts stronger projections of a new property from X to Y than from X to Z to the extent that X and Y have more known properties in common than X and Z. The Bayesian model simply implements what Mill (1874) characterised as the "conviction that the future will resemble the past." It will be shown in the remaining examples in this chapter that this simple idea can help explain a considerable range of phenomena in inductive reasoning beyond the simple similarity effect.

Typicality Effects

Whereas similarity effects in induction are highly intuitive and perhaps even obvious, the remaining effects, including typicality effects, are more subtle. Demonstrations of typicality effects in induction indicate that when it comes to premises, not all categories are created equal. Some categories, when used in premises in an inductive argument, have a greater impact than other categories. To used Goodman's (1955) term, it could be said that typical categories have greater entrenchment than atypical categories (see also Rips, 1975; Shipley, 1993). Consider the following two arguments:

```
Cows have Property P
---------------------------
All Mammals have Property P
```

and

```
Ferrets have Property P
----------------------------
All Mammals have Property P.
```

The first argument, with a typical premise category, seems stronger than the second, with an atypical premise category. There is considerable evidence that typicality effects are widespread. For example, Rips (1975) looked at a large number of inductive arguments involving pairs of mammals or birds, and found that the typicality of the premise category was a significant predictor of the overall strength of the argument. Osherson et al. (1990), among others, have obtained similar results, with adult subjects, and Carey (1985) and Lopez, Gelman, Gutheil, and Smith (1992) have obtained typicality effects with children.

One reason for these typicality effects could be that typical categories, such as cows, are representative of their superordinates, whereas atypical categories, such as ferrets, are believed to be idiosyncratic compared to their superordinates. Hence a novel property of cows may well be a property of all mammals, whereas a novel property of ferrets could be just another idiosyncratic ferret property. These beliefs about typical and atypical categories are illustrated in Table 5, with "Mammal" referring to All Mammals.

Table 5

Hyp No	Range	Degree of Prior Belief $P(H_i)$	$P(Cow \mid H_i)$	Posterior Belief $P(H_i \mid Cow)$	$P(Ferret \mid H_i)$	Posterior Belief $P(H_i \mid Ferret)$
1	**Cow->True** **Ferret->True** **Mammal->True**	.5	1	.71	1	.63
2	**Cow->True** **Ferret->True** Mammal->False	.1	1	.14	1	.13
3	**Cow->True** Ferret->False Mammal->False	.1	1	.14	0	.00
4	Cow->False **Ferret->True** Mammal->False	.2	0	.00	1	.25
5	Cow->False Ferret->False Mammal->False	.1	0	.00	0	.00

Note that there are only five logical possibilities, e.g., if all mammals have some property, then it must be true of cows and ferrets as well. The critical number here is the prior belief in hypothesis 4, in column 3. Here people would have a relatively strong belief that ferrets are idiosyncratic compared to cows and other mammals. In other words, people believe that ferrets have properties that make them different from other mammals (perhaps in terms of size and how they move). In comparison, the prior belief for hypothesis 3 is lower, suggesting that people have less of a belief that cows differ from other mammals. The fifth column of the table shows the posterior beliefs in the five hypotheses, given the premise that cows have some novel property. Likewise, the seventh column shows the posterior beliefs, given the premise that ferrets have a novel property. The beliefs in the conclusion, that all mammals have the property, are indicated by the posterior belief in hypothesis 1. Given that cows have the property, the model predicts a 71% chance that all mammals have the property. But given that ferrets have the property, the model predicts only a 63% chance that all mammals have the property. Because of the strong prior belief that ferrets have idiosyncratic properties, it is more difficult to project a novel property of ferrets to other animals.

With some algebra, it can be shown that the Bayesian model predicts typicality effects in a very wide range of situations, and the effect illustrated here does not depend on the particular numbers chosen for the prior beliefs. Indeed, cows will be seen as stronger premise categories than ferrets whenever the prior belief in hypothesis 4 is greater than the prior belief in hypothesis 3. Going beyond cows and ferrets, the Bayesian model predicts that a category will make a particularly poor premise whenever people believe that the category is so atypical that it already has a number of idiosyncratic properties, so that some new property of the category may well also be an idiosyncratic property.

Evaluation. The Bayesian model is able to explain the weakness of projections from atypical categories such as ferret by resorting to people's prior beliefs that atypical categories have other idiosyncratic properties that make them different from other categories. This conjecture may seem plausible for an animal such as ferret, which could be as encoded as being a mammal that is particularly small, elongated, and fast-moving. However, this account must assume that people know more properties (or at least more idiosyncratic properties) of atypical categories than typical categories. Do people really know more properties of ferrets than of familiar animals such as dogs?

