Knowledge and Concept Learning

Evan Heit

University of Warwick

In K. Lamberts and D. Shanks (Eds.), Knowledge, Concepts, and Categories, 7-41, 1997. Psychology Press.

It has been remarked of sophisticated computer data bases that "everything is deeply intertwingled" (Nelson, 1987). This observation also applies especially well to concept learning by humans. Conceptual knowledge has a highly interrelated nature. What a person learns about a new category is greatly influenced by and dependent on what this person knows about other, related categories.

For example, imagine two people who are learning to drive a manual transmission automobile. In effect, these people are learning about a new concept, manual transmission cars. Say that one person has had many years of experience driving cars with automatic transmissions, and the other person has never driven a car before. The first person's learning will be facilitated greatly by previous knowledge of the category automatic transmission cars, so that this person will be able to quickly find and operate the steering wheel, brakes, radio, etc. in the new car. Yet this prior knowledge would not be of much help as this person is learning about how to shift gears in manual transmission cars. In fact, all of this experience with automatic transmissions might make it especially difficult to learn to operate a manual transmission. Now imagine the situation of the second person, who has never driven before. Overall, this person will probably learn very slowly compared to the first person, because of this person's lack of relevant prior knowledge. This second person's learning will likely be a drawn-out process with much trial-and-error practice involved. On the positive side, though, the second person might have some advantage over the first person in learning how to shift gears, because the second person would not have to overcome negative transfer from experience with automatic transmissions.

As another example, imagine that you are an explorer visiting a remote island, with the purpose of writing a book about the people that you see there. You bring to this island many forms of prior knowledge that will guide you in learning about these new people. For example, based on your experiences in other places, you would expect to see males and females, younger and older people, shy people and arrogant people. You would also have certain hypotheses at a more abstract level, for example, that the clothes that someone wears may be related to the person's age and gender. (Goodman, 1955, referred to such abstract hypotheses as overhypotheses.) In a way, these biases due to previous knowledge might seem to be undesirable. After all, wouldn't be it be better to be a detached, unbiased observer? However, such biases can make learning much more efficient. Without any prior expectations about what the important categories are on this new island, you would likely spend too much time on unimportant information. For example, you might spend the first month of your visit categorizing people in terms of whether they have small ears or large

ears, and the second month trying to notice the relation between ear size and how fast people walk. Without the guidance of your prior knowledge, you could spend an interminable amount of time trying to learn about all the possible categories and the relations among categories. Clearly, some use of prior knowledge of old categories would be critical in learning about the new categories on this island. (See Keil, 1989, and Peirce, 1931-1935, for related arguments.)

The past decade has been an exciting time for categorization research. Our understanding of the "intertwingledness" or interrelatedness of concept learning has been building steadily. There are numerous situations, such as learning about new objects (like manual transmission cars) or visiting new locations (whether they are new islands or just new restaurants) in which category learning is influenced by what is already known. This chapter will review the experimental evidence for the claim that concept learning depends heavily on prior knowledge, and describe the different ways that prior knowledge has an influence. Furthermore, this chapter will discuss current models of categorization and concept learning with the aim of improving these models to address the important influences of prior knowledge. Finally, inductive reasoning and memory, cognitive abilities that are closely related to categorization, will be discussed in terms of effects of background knowledge.

Theoretical Arguments

The seminal paper concerning knowledge effects on concept learning was written by Murphy and Medin (1985). They contrasted two approaches to describing concept learning, which they referred to as similarity-based and theory-based. According to similarity-based approaches, there is a simple way to tell whether something belongs to a particular category: You assess the similarity between the item and what is known about the category (see also Rips, 1989). The more similar item X is to what is known about category C, the more likely you will place X in category C. This similarity-based approach does appear to be a reasonable idea, and it is consistent with several existing accounts of how people learn about categories. For example, take a standard prototype account (Hampton, 1993; Rosch & Mervis, 1975) of how you might learn about a category such as a novel kind of bird. You would observe members of this species of bird, and remember typical features or characteristics of these birds. These features would be summarized as a prototype, representing the average member of the species (e.g., light brown, fourteen-inch wingspan, lives in

tree-tops). To judge whether another bird belongs to this species, you would evaluate the similarity between this bird and the prototypical list of features.

