# Models of Inductive Reasoning

Evan Heit

To appear in Sun, R. (Ed.), Cambridge *handbook of computational psychology*. Cambridge University Press.

January 15, 2007
9954 words

1. Introduction

How do you make a prediction about the unpredictable? Inductive reasoning is about drawing conclusions that are not certain or logically valid, but still likely. Let's say you are buying a new CD for your friend. It's impossible to know with certainty what she will like, and it doesn't seem that the rules of logic will tell you which CD to buy. There's no correct answer. Nonetheless, you can make an informed guess, and indeed she will probably like the CD that you buy. The more you know about her taste in music, and which categories of music she likes and does not like, the more likely it is that your prediction will be correct. Our everyday experiences are filled with predictions of this nature—we use inductive reasoning to make likely but not certain predictions about how people will act and about things we have not seen, e.g., that when we open a door to a room, the room will have a floor and ceiling. In spite of the uncertainty, we manage to be fairly successful in our predictions—we can buy gifts that our friends will enjoy and avoid walking into rooms without floors.

When it comes to making predictions about the unpredictable, computational models are in a similar position to people. Because the judgments being modeled are themselves uncertain, it's unlikely that models of inductive reasoning will be perfectly correct. Any computational model of inductive reasoning could probably be improved by taking account of more knowledge or more principles of prediction. Nonetheless, current models of inductive reasoning already do a fairly good job of capturing patterns and regularities in how people make likely predictions.

This chapter will first review some of the empirical work on inductive reasoning by people, summarizing regularities that people show. The second part of the chapter will describe computational models of inductive reasoning, particularly from the psychology literature. The concluding section will address more general issues in modeling inductive reasoning and other cognitive activities.

## 2. Human inductive reasoning: The Data

Inductive reasoning is potentially an extremely large topic, especially since it is often defined as reasoning about problems that do not involve perfectly certain conclusions (Heit, in press). The class of problems that have perfectly certain conclusions is much more circumscribed, e.g., it could be defined in terms of a set of logical rules about what conclusions must follow from a given set of premises. In comparison, the set of problems for which inductive reasoning applies is potentially "everything else" and that is indeed a large and varied set. It's instructive to examine several textbooks on cognitive psychology. They will each no doubt have a section on inductive reasoning, but you are likely to find that different textbooks cover different topics under this heading. These topics will include analogical reasoning (Gentner, Holyoak, & Kokinov, 2001; for some connections to inductive reasoning see Lassaline, 1996; Wu & Gentner, 1998; Thagard, in press), categorization (Kruschke chapter), judgment and decision making (Busemeyer chapter), and causal reasoning (Sloman, 2005). Likewise, inductive reasoning has been an important issue in artificial intelligence and computer science (e.g., Collins & Michalski, 1989; Sun, 1995; Sun & Zhang, 2006). Some of these topics will be referred to indirectly over the course of this chapter, however this chapter will necessarily have a focus. This chapter will address one important aspect of inductive reasoning, namely psychological research on category-based induction, or how people use categories to make likely inferences.

Categories and inductive reasoning go hand in hand. For example, Anderson (1991) suggested that the most important function of categories is not that they allow us to categorize things, but rather that they allow us to draw inferences. Returning to the example of buying your friend a CD, let's say that you know that your friend likes some 1960s music and hates Celine Dion. In predicting which CDs she will like, it seems safer to choose something from within the 1960s category than from within the Celine Dion category. Category-based induction has been studied extensively by psychologists, although usually not with musical categories but with more structured categories that are part of people's basic knowledge, such as different kinds of animals.

In one of the earliest studies of category-based induction, Rips (1975) examined how people project properties of one category of animal to another. Subjects were told to imagine that on an island, all members of a species of mammals catch a particular disease, then they were asked what proportion of other species will catch the disease. For example, knowing that all horses have the disease, what proportion of cows will have the disease? What proportion of dogs? Mice?

This was a very useful task, but before reviewing results in inductive reasoning another important step in studying inductive reasoning will be described. A limitation of the Rips (1975) task was that it was best suited for drawing inferences from one category to another, whereas inductive reasoning usually involves more pieces of information. Osherson, Smith, Wilkie, Lopez, and Shafir (1990) made an influential contribution to the study of inductive reasoning by having subjects evaluate written inductive arguments in the format usually used for logic problems.

```
All horses have property X    (1)
All cows have property X
```

```
All mice have property X
All lions have property X
--------------------------
All mammals have property X
```

Here, in argument (1), the premise statements above the line are taken to be true, and the task is to judge the degree to which the conclusion statement, below the line, follows from the premises. Essentially this is a judgment of argument strength. This task is very flexible in the sense that any number of premises can be used, and various properties could be swapped in for property X (or the uninformative property X could be used). Osherson et al. reported 11 main phenomena or regularities regarding how people perform inductive reasoning. A later review paper by Heit (2000) split things up differently, considering some newer data, and reported 8 main phenomena in inductive reasoning. In the next sections of this chapter, some of these phenomena will be described, setting the stage for the presentation of computational models of inductive reasoning. Note that results from adults only will be discussed here. There is also a rich literature on inductive reasoning by children that could be used to constrain modeling; for reviews, see Heit (2000; Heit, Hahn, & Feeney, 2005; Heit & Hayes, 2005) as well as Hayes (in press).