What is critical to the model is not the actual number of properties known as much as the strengths of beliefs. A person might have a very strong belief that ferrets are odd, idiosyncratic mammals, without being able to give a long list of ferret properties. In terms of Table 5, people would have a high prior belief in hypothesis 4 which may go beyond known properties to also reflect a "placeholder belief" (cf., Medin & Ortony, 1989). This placeholder represents the strong belief that ferrets have a number of idiosyncratic properties that may be presently unknown to the person doing the reasoning. A person might think "Ferrets are really odd animals, but I can't possibly state all the reasons why." Recently, there has been considerable discussion of the possibility that people's beliefs about categories go beyond knowledge of properties to beliefs about category essences that do not easily correspond to listable properties (e.g., Gelman, Coley, & Gottfried, 1994). The Bayesian analysis builds on this essentialist account, with the added assumption that atypical categories have more idiosyncratic essences than typical categories. Clearly, it would be useful to evaluate this assumption further, in future empirical work.

Finally, it is interesting to point out the ties between the model's account for similarity effects and the model's account for typicality effects. The model predicts that people will make relatively weak inferences between dissimilar categories, such as cows and ferrets, whenever people know other properties that are true of cows but not true of ferrets. Likewise, the model predicts relatively weak inferences from an atypical category, such as ferrets, to all mammals, because ferrets are thought to have other properties that are not true of all mammals. To explain each effect, what is critical is the idiosyncrasy of the premise category relative to the conclusion category. If the premise category has known properties that are idiosyncratic (not true of the conclusion category), then people will be relatively reluctant to project a novel property from the premise to the conclusion.

Diversity Effects

One of the fascinating findings in inductive reasoning is that while a typical category can make a strong premise, two typical categories together can make a worse case than one typical category and one atypical category together. Consider the following arguments:

```
Cows have Property P
Horses have Property P
---------------------------
All Mammals have Property P
```

and

```
Cows have Property P
Ferrets have Property P
----------------------------
All Mammals have Property P.
```

People tend to find arguments such as the one with cows and ferrets stronger; arguments with premises that contain a diverse set of categories are favoured over arguments that contain a narrow range of categories (Osherson et al., 1990). In the present example, cows and horses together make a weaker case even though horses are more typical than ferrets. Intuitively, cows and horses are so similar that finding out that horses have some property adds little information when you already know that cows have the property. On the other hand, getting information about ferrets seems to tell you something that you don't already know.

The Bayesian analysis of the diversity effect is illustrated in Table 6. With four different categories under consideration, there are sixteen (2^4) types of hypotheses, but only four of them are relevant to the present example. The remaining hypotheses must have a zero prior belief due to logical necessity, or these hypotheses do not affect the evaluation of the conclusion that all mammals have property P. The most critical prior belief in this illustration is the relatively high belief in hypothesis 3, indicating a high number of known properties common to cows and horses.

When Bayes's Theorem is applied, a compound event of two premises is treated as the data, D. For the first argument, D is Cows have Property P AND Horses have Property P. And for the second argument, D is Cows have Property P AND Ferrets have Property P. For example, column 4 refers to the event that both cows and horses have the property in question. The fifth and seventh columns show the posterior estimates given the two different sets of premises. The Bayesian analysis predicts a stronger conclusion that all mammals have the property given cows and ferrets (.67) compared to being given cows and horses (.57).

Table 6

Hyp No	Range	Degree of Prior Belief $P(H_i)$	$P(Cow\&Horse \mid H_i)$	Posterior Belief $P(H_i \mid Cow\&Horse)$	$P(Cow\&Ferret \mid H_i)$	Posterior Belief $P(H_i \mid Cow\&Ferret)$
1	**Cow->True** **Horse->True** **Ferret->True** **Mammal->True**	.4	1	.57	1	.67
2	**Cow->True** **Horse->True** **Ferret->True** Mammal->False	.1	1	.14	1	.17
3	**Cow->True** **Horse->True** Ferret->False Mammal->False	.2	1	.29	0	.00

4	**Cow->True** Horse->False **Ferret->True** Mammal->False	.1	0	.00	1	.17

In brief, the Bayesian account predicts diversity effects because it is unlikely that if two very different categories share a property, that they will be the only categories to have this property. The properties shared by cows and ferrets, such as having fur and being warm-blooded, tend to be true of other mammals as well. Hence a novel property of cows and ferrets seems likely to be distributed the same way, and thus true of other mammals. In contrast, two very similar categories are more likely to have some idiosyncratic similarities. For example, there are properties of cows and horses, such as "typically raised on a farm," that are not true of other mammals. Hence there is a good chance that a novel property of cows and horses will likewise not extend to all mammals.

Again, this prediction of a diversity effect is highly robust across different assumptions about prior beliefs and different specific numbers. The only constraint is that the prior belief in hypothesis 3 must be higher than the prior belief in hypothesis 4. In other words, people must believe that cows and horses have more properties in common than do cows and ferrets. This constraint just about embodies the idea of diversity of categories, hence the Bayesian model predicts quite generally that less diverse sets of premises will make weaker arguments than more diverse sets of premises.

Evaluation. The Bayesian account of the diversity effect illustrates how this simple model can explain increasingly complex phenomena beyond the basic similarity effect, and its variant, the typicality effect. No new assumptions are needed by the model to account for diversity effects. Indeed, cows and horses in the premise make a weak argument here for the same reason that ferrets in the premise made a weak argument in the previous section. An atypical category (e.g., ferret) is likely to have known idiosyncratic properties, so it is hard to project a novel property from this category. Likewise, two very similar categories (e.g., cow and horse) are likely to have some idiosyncratic properties in common, so it would be relatively hard to project a novel property from this pair of categories in the premise of an inductive argument.