Murphy and Medin argued that although a similarity-based approach to categorization may be a reasonable start, it will ultimately prove to be incomplete. As illustrated by the earlier example of the explorer visiting an island, there may be so much information available that it will be difficult to simply observe and remember everything. A category learner needs some constraints or biases on what to observe. A related point is that the learner needs to figure out how to describe observations in terms of features. Except perhaps in nature books, birds do not come already labeled with tags such as "light brown" and "lives in tree-tops." Such descriptions are inferred and applied by the learner. In addition, people have knowledge about the causal relations between these features, that would not be captured by a feature list. For example, it is reasonable to expect that smaller birds will tend to live closer to the ground and larger birds would be more likely to live in tree-tops, because larger birds can better sustain exposure to wind and severe weather.

These critical influences of knowledge are not explained by similarity-based approaches, Murphy and Medin argued. In contrast, theory-based approaches would consider people's knowledge about the world, including their intuitive theories about what features are important to observe and how they are related to each other. The Murphy and Medin article did not propose a particular theory-based model of categorization so much as to lay out the challenges that researchers would face in developing a more complete account of categorization that addresses the influences of knowledge. Much of the categorization research published after Murphy and Medin (1985) has presented experimental evidence for, and more detailed empirical accounts of, knowledge effects on concept learning. Also, some work has begun to develop more complete models of categorization that address some of the issues raised by Murphy and Medin. The next two sections of this chapter will review the empirical work on knowledge and concept learning, and the following section will discuss categorization models that address these experimental results.

Experimental Evidence for Specific Influences of Knowledge

At this point, there is quite a bit of amassed evidence on ways that knowledge influences category learning. Before describing this evidence in detail, it is possible to draw some generalizations about what is known. Perhaps the most fundamental generalization is that in learning about new categories, people act as if these categories will be consistent with previous knowledge.

People seem to act with economy, so that previous knowledge structures are reused when possible. This generalization is apparent in a few different ways. In general it is easier to learn a new category when it is similar to a previously-known category, as in the earlier example of learning about manual transmission cars. Also, people's beliefs about new categories include their knowledge from other categories; in effect, there is leakage from one category to another. Likewise, people's strategies in learning new categories are consistent with their beliefs about other categories. For example, an explorer's strategies in studying people on a new island would reflect what the explorer knows about the social structure of other places. In the following sections, four different kinds of experimental results will be described, indicating different effects of prior knowledge.

Integration Effects

One of the basic influences of prior knowledge on the learning of new categories is integration of prior knowledge with new observations (Heit, 1994). That is, the initial representation of a new category is based on prior knowledge, and this representation is updated gradually as new observations are made. For example, imagine that you are walking through some forest for the first time. A nearby forest has large and aggressive birds, so you initially expect the same in the new forest. However, most of the birds you first see are small and unaggressive. As you observe more birds, you gradually revise your beliefs to reflect the local conditions. After just a few observations, your beliefs about the new category of birds might represent an average of your prior knowledge and what you observe. With an even larger number of observations, your beliefs mostly reflect the data from the new forest (small and unaggressive birds) rather than your previous beliefs based on the other forest (large and aggressive). This process is similar to an anchor-and-adjust method of estimation, which Tversky and Kahneman (1974) have argued is a widespread form of reasoning. Also, this process is similar to Bayesian statistical procedures for estimation, in which an initial estimate is revised as new data are encountered (Edwards, Lindman, & Savage, 1963; Raiffa & Schlaifer, 1961).

Recent experiments by Heit (1994, 1995) obtained results that are consistent with an integration account. Instead of being brought to a forest, the subjects in these experiments were shown descriptions of people in a fictional city. Heit assessed subject's initial beliefs about the city as well as their beliefs after they observed members of categories from this city. For example, the subjects learned about a category of joggers. Initially, subjects expected that about 75% of these joggers would own expensive running shoes. Some subjects then saw descriptions of joggers such

that 75% did own expensive running shoes, whereas other subjects saw other proportions (0%, 25%, 50%, 100%) of joggers with expensive running shoes. In their final judgments, subjects acted as if they were taking a weighted average of the expected proportion of joggers with expensive running shoes and the observed proportion. For example, subjects who observed 75% expensive running shoes continued to make judgments of about 75%. Subjects who observed only 25% running shoes ultimately made judgments of about 50%. Furthermore, Heit found that subjects who were given a larger number of descriptions of people in the city tended to discount their prior knowledge more, again consistent with the integration account. (For further experimental evidence of integration effects, see Hayes and Taplin, 1992, 1995.)