Similarity effects. The idea that similarity should guide inductive reasoning has a distinguished history. Mill (1874) argued “that what happens once, will, under a sufficient degree of similarity of circumstances, happen again.” Going back to the example of buying a CD for your friend, if you know that she likes 1960s albums by the Rolling Stones and does not like Celine Dion, the most promising strategy is no doubt to buy her a CD by a similar 1960s band rather than by someone else who sings like Celine Dion. These similarity effects are backed up by a lot of laboratory evidence—similarity effects are the most robust result in inductive reasoning. For example, Rips (1975) found a strong correlation between strength of inferences and measures based on similarity judgments. If all horses had the disease, then most cows would have the disease, fewer dogs, and yet fewer mice. Likewise Osherson et al. (2000) found that given a choice such as between argument (2) and argument (3), about 95% of people chose argument (2), due to the greater similarity of sparrows to robins and bluejays versus geese to robins and bluejays.

```
Robins use serotonin as a neurotransmitter.  (2)
Bluejays use serotonin as a neurotransmitter.
------------------------------------------------
Sparrows use serotonin as a neurotransmitter.

Robins use serotonin as a neurotransmitter.  (3)
Bluejays use serotonin as a neurotransmitter.
------------------------------------------------
Geese use serotonin as a neurotransmitter.
```

Despite this strong evidence, there are some exceptions to similarity effects and some complications, as will soon be described. Yet clearly any computational model of induction will have to address similarity effects (and ideally the exceptions and complications too).

<u>Typicality effects</u>. Another very robust finding in inductive reasoning is the typicality effect. This phenomenon is closely tied to categorization research, in particular the idea that not all category members are equal, but instead some are more prototypical than others (e.g., Rosch & Mervis, 1975). Returning to buying a CD for your friend, if you know that she likes albums by the Rolling Stones, a prototypical 1960s guitar-based rock band, there would seem to be a lot of a lot similar 1960s bands to choose from. On the other hand, if you know that she likes albums by the Moody Blues, a much less typical 1960s band that recorded with a symphony orchestra, it would seem harder to choose another 1960s band that she would like—she might only like rock bands that use classical music.

Like similarity effects, typicality effects have been well-documented in laboratory research. For example, Rips (1975) found that when a more typical mammal such as horses had the disease, people generally drew stronger inferences to other mammals, compared to a situation where a less typical mammal, such as mice, had the disease. There was an additional effect of typicality beyond what might be predicted based only on similarity. Intuitively, if a typical mammal such as horses have the disease, then perhaps all mammals have it, that is the property applies to the superordinate category. On the other hand, if mice have the disease, it might be restricted to a subcategory of mammals such as rodents.

Indeed, Osherson et al. (1990) addressed this intuition directly. They compared arguments like (4) and (5). Here, knowing a fact about robins seems to license a stronger inference about all birds in comparison to knowing a fact about penguins. More than 90% of people chose argument (4).

```
     Robins have a higher potassium concentration in their blood
than humans.   (4)
     ------------------------------------------------
     All birds have a higher potassium concentration in their
blood than humans.
```

```
     Penguins have a higher potassium concentration in their
blood than humans.  (5)
     ------------------------------------------------
     All birds have a higher potassium concentration in their
blood than humans.
```

In sum, the typicality effect is another robust phenomenon that must be addressed by models of inductive reasoning.

<u>Diversity effects</u>. The next regularity to be discussed, the diversity effect, is somewhat more elusive than similarity or typicality, but it too has a distinguished history (Heit, Hahn, & Feeney, 2005). Bacon (1620) argued that before drawing inferences about a category, such as things possessing heat, diverse instances of this category should be examined. In making this point, he listed 28 different kinds of heat and hot things, including the rays of the sun, steam, burning hay, and the insides of animals. The diversity effect is also well illustrated in the example of buying CDs. If your friend actually likes both the Rolling Stones and Celine Dion, then you might infer that she has broad tastes in music and it would be safe to buy her one of many styles of music. On the other hand, if you know she likes the Rolling Stones and The Who, another guitar-based 1960s band, you might infer that her musical tastes are fairly narrow after all and you should not stray too far from similar bands.

Studying the diversity effect in laboratory experiments requires giving people at least two pieces of information, and varying the diversity of these two items. In the study by Rips (1975), subjects were only given one fact at a time, so diversity effects were not addressed. It was the Osherson et al. (1990) study that first focused on diversity effects in inductive reasoning by adults, using arguments such as the following.

```
     Hippos have a higher sodium concentration in their blood
than humans.   (6)
     Hamsters have a higher sodium concentration in their blood
than humans.
     ------------------------------------------------
     All mammals have a higher sodium concentration in their
blood than humans.
```

```
     Hippos have a higher sodium concentration in their blood
than humans.   (7)
     Rhinos have a higher sodium concentration in their blood
than humans.
     ------------------------------------------------
```

All mammals have a higher sodium concentration in their blood than humans.

About 75% of people chose argument (6), with the more diverse set of category members, over argument (7), with the less diverse set. In essence, diverse evidence is stronger than non-diverse evidence. Osherson et al.'s own explanation for this phenomenon was in terms of coverage: Hippos and hamsters cover, or span, the category of mammals better than do hippos and rhinos. Hence a property of hippos and hamsters is more likely to generalize to other mammals.

Although diversity effects have been found in other laboratory experiments, there are more exceptions in comparison to similarity and typicality effects (see Heit, 2000, and Heit, Hahn, & Feeney, 2005, for reviews). Still, due to the very nature of inductive reasoning, namely that it is probabilistic, there are likely to be exceptions to any regularity that is found. Hence, diversity is another effect that computational models will need to address.

<u>Other phenomena, including background knowledge effects</u>. Although similarity, typicality, and diversity are three of the most important phenomena, there are several others summarized by Osherson et al. (1990) and Heit (2000), and reported by other researchers. Some of these are fairly straightforward, encompassing other points about the structure of inductive arguments. For example, Osherson et al. reported that more evidence leads to stronger generalizations than less evidence (see also Nisbett, Krantz, Jepson, & Kunda, 1983). Looking back at argument (1), this argument with four premises seems fairly strong, and it would be stronger than another argument with just a single premise, e.g., that only horses have property X.

There is another class of phenomena that are much messier but much more interesting, both in their own right and as challenges for computational models. These phenomena involve the use of more background knowledge, and as such, are closely linked to the nature of inductive reasoning. That is, because inductive reasoning is uncertain by nature, there is always room for improvement by drawing on some other source of knowledge. (This makes a sharp contrast with deductive reasoning, or logical reasoning, where the conclusion is certain and using background knowledge is usually considered an error.) Returning to the example of buying a CD that your friend will like, the more information that you can accumulate about her musical tastes and listening habits, there more successful you will be in choosing a CD for her. In fact, with enough knowledge about her habits, you might be able to able to make detailed and sophisticated predictions such as music to play in the car versus while at home studying.