Diversity effects have turned up quite robustly for adult subjects (Osherson et al., 1990), but, interestingly, Gutheil and Gelman (1997) and Lopez et al. (1992) have reported that young children do not show diversity effects. This lack of diversity effect poses a problem for the Bayesian account, which makes a quite strong prediction here. There is no easy way for the Bayesian model to not predict diversity effects. One possible avenue for explaining young children's lack of diversity effects would be to examine their prior knowledge about animals, which could differ from that of adults (and hence children's inductions would differ as well). Another possible attack on this issue would be to question young children's ability to perform complex tasks such as evaluating logical arguments with a number of premises. Perhaps arguments with multiple premises could cause memory problems for children. In such cases, one or more premise categories might be ignored or forgotten, so that diversity effects could not be a possibility. Perhaps further experiments could be devised that make the induction task easier, so that young children might have fewer performance problems and possibly show diversity effects.

Phenomena Involving Meaningful Properties

The preceding examples have focused on how induction is influenced by the categories in an inductive argument, e.g., whether there is a typical category in the premise such as cow or an atypical category such as ferret. However, just as all categories are not created equal in inductive reasoning, not all properties are equal either: People reason differently depending on what Property P actually is. Some previous work (e.g., Rips, 1975; Osherson et al., 1990), while documenting a number of interesting category-based phenomena, has only looked at fairly unfamiliar, or "blank," anatomical properties of animals, such as "has sesamoid bones" or "has BCC in its blood." For such unfamiliar properties, it seems plausible that the property itself has little influence on reasoning. But inductive reasoning with blank properties is only one facet of inductive reasoning more broadly considered, in which the property itself plays an important role. The remaining examples in this chapter show some influences of properties on induction. Importantly, the Bayesian model provides a novel framework for addressing these phenomena.

Projectible Versus Non-Projectible Properties

The first property-based phenomenon to be described is a classic in inductive reasoning. Goodman (1955), in posing his riddle of induction, noted that some properties are more projectible, or more easily projected, than other properties. Some properties seem particularly transient or idiosyncratic, and they seem unlikely to project from one category or individual to another. (Again, perception of idiosyncrasy comes up as a crucial issue in induction.) To use an example from

Gelman (1988), imagine that you see a rabbit eating alfalfa. You can probably conclude that other rabbits would like to eat alfalfa as well. But if you see a rabbit that is dirty, you probably would not conclude that other rabbits are dirty as well. What makes "eats alfalfa" a more projectible property than "is dirty"? The Bayesian model itself does not answer this question, but it does provide a way to represent information about projectibility and use it in reasoning.

Table 7 illustrates four hypotheses about a pair of rabbits and whether each one likes to eat alfalfa. (Note that here the illustration shifts to inferences about individuals, whereas the previous examples referred to inferences about categories. The model is applied in the same manner in either case.) In a way, the prior beliefs in column 3 represent a great deal of ignorance. For example, the prior belief that Rabbit 1 likes to eat alfalfa is .5 (the sum of hypotheses 1 and 2) and likewise the prior belief that Rabbit 1 does not like to eat alfalfa is .5 (sum of hypotheses 3 and 4). Similarly, people see a 50% chance that Rabbit 2 likes to eat alfalfa and a 50% chance that it does not. However, the prior beliefs in column 3 also encode a dependence between the two rabbits. From summing up hypotheses 1 and 4, there is an 80% chance that the two rabbits will be alike, either both eating alfalfa or both not eating it. Thus, by enumerating the four possible types of hypotheses and prior beliefs in each one, it is possible to represent the belief that "I don't know whether rabbits eat alfalfa, but probably they are all alike in their food preferences."

Table 7

Hypothesis No	Range of Eats Alfalfa	Degree of Prior Belief $P(H_i)$	P(Rabbit 1 Eats Alfalfa$\|H_i$)	Posterior Belief $P(H_i\|$Rabbit 1 Eats Alfalfa)
1	**Rabbit 1 -> True** **Rabbit 2 -> True**	.4	1	.8
2	**Rabbit 1 -> True** Rabbit 2 -> False	.1	1	.2
3	Rabbit 1 -> False **Rabbit 2 -> True**	.1	0	.0
4	Rabbit 1 -> False Rabbit 2 -> False	.4	0	.0

In this situation, finding out that Rabbit 1 likes to eat alfalfa is quite informative. When Bayes's Theorem is applied after this premise is considered, hypotheses 3 and 4 are ruled out entirely, and the remaining hypotheses are strengthened. The belief that Rabbit 2 likes to eat alfalfa increases from .5 to .8.

