Handbook of Thinking and Reasoning

The Problem of Induction

Steven A. Sloman

Brown University

David A. Lagnado

University College, London

Please address correspondence to:

Steven Sloman
Cognitive and Linguistic Sciences
Brown University, Box 1978
Providence, RI 02912
Email: Steven_Sloman@brown.edu.
Phone: 401-863-7595
Fax: 401-863-2255

In its classic formulation, due to Hume (1739, 1748), inductive reasoning is an activity of the mind that takes us from the observed to the unobserved. From the fact that the sun has risen every day thus far, we conclude that it will rise again tomorrow; from the fact that bread has nourished us in the past, we conclude that it will nourish us in the future. The essence of inductive reasoning lies in its ability to take us beyond the confines of our current evidence or knowledge to novel conclusions about the unknown. These conclusions may be particular, as when we infer that the next swan we see will be white, or general, as when we infer that all swans are white. They may concern the future, as in the prediction of rain from a dark cloud; or concern something in the past, as in the diagnosis of an infection from current symptoms.

Hume argued that all such reasoning is founded on the relation of cause and effect. It is this relation which takes us beyond our current evidence, whether it is an inference from cause to effect, or effect to cause, or from one collateral effect to another. Having identified the causal basis of all our inductive reasoning, Hume proceeded to raise a fundamental question now known as ‘the problem of induction’: what are the grounds for such inductive or causal inferences? In attempting to answer this question, Hume presents both a negative and a positive argument.

In his negative thesis, Hume argued that our knowledge of causal relations is not attainable through *demonstrative* reasoning, but is acquired through past experience. To illustrate, our belief that fire causes heat, and the expectation that it will do so in the future, is based on previous cases in which one has followed the other, and not on any *a priori* reasoning. However, once Hume identifies experience as the basis for inductive inference, he proceeds to demonstrate its inadequacy as a justification for these inferences. Put simply, any such argument requires the presupposition that past experience will be a good guide to the future, and this is the very claim we seek to justify.

For Hume, what is critical about our experience is the perceived similarity between particular causes and their effects: ‘From causes, which appear *similar*, we expect similar effects. This is the sum of all our experimental conclusions’ (see Goldstone & Son, Chapter 1, this volume). But this expectation cannot be grounded in reason alone, because similar causes could conceivably be followed by dissimilar effects. Moreover, if one introduces hidden powers or mechanisms to explain our observations at a deeper level, the problem just gets shifted down. What guarantees that the powers or mechanisms that underlie our current experiences will do so in the future?

In short, Hume’s negative argument undermines the assumption that the future will resemble the past. This assumption cannot be demonstrated a priori, as it is not contradictory to imagine that the course of nature may change. But neither can it be supported by an appeal to past experience, as this would be to argue in a circle.

Hume’s argument operates at two levels, both *descriptive* and *justificatory*. At the descriptive level it suggests that there is no actual process of reflective thought that takes us from the observed to the unobserved. After all, as Hume points out, even young infants and animals make such inductions, though they clearly do not use reflective reasoning. At the justificatory level, it suggests that there is no possible line of reasoning that could do so. Thus Hume argues both that reflective reasoning *does not* and *could not* determine our inductive inferences.

Hume’s positive argument provides an answer to the descriptive question of how we *actually* pass from the unobserved to the observed but not to the justificatory one. He argues that it is custom or habit that leads us to make inferences in accordance with past regularities. Thus, after observing many cases of a flame being accompanied by heat, a novel instance of a flame creates the idea, and hence an expectation, of heat. In this way a correspondence is set up between the regularities in the world and the expectations of the mind. Moreover, Hume maintains that this tendency is 'implanted in us as an instinct', because nature would not entrust it to the vagaries of reason. In modern terms, then, we are pre-wired to expect past associations to hold in the future, although what is associated with what will depend on the environment we experience. This idea of a general purpose associative learning system has inspired many contemporary accounts of inductive learning (see Cheng & Beuhner, Chapter 5, this volume).

Hume's descriptive account suffers from several shortcomings. For one, it seems to assume that there is an objective sense of similarity or resemblance that allows us to pass from like causes to like effects, and vice-versa. In fact, a selection from amongst many dimensions of similarity might be necessary for a particular case. For example, to what degree, and in what respects, does a newly encountered object, e.g., a new type of candy bar, need to be similar to previously encountered bars, for someone to expect a similar taste? If we are to acquire any predictive habits at all we must be able to generalize to some extent from one object to another, or to the same object at different times and contexts. How this is carried out is as much in need of a descriptive account as the problem of induction itself. Second, we might accept that no reflective reasoning can *justify* our inductive inferences, but this does not entail that reflective reasoning cannot be the *actual* cause of some of our inferences. Nevertheless, Hume presciently identified the critical role of both similarity and causality in inductive reasoning, the variables that, as we will see, are at the heart of work on the psychology of induction.

Hume was concerned with questions of both description and justification. In contrast, the logical empiricists (e.g., Carnap, 1950, 1966; Hempel, 1965; Reichenbach, 1938) focused only on justification. Having successfully provided a formal account of deductive logic (Frege, 1880; Russell & Whitehead, 1925), in which questions of deductive validity were separated from how people actually make deductive inferences (see Evans, Chapter 6, this volume), philosophers attempted to do the same for inductive inference by formulating rules for an inductive logic.

Central to this approach is the belief that inductive logic, like deductive logic, concerns the logical relations that hold between statements, irrespective of their truth or falsity. In the case of inductive logic, however, these relations admit of varying strengths, a conditional probability measure reflecting the *rational* degree of belief that someone should have in a hypothesis given the available evidence. For example, the hypothesis that ‘all swans are white’ is made probable (to degree p) by the evidence statement that ‘all swans in Central park are white’. Upon this basis the logical empiricists hoped to codify and ultimately justify the principles of sound inductive reasoning.

This project proved to be fraught with difficulties, even for the most basic inductive rules. Thus consider the rule of induction by enumeration, which states that a universal hypothesis H_1 is confirmed or made probable by its positive instances E. The problem is that these very same instances will also confirm a different universal hypothesis H_2 (indeed an infinity of them) which makes an entirely opposite prediction about subsequent cases. The most notorious illustration of this point was provided by Goodman

(1955), and termed ‘the new riddle of induction’. Imagine that you have examined numerous emeralds, and found them all to be colored green. You take this body of evidence *E* to confirm (to some degree) the hypothesis that ‘All emeralds are green’. However, suppose we introduce the predicate ‘grue’, which applies to all objects examined so far (before time *t*) and found to be green, *and* to all objects not examined and blue. Given this definition, and the rule that a universal hypothesis is confirmed by its positive instances, our evidence set *E* also confirms the gruesome hypothesis ‘All emeralds are grue’. But this is highly undesirable, because each hypothesis makes an entirely different prediction as to what will happen in the future (after time *t*), when we examine a new emerald. Goodman states this problem as one of *projectibility*: how can we justify or explain our preference to project predicates such as ‘green’ from past to future instances, rather than predicates such as ‘grue’?

Many commentators object that the problem hinges on the introduction of a bizarre predicate, but the same point can be made equally well using mundane predicates or simply in terms of functions (see Hempel, 1965). Indeed the problem of drawing a line or curve through a finite set of data points illustrates the same difficulty. Two curves C_1 and C_2 may fit the given data points equally well, but diverge otherwise. According to the simple inductive rule both are equally confirmed and yet we will often prefer one curve over the other. Unfortunately, an inductive logic of the kind proposed by Carnap et al. gives us no grounds to decide which predicate (or curve) to project.