Heit and Rubinstein (1994) reported one such phenomenon based on background knowledge, showing a kind of exception to the similarity effect. Suppose there are two inductive arguments as follows.

```
Bears have property X.    (8)
---------------------------
Whales have property X.

Tuna have property X.     (9)
---------------------------
Whales have property X.
```

Which argument is stronger, (8) or (9)? Heit and Rubinstein showed that the answer depends on X. If property X is filled in with an anatomical property, such as having a liver with two chambers, then (8) is considered stronger than (9), by virtue of other shared anatomical properties of the two mammals, bears and whales. On the other hand, if property X is filled in with a behavioral property, such as traveling shorter distances in extreme heat, then (9) is considered stronger than (8). Here, shared behavioral properties are considered, and the two swimmers, tuna and whales, have more in common on this basis. Heit and Rubinstein concluded that the basic similarity effect in inductive reasoning is not a singular phenomenon but instead is mediated by background knowledge.

There have been many other demonstrations of background knowledge effects in inductive reasoning (Lopez, Atran, Coley, Medin, & Smith, 1997; Medin, Coley, Storms, & Hayes, 2003; Proffitt, Coley, & Medin, 2000; Rehder, 2006; Ross & Murphy, 1999; Shafto & Coley, 2003; Sloman, 1994). For example, Medin, Coley, Storms, and Hayes (2003) reported an exception to the diversity effect, known as the non-diversity by property reinforcement effect. The idea behind non-diversity by property reinforcement is that two diverse categories may nonetheless have some characteristic in common, and tend to generalize only to other categories with this same characteristic. This phenomenon is illustrated by the following arguments.

```
Polar bears have property X.  (10)
Antelopes have property X.
--------------------------------
All animals have property X.

Polar bears have property X.   (11)
Penguins have property X.
--------------------------------
All animals have property X.
```

When given a choice between polar bears and antelopes versus polar bears and penguins, people judged the two animals from the same biological class, polar bears and antelopes, to be more similar than the two animals from different biological classes, polar bears and penguins. However, when asked to assess the inductive strength of each argument, argument (11) was judged to be less convincing than argument (10). That is, argument (10) had less diverse evidence, yet it was the stronger argument. Intuitively, although polar bears and penguins are from different biological classes, people use their knowledge that both live in a cold climate, and infer that property X does not apply to all animals but only to animals living in cold climates. (See Heit & Feeney, 2005, for further discussion.)

The more general point is that when people evaluate inductive arguments, they potentially draw upon a variety of resources. When people use background knowledge, it is illuminating to think of what people are doing as causal reasoning, that is they are reasoning about relations between causes and effects (Rehder, 2006). People are reasoning about what causes whales to move more slowly in extreme heat, what causes two cold-weather animals to have other properties in common, and so on. One consideration to keep in mind as computational models are presented is whether they have any facility for addressing not only similarity, typicality, and diversity effects, but also background knowledge effects and indeed whether they show any capacity for causal reasoning.

## 3. Human inductive reasoning: The Computational Models

Having reviewed some of the important empirical phenomena, it is time to turn to models of inductive reasoning. A representative sample will be given rather than complete details on all models. First, two earlier models, by Osherson et al. (1990) and Sloman (1993), will be described. To anticipate, these models do an excellent job of addressing many structural phenomena in inductive reasoning, but do not address background knowledge effects sufficiently. Then, an alternative kind of model, based on Bayesian hypothesis testing, is presented in some detail (Heit, 1998). Bayesian models have the potential to address some background knowledge effects, and further applications by Tenenbaum and colleagues (Kemp & Tenenbaum, 2003; Tenenbaum & Griffiths, 2001; Tenenbaum, Kemp, & Shafto, in press) are described. Other modeling work in the psychology literature, not described here, can be found in papers by Rips (1975), Smith, Shafir, and Osherson (1993), McDonald, Samuels, and Rispoli (1996), Sloutsky and Fisher (2004), Heit and Hayes (2005), and Blok, Osherson, and Medin (in press). Likewise, some of the empirical regularities documented and modeled by psychologists were anticipated in a seminal paper in artificial intelligence, by Collins and Michalski (1989).

Osherson et al. (1990). The most influential computational model of inductive reasoning was proposed by Osherson et al. (1990). This model has two main components. The first component assesses the similarity between the premise categories and the conclusion category. In the most straightforward application of the model, this information is simply taken from people's similarity judgments for various pairs of categories. The model predicts that the basic similarity effect that is pervasive in inductive reasoning. The second component measures how well the premise categories cover the superordinate category that includes all the categories mentioned in an argument. For single-premise arguments, coverage more or less reduces to typicality, but for multiple-premise arguments, coverage gives something closer to a measure of diversity. Coverage is most easily explained with examples.

```
Mice have property X     (12)
-----------------------------
All mammals have property X

Horses have property X   (13)
-----------------------------
All mammals have property X

Hippos have property X   (14)
Rhinos have property X
-----------------------------
All mammals have property X

Hippos have property X   (15)
Hamsters have property X
```

```
----------------------------
All mammals have property X
```

For arguments (12) and (13), the lowest-level superordinate that includes all the categories is mammal. Coverage is assessed in terms of the average similarity of the premise category to members of the superordinate. To the extent that horses are more typical mammals than are mice, and therefore more similar to other kinds of mammals, argument (13) will have greater coverage than argument (12). This is how the model addresses typicality effects.