Clinical psychologists sometimes show similar anchoring effects in their categorizations, or diagnoses, of patients (see Mumma, 1993, for a review). Clinicians often show suggestion effects, so that their diagnoses represent an integration of their previous knowledge and their own observations. A typical source of suggestion effects would be a diagnosis made by a colleague. For example, a clinician might categorize a patient as having borderline personality disorder if another clinician has previously reported this diagnosis, even if the patient's symptoms would fit with a number of other disorders as well. Here, the previous clinician's analysis of the patient serves as an anchor or initial representation when the new clinician learns about the patient.

A critical aspect of integration effects is the initial category representation that people assemble based on prior knowledge. Ward (1994) has developed a technique for studying these initial representations. This work sheds light on how people borrow information from related categories as they begin learning about a new category. Ward's task placed people in a creative situation in which they imagined the members of new categories. For example, subjects were asked to draw pictures of animals that might appear on another planet. These imagined animals were very likely to have familiar appendages such as arms, legs, or wings, and to have sense organs such as eyes and ears. Consistent with the idea of integration, Ward concluded that these initial category representations contained a great deal of specific, borrowed information from established categories of animals on Earth.

Selective Weighting Effects

Several researchers (Keil, 1989; Murphy & Medin, 1985; Murphy & Wisniewski, 1989) have argued that selective weighting effects of prior knowledge are critical in category learning. That is, previous knowledge leads us to selectively attend to certain features or certain observations

during concept learning, thereby narrowing the space of hypotheses to be considered. In the earlier example, an explorer could have used previous knowledge to focus on the relation between age and clothing rather than the relation between ear size and speed of walking. Without such selective weighting of relevant information, concept learning would be very slow and difficult.

Pazzani (1991) investigated the issue of selective attention by teaching subjects about categories of balloons. Subjects were instructed either to learn a category of balloons that inflate or to learn a category that was simply labeled "Alpha." A pretest showed that subjects expected that stretching a balloon would facilitate inflation and that adults would be more successful than children at inflation. It was assumed that subjects in the Inflate conditions (but not in the Alpha conditions) would be influenced by their prior knowledge of what it takes to inflate a balloon. The stimuli in this experiment were pictures of persons with balloons. The pictures varied on four dimensions: adult or child, stretched balloon or balloon dipped in water, yellow or purple balloon, and small or large balloon. In some conditions of this experiment, the Inflate (or Alpha) category was defined by a disjunctive rule: These balloons must be stretched or inflated by an adult. Note that this rule is relevant to subjects' knowledge about inflating balloons. Pazzani found that category learning was much faster in the Inflate condition than in the Alpha condition. This result may be explained by subjects in the Inflate conditions paying special attention to the age and stretching features. Prior knowledge about these relevant features would be helpful because the concept was defined in terms of age and stretching.

Several other researchers have obtained results that they explained in terms of selective weighting (e.g., Hayes & Taplin, 1992, 1995; Keleman & Bloom, 1994; Medin, Wattenmaker, & Hampson, 1987; Murphy & Wisniewski, 1989; Wisniewski, 1995). For example, Medin et al. (1987) used a sorting task to study how people construct categories. Medin et al. found that when people sorted items into groups, they were especially likely to be influenced by pairs of dimensions that were causally related according to prior knowledge. For example, in sorting medical patients who were described by several symptoms, subjects were likely to sort on the basis of a pair of related symptoms such as dizziness and earache, presumably because these dimensions were given extra weight. Considering the theoretical arguments by Keil (1989), Murphy and Medin (1985), and Peirce (1931-1935), it does seem plausible that selective weighting due to previous knowledge is a central part of category learning.