In general, then, Goodman’s problem of projectibility concerns how we distinguish projectible predicates such as ‘green’ from non-projectible ones such as ‘grue’. Although he concurs with Hume’s claim that induction consists in a mental habit formed by past regularities, he argues that Hume overlooks the further problem (the new riddle) of *which* past regularities are selected by this mental habit, and thus projected in the future. After all, it would appear that we experience a vast range of regularities and yet are prepared to project only a small subset. Goodman himself offers a solution in terms of *entrenchment*. In short, a predicate is entrenched if it has a past history of use, where both the term itself, and the extension of the term, figure in this usage. Thus 'green' is entrenched whereas 'grue' is not, because our previous history of projections involves numerous cases of the former, but none of the latter. In common with Hume, then, Goodman gives a descriptive account of inductive inference, but one grounded in the historical practices of people, and in particular their language use, rather than simply the psychology of an individual.

One shortcoming of Goodman's proposal is that it hinges on language use. Ultimately he attempts to explain our inductive practices in terms of our linguistic practices: 'the roots of inductive validity are to be found in our use of language'. But surely inductive questions, such as the problem of projectibility, arise and are solved by infants and animals without language (see Suppes, 1994). Indeed our inductive practices may drive our linguistic practices, rather than the other way around. Moreover, Goodman rules out, or at least overlooks, the possibility that the notions of similarity and causality are integral to the process of inductive reasoning. But, as we shall see, more recent analyses suggest that these are the concepts that will give us the most leverage on the problem of induction.

In his essay 'Natural Kinds' (1970), Quine defends a simple and intuitive answer to Goodman's problem: projectible predicates apply to members of a *kind*, a grouping formed on the basis of similarity. Thus ‘green’ is projectible while ‘grue’ is not because green things are more similar than grue things; that is, green emeralds form a *kind* whereas grue emeralds do not. This shifts the explanatory load onto the twin notions of similarity and kind, which Quine holds to be fundamental to inductive inference: ‘every reasonable expectation depends on similarity’. For Quine, both humans and animals possess an innate standard of similarity useful for making appropriate inductions. Without this prior notion no learning or generalization can take place.

Despite the subjectivity of this primitive similarity standard, Quine believes that its uniformity across humans makes the inductive learning of verbal behavior relatively straightforward. What guarantees, however, that our ‘innate subjective spacing of qualities’ matches up with appropriate groupings in nature? Here Quine appeals to an evolutionary explanation: without such a match, and thus the ability to make appropriate inductions, survival is unlikely.

Like Hume, then, Quine proposes a naturalistic account of inductive inference, but in addition to the instinctive habit of association, he proposes an innate similarity space. Furthermore, Quine argues that this primitive notion of similarity is supplemented, as we advance from infant to adult, and from savage to scientist, by ever more developed senses of ‘theoretical’ similarity. The development of such theoretical kinds, by the regrouping of things, or the introduction of entirely new groupings, arises through ‘trial-and-error theorizing’. In Goodman’s terms, novel projections on the basis of second-order inductions become entrenched if successful. Although this progress from primitive to theoretical similarity may actually engender a qualitative change in our reasoning processes, the same inductive tendencies apply throughout. Thus, whether we infer heat from a flame, or a neutrino from its path in a bubble-chamber, or even the downfall of an empire from the dissatisfaction of its workers, all such inferences rest on our propensity to group kindred entities, and project them into the future on this basis.

For Quine, our notions of similarity and the way in which we group things become increasingly sophisticated and abstract, culminating, he believes, in their eventual removal from mature science altogether. This conclusion seems to sit uneasily with his claims about theoretical similarity. Nevertheless, as mere humans we will always be left with a spectrum of similarity notions, and systems of kinds, applicable as the context demands; hence the co-existence of a variety of procedures for carrying out inductive inference, a plurality that appears to be echoed in recent cognitive psychology (e.g., Cheng & Holyoak, 1985).

Both Goodman and Quine say very little about the notion of causality. This is probably a hangover from the logical empiricist view of science that sought to avoid all reference to causal relations in favor of logical ones. Contemporary philosophical accounts have striven to re-instate the notion of causality into induction (Glymour, 2001; Lipton, 1991; Miller, 1987).

Miller and Lipton provide numerous examples of inductive inferences that depend on the supposition of, or appeal to, causal relations. Indeed, Miller proposes a definition of inductive confirmation as causal comparison: hypotheses are confirmed by appropriate causal accounts of the data-gathering process. Armed with this notion, he claims that Goodman's new riddle of induction is soluble. It is legitimate to project 'green' but not 'grue' because only 'green' is consistent with our causal knowledge about color constancy, and the belief that no plausible causal mechanism supports spontaneous color change. He argues that any adequate description of inductive reasoning must allow for the influence of causal beliefs. Further development of such an account, however, awaits a satisfactory theory of causality (for recent advances see Pearl, 2000).

In summary, tracing the progress of philosophical analyses suggests a blueprint for a descriptive account of inductive reasoning – a mind that can extract relations of similarity and causality and apply them to new categories in relevant ways. In subsequent sections we argue that this is the same picture that is emerging from empirical work in psychology.

Empirical background

Experimental work in psychology on how people determine the projectibility of a predicate has its roots in the study of generalization in learning. Theories of learning frequently were attempts to describe the shape of a generalization gradient for a simple predicate applied to an even simpler class, often defined by a single dimension. For example, if an organism learned that a tone predicts food, one might ask how the organism would respond to other tones. The function describing how a response (like salivation) varies with the similarity of the stimulus to the originally trained stimulus is called a generalization gradient. Shepard (1987) has argued that such functions are invariably negatively exponential in shape.

If understood as general theories of induction, such theories are necessarily reductionist in orientation. Because they only consider the case of generalization along specific dimensions that are closely tied to the senses (often spectral properties of sound or light), the assumption is, more or less explicitly, that more complex predicates can be decomposed into sets of simpler ones. The projectibility of complex predicates is thus thought to be reducible to generalization along more basic dimensions.

Reductionism of this kind is highly restrictive. It requires that there exist some fixed, fundamental set of dimensions along which all complex concepts of objects and predicates can be aligned. This requirement has been by and large rejected for many reasons. One problem is that concepts tend to arise in systems, not individually. Even a simple linguistic predicate like “is small” is construed very differently when applied to mice and when applied to elephants. Many predicates that people reason about are emergent properties whose existence depends on the attitude of a reasoning agent (consider “is beautiful” or a cloud that “looks like a mermaid”). So we can’t simply represent predicates as functions of simpler perceptual properties. Something else is needed, something that respects the information we have about predicates via the relations of objects and predicates to one another.

In the 1970’s, the answer proffered was similarity (Goldstone & Son, Chapter 1, this volume). The additional information required to project a predicate was the relative position of a category with respect to other categories; the question about one category could be decided based on knowledge of the predicate’s relation to other (similar) categories (see Medin & Rips, Chapter 6, this volume). Prior to the 1970’s, similarity had generally been construed as a distance in a fairly low dimensional space (Shepard, 1980). In 1977, Tversky proposed a new measure that posited that similarity could be computed over a large number of dimensions, that both common and distinctive features were essential to determine the similarity between any pair of objects, and, critically, that the set of features used to measure similarity were context dependent. Features depended on their diagnosticity in the set of objects being compared and on the specific task used to measure similarity. Tversky's contrast model of similarity would, it was hoped, prove to have sufficient representational power to model a number of cognitive tasks including categorization and induction.

The value of representing category structure in terms of similarity was reinforced by Rosch's (1973) efforts to construct a similarity-based framework for understanding natural categories. Her seminal work on the typicality structure of categories and on the basic-level of hierarchical category structure provided the empirical basis for her arguments that categories were mentally represented in a way that carved the world at its joints. She imagined categories as clusters in a vast high-dimensional similarity space that were devised to maximize the similarity within a cluster and minimize the similarity between clusters. Her belief that the structure of this similarity space was given by the world and was not simply a matter of subjective opinion implies that the similarity space contains a lot of information, information that can be used for a number of tasks including inductive inference.