The remaining arguments have multiple premises. When assessing similarity between members of the superordinate category and the multiple premises, only the maximum similarity for any one premise category is considered. So for argument (14), very large mammals tend to be similar to both hippos and rhinos, and smalls mammals tend not to be similar to hippos and rhinos. So including rhino as a premise category does not add much information beyond just having hippo as a premise category alone. In contrast, for argument (15), some mammals are similar to hippos and other mammals are similar to hamsters. Therefore, the hamster premise adds information, and the coverage for argument (15) is greater than for argument (14). Hence, the Osherson et al. (1990) model addresses diversity effects, to the extent that greater coverage is correlated with greater diversity.

The Osherson et al. (1990) model can be written out more formally, as show in Equation I.

$$Strength = \alpha SIM(P_1,...P_n;C) + (1-\alpha)SIM\ \ (P_1,...P_n;[P_1,...P_n,C]) \quad \text{(I)}$$

Here, $\alpha$ refers to the relative influence of the similarity component (ranging from 0 to 1) and $(1-\alpha)$ is the influence of the coverage component. This equation applies when there are n premise categories P and one conclusion category C. When the premise and conclusion categories are all at the same taxonomic level, (e.g., robins, bluejays; sparrows), then SIM returns the maximum of the pairwise similarities between each $P_i$ and C. When the conclusion category is at a higher taxonomic level than the premise categories (e.g., robins, bluejays; birds), then SIM is applied recursively to known c that are members of C, and averaged over these c. For example, SIM(robins, bluejays; birds) = AVERAGE(SIM(robins, bluejays; sparrows), SIM(robins, bluejays; penguins), SIM(robins, bluejays; chickens), SIM(robins, bluejays; pigeons),….). Finally, the bracket function [] returns the lowest-level superordinate covering all of the included categories, e.g., [robins, bluejays, sparrows]=birds; [robins, bluejays, birds]=birds; [robins, bluejays, dogs]=animals.

Generally speaking, the Osherson et al. (1990) addressed a wide variety of structural phenomena in inductive reasoning, and is particularly impressive in how it puts together information from multiple premises, due to the powerful combination of similarity and coverage components. Although the model does incorporate some information about categories and similarity, the model does not address background knowledge effects such as the differential use of similarity and properties in Heit and Rubinstein (1994), exceptions to diversity in Medin et al. (2003), or more generally any use of causal knowledge or causal reasoning.

<u>Sloman (1993)</u>. The model by Sloman (1993) is particularly interesting because it asks the question of whether the coverage component in the Osherson et al. (1990) model is really necessary. Sloman's model was implemented as a connectionist network and it can account for many of the same phenomena as the Osherson et al. model. (For comparison, see an alternate connectionist model in the artificial intelligence literature, Sun, 1995.) The way Sloman's model

works is that premises of an argument are encoded by training the connectionist network to learn associations between input nodes representing the features of the premise categories and an output node for the property to be considered, using the classic delta rule (e.g., Sutton & Barto, 1981). For example, for the model to learn that applies have property X, it would learn to associate a vector of features such as {is round, is red, is edible, ...} with an output node representing property X. Then the model is tested by presenting the features of the conclusion category and measuring the activation of the same output node. For example, to evaluate the strength of the conclusion that oranges have property X, the model would use a somewhat different input vector of features {is round, is orange, is edible, …) and measure the degree of activation for the output unit corresponding to property X.

The model accounts for similarity effects because training and testing on similar input vectors will lead to strong outputs during testing. Going back to arguments (2) and (3), the model would first be trained to associate input representations for robins and bluejays with an output node representing the property in the conclusion. Because the representation for sparrows would have a lot of overlap with representations of robins and bluejays, presenting sparrow to the network would also activate the output node strongly. In comparison, the representation of geese would have much less overlap with representations for robins and bluejays. Hence, presenting geese to the network would only weakly activate the output node.

The activation function is as follows, in Equation II.

$$a(C \mid P_i,...,P_n) = \frac{W(P_i,...,P_n) \bullet C}{|C|^2} \qquad \text{(II)}$$

This function refers to the output activation given a set of n premise categories P and a conclusion category C. W is a vector corresponding to the already-trained weights in the network after the premise categories have been learned. C is a vector corresponding to the featural representation of the conclusion category. The dot product between W and C is computed. yielding a value corresponding to the similarity between the premise categories and the conclusion category. For example, donkey and mule would have many features in common and there would be a fairly high, positive dot product between the two vectors. On the other hand, donkey and ostrich would have fewer features in common and a lower dot product, perhaps close to zero. Finally, the activation is scaled. in the denominator, by the squared length of the vector C, essentially a measure of the number of known features of C. If C corresponds to a well-known category such as dogs, it will be relatively difficult to draw a new conclusion. If C corresponds to a poorly-known category such as ocelots, it will be easier to draw new conclusions about the category.

The model accounts for diversity effects because training on a diverse set of categories will tend to strengthen a greater number of connections in W than training on a narrow range of categories. In terms of arguments (6) and (7), training the network that both hippos and hamsters have a certain property would activate a broad range of features that apply to various mammals, leading to a strong conclusion that all mammals have that property. That is, hippos and hamsters would activate different features and different connections. In comparison, training the network that hippos and rhinos have a property would only activate a narrow range of features and connections. Although this model does have a notion of breadth of features, there is no distinct component for assessing coverage of a superordinate category, as in the Osherson et al. (1990) model, and indeed Sloman’s model does not even rely on knowledge about superordinate categories (see also Sloman, 1998, and see Sun & Zhang, 2006, for an alternative account). Nonetheless, the Sloman model can account for not only diversity effects but a variety of other phenomena involving multiple premises.

The treatment of typicality effects is slightly less straightforward. The model would correctly predict that argument (4) is stronger than argument (5), namely that robins lead to stronger inferences about all birds than do penguins. The Sloman model makes this prediction in terms of overlap in representations. On the assumption that the featural representations for robins and birds are closer than the representations for penguins and birds, the model predicts greater activation for birds after training on robins compared to training on penguins. Yet the model is essentially predicting the typicality effect in the same way as the similarity effect. This seems to be at odds with a finding from Rips (1975) of an independent contribution of typicality beyond similarity. Still, at a broad level, the Sloman predict predicts a kind of typicality effect (see Heit, 2000, for further discussion).