In contrast, consider a representation of beliefs for "is dirty," shown in Table 8. Just as in Table 7, this table shows a prior belief that there is a 50% chance that Rabbit 1 has the property in question (sum of hypotheses 1 and 2) and that there is a 50% chance that Rabbit 2 has the property (sum of hypotheses 1 and 3). However, unlike Table 7, the third column of Table 8 suggests that the property in question is distributed evenly among the four hypotheses. For "is dirty" in Table 8, people see no contingency between what Rabbit 1 has and what Rabbit 2 has, but for "eats alfalfa" in Table 7, people see a positive contingency between the two individuals. (It would also be possible to represent intermediate degrees of contingency between these two possibilities, or even to represent a greater degree of contingency than shown by Table 7.) Given the premise that Rabbit 1 is dirty and applying the Bayesian model, it can be seen that the posterior belief that Rabbit 2 is dirty is the same as the prior belief: 50%. In this case, finding out information about one individual having a non-projectible property does not change beliefs about another individual, whereas in the case of the projectible property "eats alfalfa," finding out about one individual does change beliefs about another. (See Nisbett, Krantz, Jepson, and Kunda, 1983, for other examples of projectibility effects.)

Table 8

Hypothesis No	Range of Is Dirty	Degree of Prior Belief $P(H_i)$	P(Rabbit 1 Is Dirty$\|H_i$)	Posterior Belief $P(H_i\|$Rabbit 1 Is Dirty)
1	**Rabbit 1 -> True** **Rabbit 2 -> True**	.25	1	.5
2	**Rabbit 1 -> True** Rabbit 2 -> False	.25	1	.5
3	Rabbit 1 -> False **Rabbit 2 -> True**	.25	0	.0
4	Rabbit 1 -> False Rabbit 2 -> False	.25	0	.0

Evaluation. The point of this example was simply that the Bayesian account can represent information about projectibility of properties, and that the model makes sensible predictions using this information. As will be noted in the General Discussion, other models of inductive reasoning do not incorporate information about projectibility at all, so the Bayesian model does provide an advance over past work. The crucial insight provided by the Bayesian account is that beyond prior beliefs about whether or not an individual such as Rabbit 2 has a property, the distribution of beliefs across hypotheses is important. In particular, the distributional representation maintains beliefs about contingencies between different categories.

Still, it would be desirable to know how these prior beliefs concerning projectibility come about. Why does the model have different distributions of prior beliefs for “eats alfalfa” and “is dirty”? Following the spirit of Bayesian accounts, it is assumed that these priors will largely come from past observations. For example, people might know or believe that food preferences are fairly uniform within categories of animals, e.g., cows eat grass and chickens don’t eat meat. Therefore, the priors for another food preference, eating alfalfa, in Table 7 would reflect the belief that this property is likely to be uniform across rabbits, so that any two rabbits will either both eat alfalfa or both not eat alfalfa. Similarly, people have experience with transient or idiosyncratic properties like “is dirty.” They may have seen a dirty cat sitting near a clean cat. This experience that “is dirty” is not uniform within animal categories could provide the priors shown in Table 8.

This explanation for how priors come about is heavily memory-based, but the process of coming up with priors would involve reasoning as well as memory retrieval. Consider the property “has a blue cotton ball sitting on its head.” People would probably believe this property to be non-projectible, that is, seeing that one individual has a blue cotton ball on its head does not help you to make inferences about another individual. Chances are, people would not have had any direct observations of this property in the past, so the derivation of prior beliefs for this property would necessarily include some reasoning to determine what prior knowledge is relevant to the property. Goodman (1955) has suggested that we also have abstract beliefs, called overhypotheses, describing the scope of properties such as food preferences and transient properties (see also Shipley, 1993). An example of an overhypothesis would be “different kinds of animals eat characteristic foods.” Such overhypotheses could also be used as a source of priors. Thus, reasoning about particular properties is likely to be influenced by specific knowledge of known, individual properties as well as by more general, abstract knowledge of classes of properties (Heit, 1997).

Different Kinds of Properties

Although it might seem from the previous section that some properties (the projectible ones) are good for inductive reasoning and other properties (the non-projectible ones) do not promote inductive reasoning, the picture is actually more complicated and more interesting. Depending on the argument, that is, depending on the categories in an inductive argument, a particular property may be projectible or non-projectible or somewhere in between.

Consider the following example, from Heit and Rubinstein (1994). For a typical blank anatomical property, such as "has a liver with two chambers," people will make stronger inferences from chickens to hawks than from tigers to hawks. Because chickens and hawks are from the same biological category, and share many internal properties, people are quite willing to project a novel anatomical property from one bird to another. But since tigers and hawks differ in terms of many known internal biological properties, it seems less likely that a novel anatomical property will project from one to the other. This result illustrates the priority of biological categories that has been observed in inductive reasoning (e.g., Carey, 1985; Gelman, 1988).