Rosch (1978) suggested that the main purpose of category structure was to provide the evidential base for relating predicates to categories. She attempted to motivate the basic-level as the level of hierarchical structure that maximized the usefulness of a cue for choosing a category, what she called cue validity, the probability of a category given a cue. Basic-level categories were presumed to maximize cue validity by virtue of being highly differentiated; members of a basic-level category have more common attributes than members of a superordinate and they have fewer common attributes with other categories than do members of a subordinate. Murphy (1982) observed however that this won’t work. The category with maximum probability given a cue is the most general category possible (“entity”), whose probability is 1 (or at least close to it). But Rosch’s idea can be elaborated using a measure of inductive projectibility in a way that succeeds in picking out the basic level. If the level of a hierarchy is selected by appealing to the inductive potential of the category, say by maximizing category validity, the probability of a specific feature given a category, then one is driven in the opposite direction of cue validity, namely to the most specific level. Given a particular feature, one is pretty much guaranteed to choose a category with that feature by choosing a specific object known to have the feature. By trading off category and cue validity, the usefulness of a category for predicting a feature and of a feature for predicting a category, one can arrive at an intermediate level of hierarchical structure.   Jones (1983) made this suggestion, calling it a measure of “collocation.” A more sophisticated information-theoretic analysis along these lines is presented in Corter and Gluck (1992) and Fisher (1987).

Another quite different but complementary line of work going on at about the same time as Rosch’s, with related implications for inductive inference, was Tversky and Kahneman’s (1974) development of the representativeness heuristic of probability and frequency judgment. The representativeness heuristic is essentially the idea that categorical knowledge is used to make probability judgments (see Kahneman & Frederick, Chapter 10, this volume). In that sense, it is an extension of Rosch’s insights about category structure. She showed that similarity was a guiding principle in decisions about category membership; Kahneman and Tversky showed that probability judgment could, in some cases, be understood as a process of categorization driven by similarity. To illustrate, Linda is judged more likely to be a feminist bankteller than a bankteller (despite the conjunction rule of probability which disallows this conclusion) if she has characteristic feminist traits, i.e., if she seems like she is a member of the category of feminists.

In sum, the importance of similarity for how people make inductive inferences was recognized in the 1970s in the study of natural category structure and of probability judgment and manifested in the development of models of similarity per se. Rips (1975) put these strands together in the development of a categorical induction task. He told people that all members of a particular species of animal on a small island had a particular contagious disease and asked participants to guess what proportion of other species would also have the disease. For example, if all rabbits have it, what proportion of dogs would? Rips found that judgments went up with the similarity of the two categories and with the typicality of the first (premise) category.

Relatively little work on categorical induction was done by cognitive psychologists immediately following Rips's seminal work. Instead, the banner was pursued by developmental psychologists like Carey (1985). She focused on the theoretical schema that children learn through development and how they use those schema to make inductive inferences across categories. In particular, she showed that adults and 10 year-olds used general biological knowledge to guide their inductions about novel animal properties, whereas small children based their inductions on knowledge about humans. Gelman and Markman (1986) argued that children prefer to make inductive inferences using category structure rather than superficial similarity. However, it was the theoretical discussion and mathematical models of Osherson and his colleagues, discussed below, that led to an explosion of interest by cognitive psychologists with a resulting menu of models and phenomena to constrain them.

Scope of chapter

In order to limit the scope of this chapter, in the remainder we focus exclusively on the psychology of categorical induction: How people arrive at a statement of their confidence that a conclusion category has a predicate after being told that one or more premise categories do. As Goodman's (1955) analysis makes clear, this is a very general problem. Nevertheless, we will not address a number of issues related to induction. For example, we will not address how people go about selecting evidence to support an hypothesis (see Klayman & Ha, 1987; Doherty et al., 1996; Oaksford & Chater, 1994). We will not address how people discover hypotheses but rather focus only on their degree of certainty in a pre-specified hypothesis (cf. the distinction between the contexts of discovery and confirmation, Reichenbach, 1938). This rules out a variety of work on the topic of hypothesis discovery (e.g., Klahr, 2000; Klayman, 1988). Relatedly, we will not cover the variety of work on the topic of cue learning, how people learn the predictive or diagnostic value of stimuli (see the chapter by Cheng & Buehner, this volume).

Most of our discussion will concern the evaluation of categorical arguments, such as

<u>Boys use GABA as a neurotransmitter.</u>
Therefore, girls use GABA as a neurotransmitter.

that can be written schematically as a list of sentences:

$P_1 ... P_n/C$

in which the P_i are the premises of an argument and C is the conclusion. Each statement includes a category (e.g., Boys) to which is applied a predicate (e.g., use GABA as a neurotransmitter). In most of the examples discussed, the categories will vary across statements whereas the predicate will remain constant. The general question will be how people go about determining their belief in the conclusion of such an argument after being told that the premises are true. We'll discuss this question both by trying to describe human judgment as a set of phenomena and by trying to explain the existence of these phenomena in terms of more fundamental and more general principles. The phenomena will concern judgments of the strength of categorical arguments or the convincingness of an argument or some other measure of belief in the conclusion once the premises are given (reviewed by Heit, 2000).

One way to represent the problem we address is in terms of conditional probability. The issue can be construed in terms of how people make judgments of the following form:

P(Category C has some property | Categories $P_1 ... P_n$ have the property).

Indeed, some of the tasks we discuss involve a conditional probability judgment explicitly. But even those that don't, like argument strength, can be directly related to judgments of conditional probability.

Most of the experimental work we address attempts to restrict attention to how people use categories to reason by minimizing the role of the predicate in the reasoning process. To achieve this, arguments are usually restricted to "blank" predicates, predicates that use relatively unfamiliar terms (like "use GABA as a neurotransmitter") so that they don't contribute much to how people reason about the arguments (Osherson, Smith, Wilkie, López, & Shafir, 1990). They do contribute some however. For instance, all the predicates applied to animals are obviously biological in nature, thus suggesting that the relevant properties for reasoning are biological. Lo, Sides, Rozelle, and Osherson (2002) characterize blank predicates as "indefinite in their application to given categories, but clear enough to communicate the kind of property in question" (p. 183).

Philosophers like Carnap (1950) and Hacking (2001) have distinguished intensional and extensional representations of probability (sometimes called epistemic vs. aleatory representations). Correspondingly in psychology we can distinguish modes of inference that depend on assessment of similarity structure and modes that depend on analyses of set structure (see Lagnado & Sloman, in press, for an analysis of the correspondence between the philosophical and psychological distinctions). We refer to the former as the inside view of category structure and the latter as the outside view (Tversky & Kahneman, 1983; Sloman & Over, 2003). In this chapter, we focus on induction from the inside, via similarity structure. We thus neglect a host of work concerning, for example, how people make conditional probability judgments in the context of well-defined sample spaces (e.g., Johnson-Laird et al., 1999), reasoning using explicit statistical information (e.g., Nisbett, 1993), and the relative advantages of different kinds of representational format (e.g., Tversky & Kahneman, 1983).

Two theoretical approaches to inductive reasoning

A number of theoretical approaches have been taken to the problem of categorical induction in psychology. Using broad strokes, the approaches can be classified into two groups:

- Similarity-based induction

- Induction as scientific methodology

We discuss each in turn. As will become clear, the approaches are not mutually exclusive both because they overlap and because they sometimes speak at different levels of abstraction.