More importantly, the Sloman (1993) model, like the Osherson et al. (1990) model, can account for many structural phenomena in inductive reasoning, but it does not address background knowledge effects, and does not use knowledge about the properties being reasoned about to guide the use of similarity or information about causality.

<u>Bayesian model</u>. The next model to be discussed is the Bayesian model, applied to inductive reasoning problems by Heit (1998; see also Tenenbaum & Griffiths, 2001, as well as the chapter by Griffiths, Kemp, & Tenenbaum, this volume). According to the Bayesian model, evaluating an inductive argument is conceived of as learning about a property, in particular learning for which categories the property is true or false. For example, in argument (16),

```
Cows have property X     (16)
-----------------------------
Sheep have property X.
```

the goal is to learn which animals have property X and which do not. The model assumes that for a novel property X, people would rely on prior knowledge about familiar properties, to derive a set of hypotheses about what property X may be like. For example, people know some facts that are true of all mammals, including cows and sheep, but they also know some facts that are true just of cows and some facts that are true just of sheep. The question is which kind of properties is property X. Is it a cow-and-sheep property, or a cow-only property, or a sheep-only property? To answer this question, the Bayesian model treats the premise or premises in an inductive argument as evidence, which is used to revise beliefs about the prior hypotheses according to Bayes's Theorem. Once these beliefs have been revised, then the plausibility of the conclusion is estimated.

People know quite a few properties of animals, but these known properties must fall into four types: properties that are true of cows and sheep, properties that are true of cows but not sheep, properties that are true of sheep but not cows, and properties that are not true of either cows or sheep. These 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, as listed in Table 1. As shown in the table, 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 sheep. This high value could reflect the high degree of similarity between cows and sheep, and that people know many other properties that are true of both cows and sheep. (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 sheep, a 5% chance that a new property would be true of sheep and not cows, and a 20% chance that the property is true of neither category.

The next step is to combine these prior beliefs with new evidence, using Bayes's Theorem as shown in Equation III. The given premise, "Cows have property X," is used to update beliefs about the four hypotheses, so that the conclusion, "Sheep have property X," can be evaluated. In applying Bayes's Theorem, below, the premise is treated as the data, D. The prior degree of belief in each hypothesis is indicated by P(Hi). (Note that there are four hypotheses, so n = 4 here.) The task is to estimate P(Hi|D), that is the posterior degree of belief in each hypothesis given the data.

$$P(H_i \mid D) = \frac{P(H_i)P(D \mid H_i)}{\sum_{j=1}^{n} P(H_j)P(D \mid H_j)} \qquad \text{(III)}$$

The calculations are shown for all four hypotheses, given the data that cows have property X. The calculation of P(D|Hi) is 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, is calculated using Bayes's Theorem. Notably, hypothesis 1, that cows and sheep have the property, and hypothesis 2, that just cows have the property, have been strengthened, and the two remaining hypotheses have been eliminated from consideration.

Finally, the values in Table 1 may be used to evaluate the conclusion, that sheep have property X. 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 sheep have the property was only .75. hence, the premise that cows have the property led to an increase in the belief that sheep have the property.

The Bayesian model addresses many of the key phenomena in inductive reasoning. For example, the model predicts the similarity effect because novel properties would be assumed to follow the same distributions as familiar properties. Generalizing from cows to sheep seems plausible because many known properties are true of both categories. In contrast, generalizing from cows to mice seems weaker because prior knowledge indicates that there are fewer properties in common for these two categories. The Bayesian model also addresses typicality effects, under the assumption that according to prior beliefs, atypical categories, such as mice, would have a number of idiosyncratic features. A premise asserting a novel property about mice would suggest that this property is likewise idiosyncratic and not to be widely generalized. In comparison, prior beliefs about typical categories would indicate that they have many features in common with other categories, hence a novel property of a typical category should generalize well to other categories.

The Bayesian model also addresses diversity effects, with a rationale similar to that for typicality effects. An argument with two similar premise categories, such as hippos and rhinos, could bring to mind a lot of idiosyncratic properties that are true just of large mammals. Therefore a novel property of hippos and rhinos might seem idiosyncratic as well. In contrast, an argument with two diverse premise categories, such as hippos and hamsters, could not bring to mind familiar idiosyncratic properties that are true of just these two animals. Instead, the prior hypotheses would be derived from known properties that are true of all mammals or all animals. Hence a novel property of hippos and hamsters should generalize fairly broadly. More generally, Heit (1998) showed that the Bayesian model addresses about the same range of structural phenomena in inductive reasoning as does the Osherson et al. (1990) and Sloman (1993) models.

Although it is by no means a complete model of the use of background knowledge in inductive reasoning, the Bayesian model does make a start at addressing background knowledge effects. For example, when reasoning about the anatomical and behavioral properties in Heit and Rubinstein (1994), subjects could have drawn on different priors for the two kinds of properties. Reasoning about anatomical properties led people to rely on prior knowledge about familiar anatomical properties, so there was stronger generalization from bears to whales than from tunas to whales. In contrast, when reasoning about a behavioral property, the prior hypotheses could be drawn from knowledge about familiar behavioral properties. These priors would tend to promote inferences between animals such as tunas and whales that are similar behaviorally rather than anatomically. Although by no means does the Bayesian model itself perform causal reasoning, the priors used in the model could be the end-product of causal reasoning.

Indeed, a fair criticism of the Bayesian model would be that its predictions are captive to assumptions about distributions of prior beliefs. Heit (1998) responded to this criticism in part, by stating that the exact values of priors don't usually matter, e.g., the values for prior beliefs in Table 1 could be somewhat different and the same general pattern would emerge. Also, Heit argued that priors would be derived from psychologically plausible mechanisms. It could be assumed that priors are determined by the number of known properties of each type that are brought to mind in the context of evaluating an inductive argument. In this way, prior beliefs for new properties would be estimated using something like an availability heuristic (Tversky & Kahneman, 1973) based on known properties.