However, now consider the behavioural property "prefers to feed at night." Heit and Rubinstein (1994) found that inferences for behavioural properties concerning feeding and predation were weaker between the categories chicken and hawk than between the categories tiger and hawk--the opposite of the result for anatomical properties. Here, it seems that despite the considerable biological differences between tigers and hawks, people were influenced by the known similarities between these two animals in terms of predatory behaviour, thus making strong inferences about a novel behavioural property. In comparison, chickens and hawks differ in terms of predatory behaviour (with chickens tending to be pacifists), thus people were less willing to project a novel behavioural property between these two animals. Putting together these results, it seems that each property is more projectible for a different pair of animals. It is not simply the case that some properties are always more projectible than other properties.

This pattern of results can be derived in a straightforward manner using the Bayesian model. For ordinary anatomical properties such as "has a liver with two chambers," the model would predict an overall similarity effect, i.e., inferences would be stronger from hawks to chicken than from tigers to chickens. (See the earlier section on Similarity Effects, and Table 3, for a comparable example.) It is assumed that in reasoning about a novel anatomical property, the prior hypotheses would reflect the greater number of shared biological properties between the two birds, chickens and hawks, and thus the novel property would be projected more readily from chickens to hawks than from tigers to hawks.

In contrast, when reasoning about a novel behavioural property such as "prefers to feed at night," a person would derive priors based on other known behavioural properties. The critical assumption for the Bayesian analysis is that different properties will recruit different prior beliefs. In the previous section on projectibility, it was assumed that different distributions of prior hypotheses would be derived for "eats alfalfa" and "is dirty." Likewise, here it is assumed that for reasoning about a novel anatomical property, known anatomical properties will be used to derive the priors, whereas for a novel behavioural property, known behavioural properties will be considered. For example, a person might consider known facts concerning whether various species prey on weaker animals, and whether they use sharp claws for attacking, to derive priors like those illustrated in Tables 9 and 10.

Table 9 illustrates a situation where a person has retrieved a fairly large number of behavioural properties held in common by tigers and hawks, leading to a strong prior belief in hypothesis 1, that tigers and hawks have the property in common. In Table 10, there is a somewhat weaker belief in hypothesis 1, that chickens and hawks have the property in common. (Note that chickens and hawks do have a number of birdlike behaviours in common, even though they differ in terms of predation.)

Table 9

Hypothesis No	Range	Degree of Prior Belief $P(H_i)$	$P(Tigers \mid H_i)$	Posterior Belief $P(H_i \mid Tigers)$
1	**Tiger -> True** **Hawk -> True**	.50	1	.77
2	**Tiger -> True** Hawk -> False	.15	1	.23
3	Tiger -> False **Hawk -> True**	.15	0	.0

Table 10

Hypothesis No	Range	Degree of Prior Belief $P(H_i)$	$P(\text{Chickens}\mid H_i)$	Posterior Belief $P(H_i\mid \text{Chickens})$
1	**Chicken -> True** **Hawk -> True**	.40	1	.62
2	**Chicken -> True** Hawk -> False	.25	1	.38
3	Chicken -> False **Hawk -> True**	.25	0	.0

As shown in Tables 9 and 10, the Bayesian account would predict that after the premise, that tigers prefer to feed at night, is given, there would be a fairly strong posterior belief in the conclusion that hawks act the same way (.77). In contrast, the inference from chickens to hawks would be weaker, only .62. Hence, the Bayesian model can predict the pattern of results obtained by Heit and Rubinstein (1994) for anatomical and behavioural properties.

Evaluation. In accounting for this result, as well as the results on projectibility, the Bayesian account must rely on different prior beliefs for different properties. These examples are valuable because they show how inductive reasoning is influenced not only categories but by properties as well as the interaction between the two. One way of describing the results of Heit and Rubinstein (1994) is that for reasoning about anatomical properties, people favour anatomically-similar pairs of animals, such as chickens and hawks, but for reasoning about behavioural properties, people favour behaviourally-similar pairs of animals, such as tigers and hawks. As will be noted in the General Discussion, other models of inductive reasoning cannot explain this result because they rely on a single measure of similarity. In contrast, the Bayesian model predicts this pattern of results naturally, with the assumption that when reasoning about a particular kind of novel property (anatomical or behavioural), familiar properties of the same kind are recruited in order to set up the priors.

Of course, the Bayesian account is somewhat incomplete in that it does not describe this recruitment process itself. However, it does seem extremely plausible that people would be able to perform the steps in reasoning assumed by this Bayesian account. For example, it seems quite plausible that people could determine that “prefers to eat at night” concerns an animal’s behaviour, and that they could retrieve prior knowledge about other behaviours of animals.

Properties Depending on Differences Rather Than Similarity

All of the previous examples in this chapter were consistent with the principle that similarity promotes inductive reasoning. Given the premise that a category has some property, it will seem likely that another, similar category has that property as well. But for some properties, and some categories, similarity seems to block inferences rather than promote them. For example, at the time of writing this chapter, Chelsea were set to face Middlesbrough for the FA Cup, in the championship match of the English football season. Chelsea were favoured to win by most football supporters, but these people had an additional kind of prior knowledge as well. As illustrated by Table 11, the two events, "Chelsea wins" and "Middlesbrough wins" are mutually exclusive. Therefore there is zero prior belief in hypothesis 1, that both teams win, and likewise there is zero prior belief in hypothesis 4. In this situation, the information that Chelsea wins would be extremely informative. After application of Bayes's Theorem, the conclusion would be wholly in favour of hypothesis 2, that Middlesbrough would not win as well, as shown in the fifth column. The premise that Chelsea wins must rule out hypothesis 3, that Middlesbrough wins. Here, information that one football team has the property "wins the FA Cup in 1997" promotes the inference that another football team does not have this property. It is interesting to compare the priors in Table 11 to the priors in Table 7, which show a positive rather than negative contingency between two events.