Feature Interpretation Effects

Another important influence of prior knowledge on learning is to help people interpret and represent what they observe. Psychologists such as Asch (1946) made this point with stimuli that describe personality traits. According to Asch's change of meaning hypothesis, a feature such as "intelligent" would be interpreted differently in the sentences "Sara is friendly and intelligent" and "Mary is ruthless and intelligent." Sara's intelligence is of a quite a different kind than Mary's, because friendly or ruthless lead us to interpret intelligent differently (but see N. H. Anderson, 1991, for an argument against this point). If a single adjective can influence interpretation so much, then the rich knowledge that people bring to category learning might well have even stronger effects. A dramatic example of knowledge effects on learning was provided by Lesgold, Glaser, Rubinson, Klopfer, Feltovich, and Wang (1988). Lesgold et al. studied expert and novice radiologists, on the task of interpreting chest x-rays and making diagnoses. There were numerous interpretation differences between the two groups, attributable to their differences in prior knowledge about human anatomy and x-ray technology. For example, the experts were better able to distinguish the appearance of diseased tissue from the appearance of artifacts on the x-ray film. Also, the experts were more likely than the novices to describe a three-dimensional representation or model of the patient rather than simply focus on simple two-dimensional cues such as a shadow on the film.

Closely related to the work of Lesgold et al. on learning about individual cases, there has been some more recent work on learning about categories. Wisniewski and Medin (1991, 1994b) demonstrated influences of prior knowledge on interpretation of category members. In their studies, the subjects observed drawings done by children. They learned about two categories of drawings, such as drawings done by city children versus farm children, or drawings done by creative children versus noncreative children. The category labels were randomly assigned by the experimenters to a particular drawing and often had a dramatic effect on how features of the drawing were interpreted. For example, one circular configuration of lines on a drawing was interpreted as a purse when the picture was assigned to the city category; in other situations this same configuration was interpreted as a pocket. Similarly, the clothing in another drawing was interpreted as either being a farm uniform or a city uniform depending on the category assignment.

The experiments by Wisniewski and Medin (1991, 1994b) were in some ways ideally suited to study influences of knowledge on feature interpretation, because their stimuli were somewhat ambiguous drawings that indeed needed to be interpreted. In contrast, for many experiments in which subjects learn categories, the features are already given in a much less

ambiguous way. For example, in a typical experiment, subjects might learn about lists of features that are familiar medical symptoms, such as runny nose and high fever (e.g., Medin & Schaffer, 1978). In such experiments, the representation (simple feature lists) is more or less given to the subject. In contrast, in learning about ambiguous drawings, and probably in many real-word concept learning situations, people must build the representations that would be used to describe category members (see also Goldstone, 1994; Murphy, 1993; Schyns & Murphy, 1994).

Facilitation Effects

Some effects of prior knowledge are best described as simply being overall facilitation of learning. It seems plausible that learning about certain kinds of category structures might be more or less facilitated depending on the prior knowledge that is accessed, e.g., depending on the kind of category structure that is expected. Medin and Schwanenflugel (1981) distinguished between two kinds of classification structures, linearly separable and nonlinearly separable. If a pair of categories, A and B, are linearly separable, then by definition it is possible to classify a new stimulus, X, using a simple linear rule. One such linear rule would be to count whether X has more characteristic features of category A or of category B. In contrast, if A and B overlap to the extent that they are nonlinearly separable, then no linear rule will allow perfect discrimination between members of the two categories. Medin and Schwanenflugel found that people can learn both kinds of category structures, with no great advantage for one kind of category structure over the other. However, Wattenmaker, Dewey, T. Murphy, & Medin (1986) investigated the influences of background knowledge on learning these two kinds of structures (see also Nakamura, 1985). For example, if your prior knowledge leads you to expect linearly separable categories, would that facilitate the learning of a linearly separable structure?