Still, it would be a major improvement to the Bayesian model if assumptions about prior beliefs could be generated by a model rather than just simply assumed. Tenenbaum and colleagues (Griffiths et al., this volume; Kemp & Tenenbaum, 2003; Tenenbaum et al., in press) have made improvements in this regard. Their central idea is that properties of living things have come about due to a process, rather than being arbitrarily distributed. That is, due to an evolutionary process, living things can be thought of as being on the branches of a tree. Two very similar animals probably have a relatively recent common ancestor, whereas two animals that are very different probably only have a common ancestor from very long ago. The starting point for Tenenbaum and colleagues was a large set of ratings on whether different animals possess various known properties. From principles of branching evolution as well as mutation, they derived a tree structure, in effect inferring common ancestors for each pair of animals. Tenenbaum and colleagues then used the tree structure to set the priors for the Bayesian model. It was found that priors derived from the tree structure were much more successful at predicting people's judgments than priors derived from their property ratings. Interestingly, Tenenbaum and colleagues have argued that this method of setting priors is particularly successful because it represents people's causal knowledge of how properties of living things come about and the mechanisms by which they could be shared.

To sum up the description of Bayesian modeling of inductive reasoning, this kind of modeling does address many phenomena and regularities. The greatest weakness of Bayesian models, that they are subject to assumptions about priors, may also be their greatest strength, in the sense that many predictions are robust over different assumptions about priors, and furthermore the priors themselves could be derived from other reasoning processes that likewise could be modeled.

4. Causal learning and causal reasoning

An interim summary of the chapter so far could be that current models of inductive reasoning can account for much of what people do, but they especially fall short when it comes to causal knowledge. Part of the problem, from the modeling perspective, is that research on modeling of inductive reasoning has largely proceeded separately from research on modeling the use of causal knowledge. Research on modeling the use of causal knowledge has itself largely taken two separate approaches. One approach focuses on causal learning, or causal induction, namely how people infer a causal structure from a set of observations. For example, a novice DJ at a party could observe that playing some songs makes people get up and dance, whereas playing other songs makes people sit quietly, and could draw inferences about what kinds of songs are associated with different behaviors. The other approach focuses on causal reasoning, which mainly addresses how often complex knowledge structures can be used to make fresh inferences. For example, an expert with years of experience in the music industry could makes detailed predictions about which new performers will succeed and which will not, and give an elaborate causal explanation in terms of market forces, demographics, current trends, etc. Of course there is much overlap between causal learning and causal reasoning, but the emphasis in causal learning research is more on the acquisition of causal knowledge and in causal reasoning research on the use of causal knowledge.

The question of how to infer causation from a set of observations has been considered a central problem of induction at least since the time of Hume (1777). If event B tends to follow event A, does A cause B? At the simplest level, this is the most important question faced by animals seeking food or other necessities—which cues are predictive of obtaining nourishment or some other needed reward? Indeed, one of the most important models of causal learning has its origins in animal conditioning research (Rescorla & Wagner, 1972). The Rescorla-Wagner model can be written as shown in Equation IV.

$$V_A(n+1) = V_A(n) + \beta(\lambda(n) - V_A(n)) \qquad \text{(IV)}$$

This formula gives the associative strength, V, of stimulus A, on trial n+1, as a function of the associative strength of A on trial n, a learning rate $\beta$, and a level of reinforcement, $\lambda(n)$, on trial n. In the asymptote, the associative strength of A will tend towards the expected value of the level of reinforcement. The consequence is that stimuli that are followed by reinforcement tend to gain higher levels of associative strength than stimuli that are not reinforced. Of course, the Rescorla-Wagner model can also be extended to situations where there is more than one stimulus on each learning trial—in these situations the stimuli compete with each other to be the best predictor of reinforcement. For example, once the animal full learns that A is associated with reinforcement, if the compound stimulus A+X is reinforced, the animal will be blocked from learning an association between X and reinforcement, because A alone sufficiently predicts the outcome.

There is a lively debate about whether such associative mechanisms are good accounts of people's causal learning (e.g., Buehner & Cheng, 2005; Luhmann & Ahn, 2004; Novick & Cheng, 2004; Shanks, in press; Waldmann, 2000; White; 2005), to which the reader is referred. However, there is no doubt that people do acquire causal beliefs. Hence, the next question is how do people represent and reason using this causal information.

The current state of the art in cognitive science is that people represent causal knowledge in the form of a causal network, akin to a formalization known as Bayes nets (Gopnik, Glymour, Sober, Schulz, Kushnir, & Danks, 2004; Pearl, 2000; Sloman, 2005; Spirtes, Glymour, & Scheines, 2001). These networks allow for the representation of complex causal configurations, such as a causal chain (e.g., the boy threw the ball that broke the window that woke the cat), a common effect from multiple causes (e.g., a fire is caused by a spark in the presence of oxygen and some kind of fuel), and common effects from a single cause (e.g., a cold caused a runny nose, fever, and redness of the eyes). Bayes nets provide a formalism for using this structured knowledge to draw inferences, e.g., to estimate the probability that the cat will wake. Although this kind of modeling no doubt provides a valuable framework way for describing causal reasoning, further work needs to be done to integrate this framework with human data. Although there have been some efforts to relate this kind of formalism to results from inductive reasoning (Rehder, 2006, in press; Rehder & Burnett, 2005), for the most part, modeling of causal reasoning stands apart from modeling of inductive reasoning as captured by the results in this chapter. It remains a challenge for Bayes nets or some related model of causal reasoning to show that they can not only explain what the current models of inductive reasoning can explain but also do a better job by incorporating the use of causal knowledge.

## 5. Conclusion

Summary. Although inductive reasoning is by its very nature uncertain reasoning, there are nonetheless regularities and patterns that people show when performing inductive reasoning. These regularities include similarity effects, typicality effects, diversity effects, and several others. Current models of inductive reasoning, such as the influential Osherson et al. (1990)

model, are successful at explaining these regularities, but have problems addressing the exceptions to these regularities, particularly when people use other kinds of background knowledge outside the scope of these models, such as causal knowledge. There are also extant models of causal induction and causal reasoning, but these models have generally been used to address other sets of results. It remains an important task for future research to integrate computational modeling work on inductive reasoning, causal induction, and causal reasoning, as applied to a diverse set of human results.