I. Similarity-based induction

Perhaps the most obvious and robust predictor of inductive strength is similarity. In the simplest case, most people are willing to project a property known to be true of (say) crocodiles to a very similar class, like alligators, with some degree of confidence. Such willingness exists either because similarity is a mechanism of induction (Osherson et al., 1990) or because induction and similarity judgment have some common antecedent (Sloman, 1993). From the scores of examples of the representativeness heuristic at work (Tversky & Kahneman, 1974) through Rosch's (1973) analysis of typicality in terms of similarity, a strong correlation between probability and similarity is more the rule than the exception. The argument has been made that similarity is not a real explanation at all (Goodman, 1972; see the review in Sloman & Rips, 1998) and phenomena exist that contradict prediction based only on similarity (e.g., Gelman & Markman, 1986). Nevertheless, similarity remains the key construct in the description and explanation of inductive phenomena.

Consider the similarity and typicality phenomena (Rips, 1975; Osherson et al., 1990; López, Atran, Coley, Medin, & Smith, 1997):

Similarity

Arguments are strong to the extent that categories in the premises are similar to the conclusion category. For example,

Robins have sesamoid bones.
Therefore, sparrows have sesamoid bones.

is judged stronger than

Robins have sesamoid bones.
Therefore, ostriches have sesamoid bones.

because robins are more similar to sparrows than to ostriches.

Typicality

The more typical premise categories are of the conclusion category, the stronger is the argument. For example, people are more willing to project a predicate from robins to birds than from penguins to birds because robins are more typical birds than penguins.

The first descriptive mathematical account of phenomena like these expressed argument strength in terms of similarity. Osherson et al. (1990) posited the similarity-coverage model that proposed that people make categorical inductions on the basis of two principles, similarity and category coverage. Category coverage was actually cashed out in terms of similarity. According to the model, arguments are deemed strong to the degree that premise and conclusion categories are similar and to the degree that premises "cover" the lowest-level category that includes both premise and conclusion categories. The idea is that the categories present in the argument elicit their common superordinate, in particular, the most specific superordinate that they share. Category coverage is determined by the similarity between the premise categories and all the categories contained in this lowest-level superordinate.

Sloman (1993) proposed a competing theory of induction that reduces the two principles of similarity and category coverage into a single principle of feature coverage. Instead of appealing to a class inclusion hierarchy of superordinates and subordinates, this theory appeals to the extent of overlap amongst the properties of categories. Predicates are projected from premise categories to a conclusion category to the degree that the previously known properties of the conclusion category are also properties of the premise categories; specifically, in proportion to the number of conclusion category features that are present in the premise categories. Both models can explain the similarity, typicality and asymmetry phenomena (Rips, 1975):

Asymmetry

Switching premise and conclusion categories can lead to arguments of different strength:

<u>Tigers have 38 chromosomes.</u>
Therefore, buffaloes have 38 chromosomes.

is judged stronger than

<u>Buffaloes have 38 chromosomes.</u>
Therefore, tigers have 38 chromosomes.

The similarity-coverage model explains it by appealing to typicality. Tigers are more typical mammals than buffaloes and therefore tigers provide more category coverage. The feature-based model explains it by appealing to familiarity. Tigers are more familiar than buffaloes and therefore have more features. So the features of tigers cover more of the features of buffaloes than vice versa.

Differences between the models play out in the analysis of several phenomena. The similarity-coverage model focuses on relations amongst categories; the feature-based model on relations amongst properties. Consider diversity (Osherson et al., 1990):

> ***Diversity***
>
> The less similar premises are to each other, the stronger the argument tends to be. People are more willing to draw the conclusion that all mammals love onions from the fact that hippos and hamsters love onions than from the fact that hippos and rhinos do because hippos and rhinos are more similar than hippos and hamsters.

The phenomenon has been demonstrated on several occasions with Western adults (e.g., López, 1995), though some evidence suggests the phenomenon does not always generalize to other groups. López et al. (1997) failed to find diversity effects amongst Itza' Maya. Proffitt, Coley, and Medin (2000) found that parks maintenance workers did not show diversity effects when reasoning about trees although tree taxonomists did. Bailenson, Shum, Atran, Medin, and Coley (2002) did not find diversity effects with either Itza' Maya or bird experts. There is also some evidence that children are not sensitive to diversity (Carey, 1985; Gutheil & Gelman, 1997; López, Gelman, Gutheil, & Smith, 1992). However, using materials of greater interest to young children, Heit and Hahn (2001) did find diversity effects with 5- and 6-year-olds.

The data show only mixed support for the phenomenon. Nevertheless, it is predicted by the similarity-coverage model. Categories that are less similar will tend to cover the superordinate that includes them better than categories that are more similar. The feature-based model also predicts the phenomenon, as a result of feature overlap. When categories differ, their features have relatively little overlap, and thus they cover a larger part of feature space; when categories are similar, their coverage of feature space is more redundant. As a result, more dissimilar premises are more likely to show more overlap with a conclusion category. However, this isn't necessarily so and, indeed, the feature-based model predicts a boundary condition on diversity (Sloman, 1993):

Feature exclusion

A premise category that has little overlap with the conclusion category should have no effect on argument strength even if it leads to a more diverse set of premises. For example,

Fact: German shepherds have sesamoid bones.
Fact: Giraffes have sesamoid bones.
Conclusion: Moles have sesamoid bones.

is judged stronger than

Fact: German Shepherds have sesamoid bones.
Fact: Blue whales have sesamoid bones.
Conclusion: Moles have sesamoid bones.

even though the second argument has a more diverse set of premises than the first. The feature-based model explains this by appealing to the lack of feature overlap between blue whales and moles over and above the overlap between German Shepherds and moles. To explain this phenomenon, the similarity-coverage model must make the ad hoc assumption that blue whales are not similar enough to other members of the lowest-level category including all categories in the arguments (presumably mammals) to add more to category coverage than giraffes.

Monotonicity and Nonmonotonicity

When premise categories are sufficiently similar, adding a supporting premise will increase the strength of an argument. However, a counterexample to monotonicity occurs when a premise with a category dissimilar to all other categories is introduced:

Crows have strong sternums.
Peacocks have strong sternums.
Therefore, birds have strong sternums.

is stronger than

Crows have strong sternums.
Peacocks have strong sternums.
Rabbits have strong sternums.
Therefore, birds have strong sternums.

The similarity-coverage model explains nonmonotonicity through its coverage term. The lowest-level category that must be covered in the first argument is birds because all categories in the argument are birds. But the lowest-level category that must be covered in the second argument is more general – animals – because rabbits are not birds. Worse, rabbits are not similar to very many animals and therefore the category does not contribute much to argument strength. The feature-based model cannot explain this phenomenon except with added assumptions, for example that the features of highly dissimilar premise categories compete with one another as explanations for the predicate (see Sloman, 1993).

As the analysis of nonmonotonicities makes clear, the feature coverage model differs from the similarity-coverage model primarily in that it appeals to properties of categories rather than instances in explaining induction phenomena and, as a result, in not appealling to the inheritance relations of a class inclusion hierarchy. That is, it assumes that people will not in general infer that a category has a property because its superordinate does. Instead it assumes that people think about categories in terms of their structural relations, in terms of property overlap and relations amongst properties. This is surely the explanation for the inclusion fallacy (Osherson, et al., 1990; Shafir, Smith, & Osherson, 1990):

Inclusion Fallacy

Similarity relations can override categorical relations between conclusions. Most people judge

All robins have sesamoid bones.
Therefore, all birds have sesamoid bones.

to be stronger than

All robins have sesamoid bones.
Therefore, all ostriches have sesamoid bones.

Of course, ostriches are birds so that the first conclusion implies the second and therefore the second argument must be stronger than the first. Nevertheless, robins are highly typical birds and therefore similar to other birds. Yet they are distinct from ostriches. These similarity relations determine most people's judgments of argument strength rather than the categorical relation.