In the Wattenmaker et al. (1986, Experiment 1) study, half of the subjects learned about linearly separable categories of people and half of the subjects learned about nonlinearly separable categories. Also, in the Trait conditions, the stimulus dimensions were labeled to promote remindings of personality categories, such as honest versus dishonest. For example, some subjects saw person descriptions in terms of behaviors that were either honest (e.g., returning a lost wallet) or dishonest (e.g., pretending to enjoy shopping). The subjects were trained repeatedly on category members until they reached a learning criterion. In the Control conditions, the stimuli were composed of unrelated traits, such as one concerning honesty, one concerning cautiousness, and one concerning cooperativeness, which would not promote the retrieval of coherent prior

categories. The first main result was that overall, remindings of prior knowledge helped subjects learn the categories faster: People in the Trait conditions performed better than people in the Control conditions. Making the task more meaningful facilitated category learning (see also Murphy & Allopenna, 1994). Second, subjects especially showed facilitation from prior knowledge when they learned about a linearly separable category structure. It appeared that people already had simple linear rules for distinguishing between honest and dishonest people by counting up the number of honest and dishonest behaviors. Thus, learning was most efficient when the structure to be learned was compatible with the structure that was expected according to prior knowledge.

In more recent work, Wattenmaker (1995) has investigated whether these knowledge facilitation effects depend on specific category knowledge or on more general knowledge. That is, when people are facilitated in learning about a new category, is this facilitation due to a close match between specific information in the new category and specific information in a previously known concept? Or is it due to a general congruence with an abstract structure, such as the linearly separable structure? Wattenmaker compared category learning using stimuli from two different general domains, social categories and object categories. An overall difference between these two domains might suggest that people apply different general knowledge structures in learning about these two kinds of categories. Indeed, Wattenmaker found that overall, people were facilitated in learning about linearly separable categories in the social domain, and people learned object categories better when they were nonlinearly separable. However, this pattern was only evident when the new categories to be learned closely matched previously known concepts. For example, people favored learning linearly separable structures for a familiar classification such as introverts versus extroverts, but not for unfamiliar social groupings. Thus it appears that the knowledge facilitation effects reported by Wattenmaker (1995) and Wattenmaker et al. (1986) depended on remindings of rather specific knowledge of particular categories.

The question does remain though, how does more general knowledge influence category learning? Even when someone is not reminded of a specific pre-existing concept, can prior knowledge affect learning?

Influences of More General Knowledge

Children's Learning of Concepts and Names

Perhaps the most dramatic example of concept learning is the performance of young children, who can learn up to 15,000 new words for things by age six (Carey, 1978). Of course, learning a new word and learning a new concept are not the same, but they are closely related (Clark, 1983). For example, a child's knowing the word "dog" and having the concept of dog are two different achievements. Knowing a concept might precede learning its name or alternately, hearing a name for an object might lead to further investigation of the concept (e.g., Waxman, Shipley, & Shepperson, 1991). Early concept learning by children appears to be guided by rather general principles or knowledge structures. Given the large number of concepts learned by children and the systematic biases that are apparent in this learning, it is plausible that the children are being influenced by general knowledge rather than by specific knowledge about other categories.

Markman (1989, 1990) suggested, and reviewed evidence for, certain constraints that would guide category learning by children. First, according to the whole object assumption, a novel category label is more likely to refer to a whole object than to its parts. Upon hearing a category label such as "dog" for the first time, a child would assume that this label refers to a dog rather than to some part of a dog such as its wagging tail. Second, according to the taxonomic assumption, learners will tend to use new words as taxonomic category labels rather than as ways to group things by other relations. For example, after a child has learned about his or her first dog, the child would extend this label to other animals that appear to be in the same taxonomic category--other dogs--rather than extending the label to objects that are otherwise associated with the dog. That is, the child would not call the dog's leash a "dog," or call the dog's owner a "dog." Third, the mutual exclusivity assumption would provide further guidance in early category learning. In following this assumption, a child would favor associating particular objects with just one category label. Thus, when learning a new category label, the child would look for some object for which he or she does not already know a label. For example, say that a child already knows the word "dog," and sees a dog being pulled on a leash. Upon hearing the word "leash" for the first time, the child might hypothesize that this term refers to the leash rather than to the dog, because the dog already has a known category label.