This chapter will conclude by discussing two general issues that arise in modeling inductive reasoning, but also arise in computational modeling of other cognitive activities. The first issue is that cognitive activities do not neatly fall into pigeonholes. The second is that putting background knowledge into models is the necessary next step.

Everything is intertwingled. As Nelson (1987) noted, everything is deeply intertwingled. Although it is convenient for books to have separate chapters about different cognitive activities, and likewise it is convenient for researchers to have separate models for different cognitive activities, the truth is that cognitive activities are not as separate as they are portrayed. For example, there is a traditional split between inductive reasoning and deductive reasoning, yet it is still not known where the dividing line is or even if inductive reasoning and deductive reasoning involve distinct cognitive processes (Heit, in press; Rotello & Heit, 2005). The well-known mental models theory of reasoning (Johnson-Laird & Yhang, this volume), usually applied to deductive reasoning problems, has also been applied to inductive reasoning problems (Johnson-Laird, 1994; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999;). An alternative to mental model theory is the probabilistic account, which aims to account for a variety of reasoning phenomena, particularly traditional deduction problems, in terms of probabilistic formulas (Chater & Oaksford, 2000). The probabilistic account is saying that people solve deduction problems by means of induction processes. Likewise, Osherson et al. (1990) applied their model of inductive reasoning to some problems involving deductively valid conclusion, and some modeling work in the artificial intelligence literature has addressed both deductive reasoning and inductive reasoning to some extent (Collins & Michalski, 1989; Sun; 1995). Although there are likely some important differences in the reasoning processes involved in induction versus deduction (Rips, 2001), it would be ideal if models of reasoning addressed both rather than focusing on one or the other. Likewise, causal reasoning is another kind of probabilistic reasoning, and as noted models of inductive reasoning would no doubt be improved by capturing causal reasoning, particularly when addressing effects of background knowledge. The same point can be made about models of judgment and decision making (e.g., Busemeyer chapter)—these models are essentially addressing inductive phenomena.

Inductive reasoning is related to not only other kinds of reasoning, but also to other cognitive activities such as categorization and recognition memory. For example, in a clever set of experiments, Sloutsky and Fisher (2004) examined relations between inductive reasoning, categorization, and recognition memory. For sets of various animals, subjects either made inductive reasoning judgments, inferring hidden properties, or made categorization judgments, assigning the animals to categories. There was a strong correlation between inductive reasoning and categorization (see also Rehder & Hastie, 2001). Moreover, Sloutsky and Fisher found systematic relations between inductive reasoning and recognition memory. For example, after seeing some pictures of cats with a particular property and inferring that all cats have this property, people tended to falsely recognize other pictures of cats, that is, remember them as having been seen when they had not been seen. In assessing this work, Heit and Hayes (2005) argued that the boundaries between induction, categorization, and recognition are fuzzy. Inductive reasoning can be thought of as a kind of categorization, categorization itself can be said to involve reasoning, and likewise a recognition judgment can be thought of as a kind of categorization. Heit and Hayes concluded that induction, categorization, and recognition should not be modeled separately, but instead it would be desirable for models to capture the theoretical relations among these activities and likewise the empirical relations, such as how one kind of judgment is correlated with another or even how one kind of judgment directly affects another.

Knowledge is power. The second, and final, general issue to be raised about modeling is that incorporating background knowledge into models is usually the most important next step and the most difficult. Taking another point from Sir Francis Bacon, knowledge is power. Again, comparing inductive reasoning to categorization is useful, in terms of historical progression. Following years of research in which computational models of categorization had been compared to each other, e.g., exemplar models versus prototype models, Murphy and Medin (1985) concluded that all the current models were wrong! Exemplar models, prototype models, and alternatives would not be able to, say, categorize someone who jumps into a swimming pool with all his clothes on as a drunk, because categorization models did not take account of people's background knowledge, intuitive theories, and use of causal reasoning and explanations. Since that time, most research on computational modeling of categorization has still focused on structural issues that do not depend heavily on background knowledge and for which models might make different predictions only in the fine details. Still there have been some efforts to incorporate background knowledge into categorization models (Heit, 1997; Heit & Bott, 2000, Heit, Briggs, & Bott, 2004; Rehder & Murphy, 2003).

Computational modeling of inductive reasoning is now at a similar point in its own history. The models reviewed in this chapter, by Osherson et al. (1990), Sloman (1993), and Heit (1998) are successful at addressing largely the same set of structural phenomena in inductive reasoning, without addressing important background knowledge effects. (See also Heit and Hayes, 2005, for further examples of current models of induction making very similar predictions.) Although it is no doubt possible to conduct experiments splitting apart the detailed predictions of these models, what is needed most is new efforts on modeling, incorporating background knowledge and causal reasoning.

Returning to the themes at the start of this chapter, models of inductive reasoning have a particularly hard job, because they are addressing a form of reasoning that is itself uncertain and does not have a correct answer. In such a situation, the success of the models can only be improved by taking account of additional knowledge. Doing so is especially important because the structural phenomena such as similarity, typicality, and diversity effects have exceptions and can indeed be overridden by other knowledge. In addressing these knowledge effects, models of inductive reasoning will need to become closer to models of causal reasoning. Going back to the example of buying a CD for your friend, it may be useful to buy CDs that are similar to what she already has. But the most successful strategy for predicting the unpredictable would no doubt to be to discover why she likes some CDs and does not like others.

# References

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

Bacon, F. (1620). *Novum organum*. London: George Bell and Sons.

Buehner, M. & Cheng, P.W. (2005). Causal learning. In K.J. Holyoak & R.G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning*, 143-168. New York, N.Y.: Cambridge University Press.