An even more direct demonstration of failure to consider category inclusion relations is the following (Sloman, 1993; 1998):

Inclusion Similarity

Similarity relations can override even transparent categorical relations between premise and conclusion. People do not always judge

Every individual body of water has a high number of seiches.
Every individual lake has a high number of seiches.

to be perfectly strong even when they agree that a lake is a body of water. Moreover, they judge

Every individual body of water has a high number of seiches.
Every individual reservoir has a high number of seiches.

to be even weaker, presumably because reservoirs are less typical bodies of water than lakes.

These examples suggest that category inclusion knowledge has only a limited role in inductive inference. This might be related to the limited role of inclusion relations in other kinds of categorization tasks. For example, Hampton (1982) showed intransitivities in category verification using everyday objects. He found, for example, that people affirmed that "A car headlight is a kind of a lamp" and that "A lamp is a kind of furniture" but not "A car headlight is a kind of furniture."

People are obviously capable of inferring a property from a general to a more specific category. Following an explanation that appeals to inheritance is not difficult (I know naked mole rats have livers because all mammals have livers). But the inclusion fallacy and the inclusion similarity phenomenon show that such information is not inevitably and therefore not automatically included in the inference process.

Gelman and Markman have shown that children use category labels to mediate induction:

Naming effect

Children prefer to project predicates between objects that look similar than objects that look dissimilar. However, this preference is overridden when the dissimilar objects are given similar labels.

Gelman and Coley (1990) have shown that children as young as 2 years-old are also sensitive to the use of labels. So, on one hand, people are extremely sensitive to the information provided by labels when making inductive inferences. On the other hand, the use of structured category knowledge for inductive inference seems to be a derivative ability, not a part of the fabric of the reasoning process. This suggests that the naming effect does not concern how people make inferences using knowledge about category structure per se, because if the use of structural knowledge is not automatic, very young children would not be expected to use it. Rather, the effect seems to be about the pragmatics of language, in particular how people use language to mediate induction. The naming effect probably results from people's extreme sensitivity to experimenters' linguistic cues. Even young children apparently have the capacity to note that when an experimenter gives two objects similar labels, the experimenter is giving a hint, a hint that the objects should be treated similarly at least in the context of the experiment. This ability to take cues from others, and to use language to do so, may well be key mechanisms of human induction.

This is also the conclusion of cross-cultural work by Coley, Medin, and Atran (1997). Arguments are judged stronger the more specific the categories involved. If told that dalmations have an ulnar artery, people are more willing to generalize ulnar arteries to dogs than to animals (Osherson et al. 1990). Coley et al. compared people's willingness to project predicates from various levels of the hierarchy of living things to a more general level. For example, when told that a sub-specific category like "male black spider monkey" is susceptible to an unfamiliar disease, did participants think that the members of the folk-specific category "black spider monkey" were susceptible? And if members of the specific category were susceptible, then were members of the folk-generic category ("spider monkey")? And if members of the generic category were, then were members of the life-form category ("mammal")? Finally, assuming the life-form category displayed susceptibility, then did the kingdom ("animal")? Coley et al. found that both American college students and members of a traditional Mayan village in lowland Guatemala showed a sharp drop off at a certain point:

> ***Preferred level of induction***
>
> People are willing to make an inductive inference with confidence from a subordinate to a near superordinate up to the folk-generic level; their willingness drops off considerably when making inferences to categories more abstract.

These results are consistent with Berlin's (1992) claim that the folk-generic level is the easiest to identify, the most commonly distinguished in speech, and that it serves best to distinguish categories. One might imagine therefore that the folk-generic level would constitute the basic-level categories that are often used to organize hierarchical linguistic and conceptual categories (Rosch et al., 1976; Brown, 1958; see Murphy, 2002, for a review). Nevertheless, the dominance of generic categories was not expected by Coley et al. (1997) because Rosch et al. had found that for the biological categories tree, fish, and bird, the life-form level was the category level satisfying a number of operational definitions of the basic level. For example, Rosch et al.'s American college students preferred to call objects they were shown "tree," "fish," or "bird" rather than "oak," "salmon," or "robin."

Why the discrepancy? Why do American college students prefer to name an object a tree over an oak yet prefer to project a property from all red oaks to all oaks than from all oaks to all trees? Perhaps they simply can't identify oaks and therefore fall back on the much more general "tree" in order to name. But this begs the question: If students consider "tree" to be informative and precise enough to name things, why are they unwilling to project properties to it? Coley et al.'s (1997) answer to this conundrum is that naming depends on knowledge, names are chosen that are precise enough to be informative given what people know about the object being named. Inductive inference, they argue, also depends on a kind of conventional wisdom. People have learned to maximize inductive potential at a particular level of generality (the folk-generic) level because culture and linguistic convention specify that that's the most informative level for projecting properties (see Greenfield, Chapter 26, this volume). For example, language tends to use a single morpheme for naming generic-level categories. This is a powerful cue that members of the same generic-level have a lot in common and that therefore it's a good level for guessing that a predicate might hold across it. This idea is related to Shipley's (1993) notion of overhypotheses (cf. Goodman, 1955): that people use category-wide rules about certain kinds of properties to make some inductive inferences. For example, upon encountering a new species, people might assume that members of the species will vary more in degree of obesity than in, say, skin color (Nisbett et al., 1983) despite having no particular knowledge about the species.

This observation poses a challenge to feature- and similarity-based models of induction (Heit, 1998; Osherson et al., 1990; Sloman, 1993). These models all start from the assumption that people induce new knowledge about categories from old knowledge about the same categories. But if people make inductive inferences using not only specific knowledge about the categories at hand, but also distributional knowledge about the likelihood of properties at different hierarchical levels, knowledge that is in part culturally transmitted via language, then more enters the inductive inference process than models of inductive process have heretofore allowed.

Mandler and McDonough (1996, 1998) argue that the basic-level bias comes relatively late, and demonstrate that 14-month-old infants show a bias to project properties within a broad domain (animals or vehicles) rather than at the level usually considered to be basic. This finding is not inconsistent with Coley et al.'s (1997) conclusion for the distributional and linguistic properties that they claim mediate induction presumably have to be learned, and so finding a basic-level preference only amongst adults is sufficient for their argument. Mandler and McDonough argue that infants' predilection to project to broad domains demonstrates an initial propensity to rely on "conceptual" as opposed to "perceptual" knowledge as a basis for induction, meaning that infants rely on the very abstract commonalities amongst animals as opposed to the perhaps more obvious physical differences amongst basic-level categories (pans versus cups and cats versus dogs). Of course, pans and cups do have physical properties in common that distinguish them from cats and dogs (e.g., the former are concave, the latter have articulating limbs). And the distinction between perceptual and conceptual properties is anyway tenuous. Proximal and distal stimuli are necessarily different, i.e., even the eye engages in some form of interpretation, and a variety of evidence shows that beliefs about what is being perceived affects what is perceived (e.g., Gregory, 1973). Nevertheless, as suggested by the phenomena discussed below, induction is mediated by knowledge of categories' role in causal systems; beliefs about the way the world works influence induction as much as overlap of properties does. Mandler and McDonough's data provide evidence that this is true even for 14 month olds.

II. Induction as scientific methodology

Induction is of course not merely the province of individuals trying to accomplish everyday goals, but also one of the main activities of science. According to one common view of science (Carnap, 1966; Nagel, 1961; Hempel, 1965; but for opposing views see Popper, 1963; Hacking, 1983) scientists spend much of their time trying to induce general laws about categories from particular examples. It is natural, therefore, to look to the principles that govern induction in science to see how well they describe individual behavior (for a discussion of scientific reasoning, see Dunbar & Fugelsang, Chapter 28, this volume). Psychologists have approached induction as a scientific enterprise in three different ways.