These three constraints might seem obvious to an adult who has already learned a language. Yet imagine a child trying to learning thousands of category labels without these assumptions (Quine,

1960). In a relatively simple situation of a girl walking in a park with a dog on a leash, the category label "dog" might refer to the girl, the park, the dog, the leash, some part of the girl, the park, the dog, or the leash, or some relation between any of these things. It appears that some application of general knowledge to this potentially confusing situation would be extremely helpful and indeed necessary.

Closely related to Markman's whole object assumption is the shape bias (see Landau, 1994, and Ward, 1993, for reviews). The shape bias is another proposed general constraint on the learning of category labels, such that young children would tend to pay attention to overall shape of an object rather than its texture or size. The shape bias is a kind of selective weighting effect, and as such it fits well with the proposals of Keil (1989) and Murphy and Medin (1985) regarding the selective effects of prior knowledge on category learning. In one study demonstrating the shape bias, Landau, Smith, and Jones (1988) taught young children that some object was called a "dax." When asked to find another "dax," the children tended to choose another object with the same shape even if it had a different size or texture. Likewise, the children tended to reject other objects with different shapes, even if they had the same size and texture as the original "dax." Interestingly, young children seem to limit their use of the shape bias to situations in which new category labels are learned. When the Landau et al. (1988) procedure was repeated except without using the "dax" label, the shape bias was reduced or eliminated. In general, it appears that children are guided by the principle that an object's overall shape is a good predictor of its category label, so children especially pay attention to shapes when learning new labels. However, as the articles by Ward (1993) and Landau (1994) show, the patterns of results for the shape bias, and the underlying general knowledge applied by children in learning category labels, are even more complex and sophisticated than the examples here illustrate.

Knowledge of Category Essences

In addition to general biases such as the taxonomic constraint and the shape bias that would affect children's learning of category labels, it appears that category learning by children and adults is guided by other rich sources of general knowledge. One set of beliefs, referred to as psychological essentialism (Medin & Ortony, 1989), seems to be wide-ranging in its influence. The main idea of psychological essentialism is that (at least for the biological domain) people act as if things in the world have a true underlying nature that imparts category identity. Furthermore, this essence is thought to be the causal mechanism that generates visible properties. Therefore, surface

features provide clues about category membership. This view is known as psychological essentialism because it is concerned with people's assumptions about how the world is, not how the world truly is.

Keil (1989) has provided evidence that children are guided by essentialist assumptions as they learn about members of natural kind categories such as animals and precious metals. In one study, Keil described to children how an animal might undergo some superficial transformations, such as transforming a racoon by painting a white stripe on its back and surgically inserting a sac that contains a smelly substance. The key question was whether this transformed animal was a racoon or a skunk. Children as young as age seven tended to maintain the identity of the animal as a racoon, even though it had been given characteristic features of a skunk. Keil's explanation was that children's biological knowledge led them to discount these superficial features, and instead selectively pay attention to other, deeper anatomical properties. For example, a racoon that resembles a skunk would give birth to other racoons rather than skunks. In related research, Keil described to children artifacts, such as pipes and coffee pots, that underwent transformations. Here it seemed that an object's function was critical to its category membership, again pointing to general beliefs that constrain categorization. (However, for a critique of this line of research, especially with regard to artifact categories, see Malt, 1993).

To summarize, people, even young children, appear to have rather deep pools of knowledge about biological categories as well as artifact categories, that are applied to learning about particular category members (see also Carey, 1985). One fairly general aspect of this knowledge is that certain categories have essences or essential features that are critical for determining category membership. Psychological essentialism has received a great deal of recent attention (also see Gelman, Coley, & Gottfried, 1994; Medin & Heit, in press, for reviews), but other general knowledge about animals, plants, and people also appears to be critical in guiding categorization and category learning. For example, see work by Springer and Belk (1994) on knowledge of contagion in biological categories, work by Coley (1995) on knowledge about biological and psychological properties, and work by Hirschfeld (1995) on knowledge about racial categories.

Implications for Categorization Models

Why Develop Models of Knowledge Effects?