Blok, S. V., Osherson, D. N., & Medin, D. L. (in press). From similarity to chance. In A. Feeney & E. Heit (Eds.), *Inductive reasoning*. Cambridge University Press.

Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behavior. *Synthese, 122*, 93-131.

Collins, A., & Michalski, R. S. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science, 13*, 1-49.

Gentner, D., Holyoak, K., Kokinov, B., eds. (2001). The analogical mind: Perspectives from cognitivesScience. Cambridge, MA: MIT Press.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 1-31.

Hayes, B. K. (in press). The development of inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning*. Cambridge University Press.

Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories*, 7-41. London: Psychology Press.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition*, 248-274. Oxford: Oxford University Press.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7,* 569-592.

Heit, E. (in press). What is induction and why study it? In A. Feeney & E. Heit (Eds.), *Inductive reasoning*. Cambridge University Press.

Heit, E., & Bott, L. (2000). Knowledge selection in category learning. In D. L. Medin (Ed.), *Psychology of Learning and Motivation* (Vol. 39), 163-199. San Diego: Academic Press.

Heit, E., Briggs, J., & Bott, L. (2004). Modeling the effects of prior knowledge on learning incongruent features of category members. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1065-1081.

Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin & Review, 12*, 340-344.

Heit, E., Hahn, U., & Feeney, A. (2005). Defending diversity. In W. Ahn, R. Goldstone, B. Love, A. Markman, & P. Wolff (Eds.), *Categorization inside and outside of the laboratory: Essays in honor of Douglas L. Medin*, 87-99. Washington DC: APA.

Heit, E. & Hayes, B. K. (2005). Relations among categorization, induction, recognition, and similarity. *Journal of Experimental Psychology: General*, *134*, 596-605.

Heit, E., & Rotello, C. M. (2005). Are there two kinds of reasoning? In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

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

Hume, D. (1777). *An enquiry concerning human understanding*. Oxford: Clarendon Press.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition, 50*, 189-209.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. A., & Caverni, J. P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62-88.

Kemp, C. S. and Tenenbaum, J. B. (2003). Theory-based induction. In Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society.

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 754-770.

Lopez, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology, 32*, 251-295.

Luhmann, C. C. & Ahn, W. (2005). The meaning and computation of causal power: A critique of Cheng (1997) and Novick and Cheng (2004). *Psychological Review, 112*, 685-693.

McDonald, J., Samuels, M., & Rispoli, J. (1996). A hypothesis-assessment model of categorical argument strength. *Cognition, 59*, 199-217.

Medin, D. L., Coley, J. D., Storms, G., Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10*, 517-532.

Mill, J. S. (1874). *A System of Logic, Ratiocinative and Inductive.* New York: Harper & Row.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nelson, T. H. (1987). *Computer lib / Dream machines*. Redmond, WA: Microsoft.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339-363.

Novick, L.R., & Cheng, P.W. (2004). Assessing interactive causal influence. Psychological Review, 111, 455-485.

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

Pearl, J. (2000). *Causality*. New York: Oxford University Press.

Proffitt, J. B., Coley, J. D., Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 811-828.

Rehder, B. (2006). When causality and similarity compete in category-based property induction. *Memory & Cognition, 34*, 3-16.

Rehder, B. (in press). Property generalization as causal reasoning In A. Feeney & E. Heit (Eds.), *Inductive reasoning*. Cambridge University Press.

Rehder, B. & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology, 50*, 264-314.

Rehder, B. & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology*: General, 130, 323-360.

Rehder, B., & Murphy, G. L. (2003). A knowledge-resonance (KRES) model of knowledge-based category learning. *Psychonomic Bulletin & Review, 10*, 759-784.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy

(Eds.), *Classical conditioning II: Current research and theory*, 64-99. New York: Appleton-Century-Crofts.

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

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science, 12*, 129-134.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology, 38,* 495-553.

Shafto, P. & Coley, J.D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naıve similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 641-649.

Shanks, D. R. (in press). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*.

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

Sloman, S. A. (1994). When explanations compete: The role of explanatory coherence on judgments of likelihood. *Cognition, 52*, 1-21.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35*, 1-33.

Sloman, S. A. (2005). *Causal models: How we think about the world and its alternatives*. New York: Oxford University Press.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166-188.

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

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd edition). Cambridge, MA: MIT Press.

Sun, R. (1995). Robust reasoning: integrating rule-based and similarity-based reasoning. *Artificial Intelligence, 75*, 241-296.

Sun, R., & Zhang, X. (2006). Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental & Theoretical Artificial Intelligence*, *18*, 169-191.

Sutton, R., & Barto, A. (1981). Towards a modern theory of adaptive networks: expectation and prediction. *Psychological Review, 88*, 135-170.

Tenenbaum, J.B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*,629-641.

Tenenbaum, J.B., Kemp, C. & Shafto, P. (in press). Theory-based Bayesian models of inductive reasoning. In Feeney, A. & Heit, E. (Eds.), *Inductive reasoning*. Cambridge University Press.

Thagard, P. (in press). Abductive inference: From philosophical analysis to neural mechanisms. In A. Feeney & E. Heit (Eds.), *Inductive reasoning*. Cambridge University Press.

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

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 53-76.

White, P. A. (2005). “The power PC theory and causal powers: reply to Cheng (1997) and Novick and Cheng (2004)”. *Psychological Review*, *112*, 675-682.

Wu, M., & Gentner, D. (1998). Structure in category-based induction. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1154-1158). New Jersey: Erlbaum.

Table 1
Sample Application of the Heit (1998) Model

| Hypothesis Number | Cows? | Sheep? | Degree of Prior Belief $P(H_i)$ | $P(D \mid H_i)$ | Posterior Belief $P(H_i \mid D)$ |
|---|---|---|---|---|---|
| 1 | True | True | .70 | 1 | .93 |
| 2 | True | False | .05 | 1 | .07 |
| 3 | False | True | .05 | 0 | .00 |
| 4 | False | False | .20 | 0 | .00 |