The rules of induction

First, some have examined the extent to which people abide by the normative rules of inductive inference that are generally accepted in the scientific community. One such rule is that properties that don't vary much across category instances are more projectible across the whole category than properties that vary more. Nisbett, Krantz, Jepson, and Kunda (1983) showed that people are sensitive to this rule:

> ***Variability/Centrality***
>
> People are more willing to project predicates that tend to be invariant across category instances than variable predicates. For example, people who are told that one Pacific island native is overweight tend to think it is unlikely that all natives of the island are

overweight because weight tends to vary across people. In contrast, if told the native has dark skin, they are more likely to generalize to all natives because skin color tends to be more uniform within a race.

Sensitivity to variability does not imply however that people consider the variability of predicates in the same deliberative manner that a scientist should. This phenomenon could be explained by a sensitivity to centrality (Sloman, Love, & Ahn, 1998). Given two properties A and B such that B depends on A but A does not depend on B, people are more willing to project property A than property B because A is more causally central than B, even if A and B are equated for variability (Hadjichristidis, Sloman, Stevenson, & Over, in press). More central properties tend to be less variable. Having a heart is more central and less variable amongst animals than having hair. Centrality and variability are almost two sides of the same coin (the inside and outside views, respectively). In Nisbett et al.'s case, having dark skin may be seen as less variable than obesity by virtue of being more central, having more apparent causal links to other features of people.

The diversity principle is sometimes identified as a principle of good scientific practice (e.g., Heit & Hahn, 2001; Hempel, 1965; López, 1995). Yet Lo, Rozelle, and Osherson (2002) argue against the normative status of diversity. They consider the following argument:

Housecats often carry the parasite Floxum.
Fieldmice often carry the parasite Floxum.
All mammals often carry the parasite Floxum.

which they compare to

Housecats often carry the parasite Floxum.
Tigers often carry the parasite Floxum.

All mammals often carry the parasite Floxum.

Even though the premise categories of the first argument are more diverse (housecats are less similar to fieldmice than to tigers), the second argument might seem stronger because housecats could conceivably become infected with the parasite Floxum while hunting field mice. Even if you don’t find the second argument stronger, merely accepting the relevance of this infection scenario undermines the diversity principle which prescribes that the similarity principle should be determinative for all pairs of arguments. At minimum, it shows that the diversity principle doesn’t dominate all other principles of sound inference.

Lo et al. (2002) prove that a different and simple principle of argument strength does follow from the Bayesian philosophy of science. Consider two arguments with the same conclusion in which the conclusion implies the premises. For example, the conclusion “every single mammal carries the parasite Floxum” implies that “every single tiger carries the parasite Floxum” (on the assumption that “mammal” and “tiger” refer to natural, warm-blooded animals). In such a case, the argument with the less likely premises should be stronger. Lo et al. refer to this as the Premise Probability Principle (PPP). In a series of experiments, they show that young children in both the United States and Taiwan make judgments that conform to this principle.

Induction as naïve scientific theorizing

A second approach to induction as a scientific methodology examines the contents of beliefs, what knowledge adults and children make use of when making inductive inferences. Because knowledge is structured in a way that has more or less correspondence to the structure of modern scientific theories, sometimes to the structure of old or discredited scientific theories, such knowledge is often referred to as a "naïve theory" (Carey, 1985; Gopnik & Meltzoff, 1997; Keil, 1989; Murphy & Medin, 1985). One strong, contentful position (Carey, 1985) is that people are born with a small number of naïve theories that correspond to a small number of domains like physics, biology, psychology, etc. and that all other knowledge is constructed using these original theories as a scaffolding. Perhaps, for example, other knowledge is a metaphorical extension of these original naïve theories (cf. Lakoff & Johnson, 1980).

One phenomenon studied by Carey (1985) to support this position is:

> ***Human bias***
>
> Small children prefer to project a property from people than from other animals. Four-year-olds are more likely to agree that a bug has a spleen if told that a person does than if told that a bee does. Ten-year-olds and adults do not show this asymmetry, and project as readily from non-human animals as from humans.

Carey argues that this transition is due to a major re-organization of the child's knowledge about animals. Knowledge is constituted by a mutually constraining set of concepts that make a coherent whole in analogy to the holistic coherence of scientific theories. As a result, concepts don't change in isolation but instead as whole networks of belief are re-organized (Kuhn, 1962). On this view, the human bias occurs because a four-year-old's understanding of biological functions is framed in terms of human behavior whereas older children and adults possess an autonomous domain of biological knowledge.

A different enterprise is more descriptive; it simply shows the analogies between knowledge structures and scientific theories. For example, Gopnik and Meltzoff (1997) claim that just like scientists, both children and lay people construct and revise abstract lawlike theories about the world. In particular, they maintain that the general mechanisms that underlie conceptual change in cognitive development mirror those responsible for theory change in mature science. More specifically, even very young children project properties amongst natural kinds on the basis of latent, underlying commonalities between categories rather than superficial similarities (e.g., Gelman & Coley, 1990). So children behave like "little scientists" in the sense that their inductive inferences are more sensitive to the causal principles that govern objects' composition and behavior than to objects' mere appearance, even though appearance is, by definition, more directly observable.

Of course, analogies between everyday induction and scientific induction have to exist. As long as both children and scientists have beliefs that have positive inductive potential, those beliefs are likely to have some correspondence to the world, and the knowledge of children and scientists will therefore have to show some convergence. If children did operate merely on the basis of superficial similarities, such things as photographs and toy cars would forever stump them. Children have no choice but to be "little scientists" merely to walk around the world without bumping into things. Because of the inevitability of such correspondences, and because scientific theories take a multitude of different forms, it's not obvious that this approach, in the absence of a more fully specified model, has much to offer theories of cognition. Furthermore, proponents of this approach typically present a rather impoverished view of scientific activity, which neglects the role of social and cultural norms and practices (see Faucher et al., 2002). Efforts to give the approach a more principled grounding have begun (e.g., Gopnik, Glymour, Sobel, Schultz, Kushnir & Danks, 2004; Rehder & Hastie, 2001; Sloman, Love, & Ahn, 1998).

Lo, Rozelle, and Osherson (2002) reject the approach outright. They argue that it just doesn't matter whether people have representational structures that in one way or another are similar to scientific theories. The question that they believe has both prescriptive value for improving human induction and descriptive value for developing psychological theory is whether whatever method people use to update their beliefs conforms to principles of good scientific practice.

Computational models of induction

The third approach to induction as a scientific methodology is concerned with the representation of inductive structure without concern for the process by which people make inductive inferences. The approach takes its lead from Marr's (1982) analysis of the different levels of psychological analysis. Models at the highest level, those that concern themselves with a description of the goals of a cognitive system without direct description of the manner in which the mind tries to attain those goals or how the system is implemented in the brain are called computational models. Three kinds of computational models of inductive inference have been suggested, all of which find their motivation in principles of good scientific methodology.

Induction as hypothesis evaluation

McDonald, Samuels and Rispoli (1996) propose an account of inductive inference that appeals to several principles of hypothesis evaluation. They argue that when judging the strength of an inductive argument, people actively construct and assess hypotheses in the light of the evidence provided by the premises. They advance three determinants of hypothesis plausibility: the scope of the conclusion, the number of premises that instantiate it, and the number of alternatives to it suggested by the premises. In their experiments, all three factors were good predictors of judged argument strength, although certain pragmatic considerations, and a fourth factor -- 'acceptability of the conclusion' -- were also invoked to fully cover the results.

Despite the model's success in explaining some judgments, others, such as nonmonotonicity, are only dealt with by appeal to pragmatic postulates that are not defended in any detail. Moreover, the model is restricted to arguments with general conclusions. Because the model is at a computational level of description, it does not make claims about the cognitive processes involved in induction. As we'll see next, other computational models do offer something in place of a process model that McDonald et al.'s framework does not: a rigorous normative analysis of an inductive task.