Table 11

Hypothesis No	Range	Degree of Prior Belief P(H_i)	P(Chelsea|H_i)	Posterior Belief P(H_i|Chelsea)
1	**Chelsea -> True** **Middlesbrough -> True**	0	1	0
2	**Chelsea -> True** Middlesbrough -> False	.6	1	1
3	Chelsea -> False **Middlesbrough -> True**	.4	0	0
4	Chelsea -> False Middlesbrough -> False	0	0	0

Smith, Shafir, & Osherson (1993) provided a more complex and subtle example where inferences go in the opposite direction of what overall similarity would predict. Consider the following two arguments:

```
Poodles can bite through barbed wire
---------------------------------------------
German Shepherds can bite through barbed wire
```

and

```
Dobermans can bite through barbed wire
----------------------------------------------
German Shepherds can bite through barbed wire.
```

Clearly there is greater similarity between Dobermans and German shepherds than between poodles and German shepherds. Yet people find the first argument stronger than the second. An informal way to justify this reasoning is that if poodles, a rather weak and tame kind of dog, can bite through barbed wire, then obviously German shepherds, which are much stronger and more ferocious, must be able to bite through barbed wire as well. This property, “can bite through barbed wire,” seems to depend on the magnitude of other dimensions such as strength and ferocity.

In Table 12, prior beliefs about this property and the three kinds of dogs are illustrated. These prior beliefs in the third column reflect a considerable degree of uncertainty about this property, in that the priors are fairly widely distributed over the eight possible types of hypotheses. Most critically, however, people would have little or no prior belief in hypothesis 2, that the weak poodles can bite through wire and the stronger dogs cannot. Thus, as shown in the fifth column, the premise that poodles can bite through barbed wire would strongly promote the inference that the other dogs can do so as well. The posterior estimate that German shepherds can bite through barbed wire would be .84, adding up the values for hypotheses 1 and 6.

In contrast, if a person is given the premise that Dobermans can bite through wire, this may well be a property of Dobermans alone. The posterior estimate for hypothesis 3, that Dobermans alone can bite through wire, would remain at .18 after application of Bayes’s Theorem to derive the seventh column. This hypothesis would compete with the others, and make it more difficult to conclude that German shepherds can bite through barbed wire. The posterior estimate for German shepherds having this property would be .72, taking the sum of the values in the seventh column for hypotheses 1 and 7.

Table 12

p o	Range	Degree of rior Belief $P(H_i)$	P(Poodle| H_i)	sterior Belief $P(H_i$| Poodle)	P(Doberman| H_i)	Posterior elief $P(H_i$| Doberman)
	Poodle->True **Doberman->True** **G. Shepherd->True**	.20	1	.67	1	.36
	Poodle->True Doberman->False G. Shepherd->False	.0	1	.00	0	.00
	Poodle->False **Doberman->True** G. Shepherd->False	.10	0	.00	1	.18
	Poodle->False Doberman->False **G. Shepherd->True**	.10	0	.00	0	.00
	Poodle->True **Doberman->True** G. Shepherd->False	.05	1	.17	1	.09
	Poodle->True Doberman->False **G. Shepherd->True**	.05	1	.17	0	.00
	Poodle->False **Doberman->True** **G. Shepherd->True**	.20	0	.00	1	.36
	Poodle->False Doberman->False G. Shepherd->False	.30	0	.00	0	.00

To summarise, the Bayesian account depends on the prior belief that it is unlikely that poodles alone would have an idiosyncratic property requiring great strength, so the prior belief in hypothesis 2 is quite low. Hence, finding out that poodles can bite through wire leads to a strong conclusion that other dogs can perform this feat as well. In contrast, it seems more likely that Dobermans will be able to perform some feats of strength unmatched by other dogs. It seems more plausible that biting through wire could be a Doberman-only property, thus it is harder to project this property from Dobermans to other dogs.

Evaluation. These examples illustrate the flexibility of the Bayesian account. Not only can the Bayesian account predict a variety of similarity-based effects, but it can also explain a number of results in which similarity does not seem to promote induction. Although similarity-based reasoning is still probably the norm, there are a number of exceptions beyond these two examples. Smith et al. (1993) suggested a number of properties similar to the "can bite through barbed wire." And there are a number of real-world situations like the FA Cup where than can only be one winner, such as elections and other contests, or where there is a negative contingency between events.