Bayesian models of inductive inference

Heit (1998) has proposed that Bayes' rule provides a representation for how people determine the probability of the conclusion of a categorical inductive argument given that the premises are true. The idea is that people combine degrees of prior belief with the data given in the premises to determine a posterior degree of belief in the conclusion.

Prior beliefs concern relative likelihoods that each combination of categories in the argument would all have the relevant property. For example, for the argument

Cows can get disease X.
Sheep can get disease X.

Heit assumes that people can generate beliefs about the relative prior probability that both cows and sheep have the disease, that cows do but sheep don't, etc. These beliefs are generated heuristically; people are assumed to bring to mind properties shared by cows and by sheep, properties that cows have but sheep do not, etc. The prior probabilities reflect the ease of bringing each type of property to mind. Premises contribute other information as well, in this case that only states in which cows indeed have the disease are possible. This can be used to update priors to determine a posterior degree of belief that the conclusion is true.

On the basis of assumptions about what people's priors are, Heit (1998) is able to describe a number of the phenomena of categorical induction: similarity, typicality, diversity, and homogeneity. However, the model is inconsistent with nonmonotonicity effects. Furthermore, because it relies on an extensional updating rule, Bayes' rule, the model cannot explain phenomena that are non-extensional like the inclusion fallacy or the inclusion-similarity phenomenon.

Sanjana and Tenenbaum (2003) offer a Bayesian model of categorical inference with a more principled foundation. The model is applied only to the animal domain. They derive all their probabilities from an hypothesis space that consists of clusters of categories. The model's prediction for each argument derives from the probability that the conclusion category has the property. This reflects the probability that the conclusion category is an element of likely hypotheses, namely that the conclusion category is in the same cluster as the examples shown, i.e., as the premise categories, and that those hypothesized clusters have high probability. The probability of each hypothesis is assumed to be inversely related to the size of the hypothesis (the number of animal types it includes) and to its complexity, the number of disjoint clusters that it includes. This model performed well in quantitative comparisons against the similarity-coverage model and the feature-based model although its consistency with the various phenomena of induction has not been reported and is rather opaque.

The principled probabilistic foundation of this model and its good fit to data so far yield promise that the model could serve as a formal representation of categorical induction. The model would show even more promise and power to generalize however if its predictions had been derived using more reasonable assumptions about the structure of categorical knowledge. The pairwise cluster hierarchy Sanjana and Tenenbaum use to represent knowledge of animals is poorly motivated (though see Kemp & Tenenbaum, 2003, for an improvement), and there would be even less motivation in other domains (cf. Sloman, 1998). Moreover, if and how the model could explain fallacious reasoning is not clear.

Summary of induction as scientific methodology

Inductive inference can be fallacious, as demonstrated by the inclusion fallacy described above. Nevetheless, much of the evidence that has been covered in this section suggests that people in the psychologist's laboratory are sensitive to some of the same concerns as scientists when they make inductive inferences. People are more likely to project nonvariable over variable predicates, they change their beliefs more when premises are a priori less likely, and their behavior can be modeled by probabilistic models constructed from rational principles.

Other work reviewed shows that people, like scientists, use explanations to mediate their inference. They try to understand why a category should exhibit a predicate based on nonobservable properties. These are valuable observations to allow psychologists to begin the process of building a descriptive theory of inductive inference. Unfortunately, current ideas and data place too few constraints on the cognitive processes and procedures that people actually use.

Conclusions and Future Directions

We have reviewed two ways that cognitive scientists have gone about trying to describe how people make inductive inferences. We limited the scope of the problem to that of categorical induction, how people generate degrees of confidence that a predicate applies to a stated category from premises concerning other categories that the predicate is assumed to apply to. Nevertheless, neither approach is a silver bullet. The similarity-based approach has been the most productive of well-specified models and phenomena, though consideration of the relation between scientific methodology and human induction may prove the most important prescriptively and may turn out to provide the most enduring principles to distinguish everyday human induction from ideal – or at least other – inductive processes.

A more liberal way to proceed is to accept the apparent plurality of procedures and mechanisms that people use to make inductions, and see this pluralism as a virtue rather than a vice.

The bag of tricks

Many computational problems are hard because the search space of possible answers is so large. Computer scientists have long used educated guesses or what are often called heuristics or rules-of-thumb to prune the search space, making it smaller and thus more tractable at the risk of making the problem insoluble by pruning off the best answers. Kahneman and Tversky imported this notion of heuristics into the study of probability judgment (see Kahneman and Frederick, Chapter 10, this volume). They suggested that people use a set of cognitive heuristics to estimate probabilities, heuristics that were informed, that made people's estimates likely to be reasonable, but left open the possibility of systematic error in cases where the heuristics that came naturally to people had the unfortunate consequence of leading to the wrong answer.

Kahneman and Tversky suggested the heuristics of availability, anchoring and adjustment, simulation and causality to describe how people make probability judgments. They also suggested that people make judgments according to representativeness, the degree to which a class or event used as evidence is similar to the class or process being judged. Representativeness is a very abstract heuristic that is compatible with a number of different models of the judgment process. We understand it not so much as a particular claim about how people make probability judgments as the claim that processes of categorization and similarity play central roles in induction. This is precisely the claim of the similarity-based model outlined above.

We believe that the bag of tricks describes most completely how people go about making inductive leaps. People seem to use a number of different sources of information for making inductive inferences including the availability of featural informational and knowledge about feature overlap, linguistic cues about the distribution of features, the relative centrality of features to one another, the relative probability of premises, and objects' roles in causal systems.

Causal induction

Our guess is that the treasure trove for future work in categorical induction is in the development of the latter mode of inference. How do people go about using causal knowledge to make inductions? That they do is indisputable. Consider the following phenomenon due to Heit and Rubinstein (1994):

> ***Relevance***
>
> People's willingness to project a predicate from one category to another depends on what else the two categories have in common. For example, people are more likely to project "has a liver with two chambers" from chickens to hawks than from tigers to hawks but more likely to project "prefers to feed at night" from tigers to hawks than from chickens to hawks.

More specifically, argument strength depends on how people explain why the category has the predicate. In the example, chickens and hawks are known to have biological properties in common and therefore think it likely that a biological predicate would project from one to the other; tigers and hawks are known to both be hunters and carnivores and therefore "prefers to feed at night" is more likely to project between them. Sloman (1994) has shown that the strength of an argument depends on whether the premise and conclusion are explained in the same way. If the premise and conclusion have different explanations, the premise can actually reduce belief in the conclusion.

The explanations in these cases are causal; they refer to more or less well-understood causal processes. Medin, Coley, Storms, and Hayes (in press) have demonstrated 5 distinct phenomena that depend on causal intuitions about the relations amongst categories and predicates. For example, they showed

Causal asymmetry

Switching premise and conclusion categories will reduce the strength of an argument if a causal path exists from premise to conclusion. For example,

<u>Gazelles contain retinum.</u>
Lions contain retinum.

is stronger than

<u>Lions contain retinum.</u>
Gazelles contain retinum.

because the food chain is such that lions eat gazelles and retinum could be transferred in the process.

What's striking about this kind of example is the exquisite sensitivity to subtle (if mundane) causal relations that it demonstrates. The necessary causal explanation springs to mind quickly, apparently automatically, and it does so even though it depends on one fact that most people are only dimly aware of (that lions eat gazelles) amongst the vast number of facts that are at our disposal.

We do not interpret the importance of causal relations in induction as support for psychological essentialism, the view that people base judgments concerning categories on attributions of "essential" qualities: of a true underlying nature that confers kind identity unlike, for example, Kornblith (1993), Medin and Ortony (1989), and Gelman and Hirschfeld (1999). We rather follow Strevens (2001) in the claim that it's causal structure per se that mediates induction, no appeal to essential properties is required (cf. Rips, 2001; Sloman & Malt, 2003). Indeed, the causal relations that support inductive inference can be based on very superficial features that might be very mutable. To illustrate, the argument

> Giraffes eat leaves of type X.
> African tawny eagles eat leaves of type X.

seems reasonably strong only because both giraffes and African eagles can reach high leaves and both are found in Africa, hardly a central property of either species.