One strength of the Bayesian model are that it provides a way for representing the prior knowledge that similar categories do not always promote each other, and that it predicts the right results given this prior knowledge. Another strength is that this fairly simple model provides a link between a varied set of phenomena. Here, people are relatively reluctant to project a property from Dobermans to other dogs, because they have a prior belief that Dobermans are idiosyncratic in other ways. This idiosyncrasy-based account is at the heart of the Bayesian model's predictions for most of the phenomena in this chapter. However, the main limitation of the Bayesian account is, again, that it does not describe how the prior knowledge in column 3 would be assembled. Still, it does seem extremely plausible that people do have beliefs compatible with the prior knowledge assumed in Tables 11 and 12.

General Discussion

Comparison to Other Models

The Bayesian model presented here is not the only formal model of how people evaluate inductive arguments. Indeed, it is predated by at least three recent models: Osherson et al. (1990), Smith et al. (1993), and Sloman (1993). It is beyond the scope of this chapter to describe these models in detail. Briefly, however, the Osherson et al. (1990) and Sloman (1993) models are well-developed mathematical models that use information about similarity between categories to predict a wide range of phenomena with unfamiliar, blank properties, including the similarity, typicality, and diversity effects described in this chapter. Indeed, these two models make very similar predictions; one main difference between the two models is that Sloman's model is implemented as a connectionist network. In spite of these models' success at accounting for a range of results, these models do not address how the content of the property to be projected affects inductive reasoning. That is, these models do not distinguish between projectible and non-projectible properties, they would rely on a single measure of similarity for different kinds of properties, and they would tend to make similarity-based predictions in cases where similarity does not actually promote induction. Out of fairness to these models, it should be noted that these models could conceivably be extended to address how properties influence inductive reasoning, but as presented the models treat all properties the same, and do not address the phenomena in the second half of this chapter.

In contrast, the Smith et al. (1993) model does address some of the phenomena with meaningful properties. Indeed this model uses information about dependencies between properties to provide an elegant account of reasoning about properties, such as "bites through wire" that depend on the magnitudes of other dimensions, and this model can also account for basic similarity effects. However, this model does not address other central phenomena such as typicality and diversity effects, or the differences between projectible and non-projectible properties.

Although it is tempting to treat the Bayesian model as a competitor for these other three models, it is better to think about how these models support and complement each other. The Bayesian model may account for a wider range of results than any one of these models taken alone, but for any particular phenomenon that the Bayesian model and another model do make predictions, the two models would generally predict the same pattern of results. For example, both the Bayesian model and the Osherson et al. (1990) model can account for simple similarity effects, and it is unlikely that results on similarity effects could be used to distinguish one model from the other.

Instead, it is useful to note that the Bayesian model provides a computational-level description and justification for why some phenomena occur. The other three models are somewhat closer to Marr's (1982) algorithmic level, and they may provide a richer description of the psychological processes that might take place in performing inductive reasoning. In other words, the Bayesian model provides a broad framework for thinking about the more specialised models, and the Bayesian model can be used to explain why the other accounts are successful. For example, the account from Osherson et al. (1990) of similarity effects is successful because it does lead to the same results as the Bayesian model, which might be interpreted as representing the optimal or normative solution to the reasoning task.

This enterprise of using the models to complement each other rather than compete with each other is supported by the large number of phenomena that can be explained by both the Bayesian model and at least one other model. Indeed, the Bayesian model can account for several other phenomena described by Osherson et al. (1990), but space considerations have required these results to be omitted. Likewise, there seem to be deeper connections between how the Bayesian model works and how these other models operate, but these connections will be a subject for future work.

Testing the Bayesian Model

Despite this plea for cooperation rather than competition between models, it is natural to want to test the Bayesian model further. First off, it is clear that sometimes the Bayesian model makes the wrong predictions. For example, as mentioned earlier, the Bayesian model makes a strong prediction of diversity effects, but young children have not shown diversity effects (Gutheil & Gelman, 1997; Lopez et al., 1992) and the Bayesian model does not have an easy account for this non-finding. There are other cases where people seem to make fallacious inferences and the Bayesian model can predict that people will get the answer right. Consider the following arguments, from Osherson et al. (1990):

```
Flies have Property P
---------------------
Bees have Property P
```

and

```
Flies have Property P
Orangutans have Property P
--------------------------
Bees have Property P.
```

People tend to find the second argument weaker than the first, even though it seems hard to justify this point logically, given that the second argument includes the premise of the first argument as well as additional positive information about the property. In contrast, the Bayesian model can predict that the second argument is at least as strong as the first argument. Thus, the Bayesian model may be falsifiable if it does not capture the conditions under which people show this non-monotonic result.

Still, for the large number for phenomena for which the Bayesian model does predict the right results, it would seem desirable to test its account further. One challenge in testing the Bayesian model is that its predictions depend on the values used for prior beliefs, i.e., the numbers in column 3 in the tables in this chapter. (However, as noted in many cases the exact values are not too critical.) Similarly, the other models of induction (Osherson et al., 1990; Smith et al., 1993; Sloman, 1993) also depend on information about features or similarity being fed in, so the requirement of specifying people's prior knowledge is not unique to the Bayesian model.

One possible way to test the Bayesian model would be to collect information from subjects about a large number of familiar properties, for categories such as cow, ferret, horse, and so on. This information could be used to estimate the priors, for example, in Tables 1 through 6 in this chapter. If 70% of the familiar animal properties tested are believed to be true of both cows and horses, then a value of .7 would be assigned as the prior belief for hypothesis 1 in Table 1. There would be some practical issues to be overcome, such as deciding which properties to use for this enterprise. (Perhaps a set of properties could be culled from a science textbook.) Also, some method might be needed to weight properties differentially, rather than counting all properties as the same. Another, more theoretical, consideration is that prior beliefs would not always correspond to known properties. As noted in the section on typicality effects, people might sometimes have strong prior beliefs about deeper essences of categories in addition to beliefs about known properties. For example, people might have a strong belief that ferrets are an unusual and idiosyncratic mammal, without actually knowing a large number of ferret properties.

A different approach for testing the Bayesian model would be to assess derived predictions of the model, rather than applying the model directly to feature listings or feature ratings. For example, the Bayesian model explains several effects, including similarity effects, typicality effects, diversity effects, and projectibility effects, in terms of people's prior beliefs about idiosyncrasy of properties. It could be possible to collect people's direct judgements about idiosyncrasy, bypassing the need to ask people to rate a large number of known properties for a set of categories. The Bayesian model would predict that suitable judgements of idiosyncrasy would be highly correlated with inductive strength.

While it would be quite desirable in future work to collect information about people's prior beliefs in order to test the Bayesian model further, the examples in this chapter using hypothetical assumptions about prior beliefs still have considerable value. These example show that under rather plausible assumptions about prior beliefs, the Bayesian model can predict a range of phenomena in inductive reasoning. Furthermore, these predictions seem quite robust; in many cases the predictions hardly depend at all on the specific values used for most of the prior beliefs and only a few minor constraints must apply to obtain the right results.

Conclusion

The work in this chapter represents initial steps towards applying a Bayesian model to the issue of how people evaluate inductive arguments. The Bayesian analysis requires a reconception of the inductive reasoning task in terms of learning about the range of a novel property. The model makes the critical assumption that beliefs about a novel property are derived from prior knowledge about familiar properties. That is, it is assumed that "the future will resemble the past." Following on from this reconception and this assumption, the application of Bayes's Theorem is fairly straightforward. Indeed, it is somewhat surprising that a model with just one equation (Equation 1) can account for such a range of results. While these initial steps do seem promising, it is clear that further work needs to be done. These areas for future work include deriving additional information about under what conditions the Bayesian model can predict the correct results, extending the model so that it can give an explanation for why people sometimes make fallacious conclusions, investigating the links between the Bayesian model and other, more process-level models of induction, and subjecting the Bayesian model to empirical tests using data about people's prior beliefs.

References

Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. Psychological Review, 98, 409-429.

Billman, D., & Heit, E. (1988). Observational learning without feedback: A simulation of an adaptive method. Cognitive Science, 12, 587-625.

Box, G. E. P., & Tiao, G. C. (1973). Bayesian inference in statistical analysis. London: Addison-Wesley.

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: Bradford Books.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. Cognitive Psychology, 20, 65-95.

Gelman, S. A., Coley, J. D., & Gottfried, G. M., (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In L. A. Hirschfeld & S. A. Gelman (Eds.), Mapping the Mind. Cambridge, England: Cambridge University Press. 341-367.

Goodman, N. (1955). Fact, fiction, and forecast. Cambridge, MA: Harvard University Press.

Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. Journal of Experimental Child Psychology, 64, 159-174.

Heit, E. (1992). Categorization using chains of examples. Cognitive Psychology, 24, 341-380.

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. Shanks (Eds.), Knowledge, concepts, and categories London: Psychology Press.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 411-422.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. Child Development, 63, 1070-1090.

Marr, D. (1982). Vision. San Francisco: W. H. Freeman.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 179-195). Cambridge: Cambridge University Press.

Mill, J. S. (1874). A system of logic. New York: Harper.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 339-363.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopéz, A., & Shafir, E. (1990). Category-based induction. Psychological Review, 97, 185-200.

Osherson, D. N., Stern, J., Wilkie, O., Stob, M., & Smith, E. E. (1991). Default probability. Cognitive Science, 15, 251-269.

Raiffa, H., & Schlaifer, R. (1961). Applied statistical decision theory. Boston: Harvard University, Graduate School of Business Administration.

Rips, L. J. (1975). Inductive judgments about natural categories. Journal of Verbal Learning and Verbal Behavior, 14, 665-681.

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: The influence of uncertainty and feature associations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 7360753.

Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.), The Psychology of Learning and Motivation (pp. 265-301). Orlando: Academic Press.

Sloman, S. A. (1993). Feature-based induction. Cognitive Psychology, 25, 231-280.

Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. Cognition, 49, 67-96.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Acknowledgements

This chapter has benefited from comments by Nick Chater, Ulrike Hahn, Yaakov Kareev, Douglas Medin, Mike Oaksford, David Over, and Richard Shiffrin.