The appeal to causal structure is instead intended to appeal to the ability to pick out invariants and act as agents to make use of those invariants. Organisms have a striking ability to find the properties of things that maximize their ability to predict and control and humans seem to have the most widely applicable capacity of this sort. But prediction and control comes from knowing what variables determine the values of other variables, that is how one predicts future outcomes and knows what to manipulate to achieve an effect. And this is of course the domain of causality. It seems only natural that people would use this talent to reason when making inductive inferences.

The appeal to causal relations is not necessarily an appeal to scientific methodology. In fact, some philosophers like Russell (1921) have argued that theories aren't scientific until they're devoid of causal reference, and the logical empiricists attempted to exorcise the notion of causality from 'scientific' philosophy. Of course, to the extent that scientists behave like other people in their appeal to causality, then the appeal to scientific methodology is trivial.

Normative models of causal structure have recently flowered (cf. Pearl, 2000; Spirtes, Glymour, & Scheines, 1993) and some of the insights of these models seem to have some psychological validity (Sloman & Lagnado, in press). Bringing them to bear on the problem of inductive inference will not be trivial. But the effort should be made because causal modeling seems to be a critical element of the bag of tricks that people use to make inductive inferences.

References

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: MIT Press.

Carnap, R. (1950). The logical foundations of probability. Chicago: University of Chicago Press.

Carnap, R. (1966). Philosophical foundations of physics. Ed. M. Gardner, New York: Basic Books.

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996) On people's understanding of the diagnostic implications of probabilistic data. Memory and Cognition, 24, 644-654.

Faucher, L., Mallon, R., Nazer, D., Nichols, S., Ruby, A., Stich, S. & Weinberg, J. (2002). The Baby in the Lab Coat In P. Carruthers, S. Stich & M. Siegal (Eds.) The Cognitive Basis of Science. Cambridge: Cambridge University Press.

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. Machine Learning, 2, 139-172.

Frege, G. W. (1880). Posthumous writings. Blackwell, 1979.

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

Gelman, S. A. & Hirschfeld, L. A. (1999). How biological is essentialism? In D. L. Medin & S. Atran (Eds.), Folkbiology. Cambridge: MIT Press.

Gelman, S. A., & Coley, J.D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. Developmental Psychology , 26, 796-804.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. Cognition, 23, 183-209.

Glymour, C. (2001). The mind's arrows: Bayes nets and graphical causal models in psychology. Cambridge, MA: Bradford Books.

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

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (ed.), Problems and projects. New York: Bobbs-Merrill.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T, & Danks, D. (in press). A theory of causal learning in children: Causal maps and Bayes nets. Psychological Review.

Gopnik, A. & Meltzoff, A. N. (1997). Words, thoughts, and theories. Cambridge: MIT Press.

Gregory, R. L. (1973). The intelligent eye. New York: McGraw-Hill.

Gutheil, G., & Gelman, S. A. (1997). The use of sample size and diversity in category-based induction. Journal of Experimental Child Psychology, 64, 159-174.

Hacking, I. (1983). Representing and intervening: Introductory topics in the philosophy of natural science. Cambridge, England: Cambridge University Press.

Hacking, I. (2001). An introduction to probability and inductive logic. Cambridge: Cambridge University Press.

Hadjichristidis, C., Sloman, S. A., Stevenson, R. J., & Over D. E. (in press). Feature centrality and property induction. Cognitive Science.

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 University Press.

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

Heit, E. & Hahn, U. (2001). Diversity-based reasoning in children. Cognitive Psychology, 47, 243-273.

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

Hempel, C. (1965). Aspects of scientific explanation. New York: Free press.

Hume, D. (1739). A treatise of human nature. Ed. D. G. C. Macnabb, London: Collins, 1962.

Hume, D. (1748). An enquiry concerning human understanding. Oxford: Clarendon.

Jones, G. (1983). Identifying basic categories. Psychological Bulletin, 94, 423-428.

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

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. Psychological Review, 94, 211- 228.

Kornblith, H. (1993). Inductive inference and its natural ground. Cambridge: MIT Press.

Kuhn, T. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Lagnado, D. & Sloman, S.A., (in press). Inside and outside probability judgment. D. J. Koehler and N. Harvey (Eds.) Blackwell Handbook of Judgment and Decision Making.

Lipton, P. (1991). Inference to the best explanation. New York: Routledge.

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. Cognitive Science, 26, 181-206.

López, 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.

Mandler, J. M. & McDonough, L. (1998). Studies in inductive inference in infancy. Cognitive Psychology, 37, 60-96.

Marr, D. (1982). Vision. New York: W.H. Freeman and Co.

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. (in press). A relevance theory of induction. Psychonomic Bulletin and Review.

Medin, D. L. & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony

(Eds.), Similarity and analogical reasoning. New York: Cambridge University Press.

Miller, R.W. (1987). Fact and method. Princeton: Princeton University Press.

Murphy, G. L. (2002). The big book of concepts. Cambridge, MA: MIT Press.

Nisbett, R. E. (Ed.) (1993). Rules for reasoning. Hillsdale, NJ: Erlbaum.

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

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. Cambridge, MA: Cambridge University Press.

Quine, W.V. (1970). Natural kinds. In N. Rescher (Ed.) Essays in honor of Carl G. Hempel. Dordrecht: D. Reidel.

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.

Reichenbach, H. (1938). Experience and prediction. Chicago: University of Chicago Press.

Rips, L. (2001). Necessity and natural categories. Psychological Bulletin, 127, 827-852.

Rosch, E.H. (1973) Natural categories. Cognitive Psychology, 4, 328-350.

Russell, B. & Whitehead, A..N. (1925). Principia Mathematica. Cambridge: Cambridge University Press.

Sanjana, N. E. & Tenenbaum, J. B. (2003).Bayesian models of inductive generalization. In Becker, S., Thrun, S., and Obermayer, K. (Eds.), Advances in Neural Processing Systems 15. MIT Press.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. Science, 210, 390-398.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. Science, 237, 1317-1323.

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., & Lagnado, D.A. (in press). Causal invariance in reasoning and learning. In B. Ross (Ed.) Handbook of Learning and Motivation.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. Cognitive Science, 22, 189-228.

Sloman, S. A. & Malt, B. C. (2003). Artifacts are not ascribed essences, nor are they treated as belonging to kinds. Language and Cognitive Processes, 18, 563-582.

Sloman, S. A. & Over, D. (2003). Probability judgment from the inside and out. In D. Over

(Ed.) Evolution and the Psychology of Thinking: The Debate, pp. 145-169. New York: Psychology Press.

Sloman, S. A. & Rips, L. J. (1998). Similarity as an explanatory construct. Cognition, 65, 87-101.

Spellman, B. A., López, A., & Smith, E. E. (1999). Hypothesis testing: Strategy selection for generalizing versus limiting hypotheses. Thinking & Reasoning, 5, 67-91.

Spirtes, P., Glymour, C. & Scheines, R. (1993). Causation, prediction, and search. New York: Springer-Verlag.

Strevens, M. (2001). The essentialist aspect of naive theories. Cognition, 74, 149-175.

Suppes, P. (1994). Learning and projectibility. In D. Stalker (Ed.) Grue: the new riddle of induction. Chicago and La salle: Open Court.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. Science, 185, 1124-1131.

Acknowledgements

We thank Uri Hasson and Marc Buehner for comments on an earlier draft. This work was funded by NASA grant NCC2-1217.

Keywords

Induction; projectibility; similarity; causality; categorization