Considering these widespread influences of both specific and general knowledge on category learning, it would be desirable to address and even try to explain these effects in terms of models of categorization. After all, any model of category learning that does not address these influences is not a complete account of category learning (Murphy & Medin, 1985). In research on categorization, there is a tradition of implementing theoretical ideas as computational or mathematical models. This development of models of categorization has had multiple purposes. For one, a categorization model is a precise statement of an account of categorization that facilitates communication among researchers. A model of category learning that addresses these influences of knowledge would be an explicit and testable statement of theory. Furthermore, modeling provides a reasoning tool; it is often difficult for a researcher to know what some theory will predict until the theory is implemented as a model (Hintzman, 1991). Thus, developing a model of some hypothesized categorization process would facilitate its evaluation in terms of how well it accounts for various experimental results. In this way, a model can provide the link between a psychological account of how knowledge influences category learning and the results of experiments such as those reviewed in this chapter.

Despite the promise and appeal of addressing knowledge effects in categorization with computational models, this issue has only recently begun to receive attention. In fact, in 1993, Murphy suggested that most categorization researchers either work on computational models that do not address prior knowledge effects, or they work on issues in categorization that address the richness of people's background knowledge but do not create formal models! Psychological models of categorization have been applied mainly to studies of category learning in isolated contexts (e.g., J. R. Anderson, 1991; Estes, 1986; Gluck & Bower, 1988; Heit, 1992; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1988; Nosofsky, Palmeri, & McKinley, 1994). Typically in these studies, subjects learned isolated categories that were intended by the experimenter to be as unrelated as possible to prior knowledge (e.g., categories of geometric figures or fictional diseases). Of course, categorization researchers have been interested in other important issues in addition to influences of background knowledge, and the strategy of teaching subjects isolated categories would have some value in allowing a researcher to focus on other variables.

Therefore, the task of addressing the widespread influences of knowledge on category learning is a new and important challenge for categorization models.

Exemplar Models

The integration model (Heit, 1994) is an exemplar model of categorization (Medin and Schaffer, 1978) that addresses some effects of prior knowledge. According to exemplar models, a decision whether to categorize some object X as a member of category A depends on the similarity of X to retrieved exemplars for category A. To the extent that X is similar to category A exemplars rather than to exemplars of alternative categories, X will be classified as an A. The novel assumption of the integration model is that two kinds of exemplars influence judgment of whether some stimulus belongs in a category: exemplars of that category as well as prior examples from other related categories. Prior examples are memories from other contexts; in many situations the prior examples would simply be observed members of other categories (Johnson, Hashtroudi, & Lindsay, 1993). For example, imagine that you move to a new city and you are looking for friends to join you in jogging. In effect, you are trying to learn about a new category, of joggers in this city. Say that you have already met a few joggers in the new city, then you meet a new person and you want to predict whether this person is a jogger. To make this evaluation, you would sum up two sources of evidence, the similarity of the new person to prior examples of joggers from other cities and the similarity of the new person to actual joggers you have observed in the new city.

For several experiments simulating this experience of category learning in a new context, Heit (1994, 1995) found that the integration model gave a good qualitative and qualitative account. Figure 1 shows the results of one experiment in which subjects learned about new categories and made judgments about whether some description X belongs in category A. The data points in each graph refer to subjects' average judgments in various conditions. The congruent points refer to test questions that are congruent with prior knowledge, e.g., "How likely is someone with expensive running shoes to be a jogger?" The incongruent points refer to test questions that involve an incongruent pairing, such as "How likely is someone who attends many parties to be shy?" The other variable in the experiment was the proportion of times X actually appeared in category A, e.g., the proportion of people with expensive running shoes who were joggers. The lines in each graph refer to the predictions of the integration model. Note the close correspondence between the data points and the model predictions. As predicted by the integration model, people were influenced by prior knowledge, as indicated by the difference between the congruent and incongruent lines, and

they were influenced by what they actually observed, as indicated by the positive slopes of these lines. Also, these two influences appear to combine independently, as evidenced by the parallel pattern of lines. This independence is consistent with the integration model's assumption that people sum up evidence derived from prior knowledge and evidence derived from actual observations.

Insert Figure 1 About Here

In addition to the integration of prior examples and observed examples, Heit (1994) developed exemplar models of other possible processes by which prior knowledge might affect category learning. First, prior knowledge may lead to selective weighting of category members so observations that fit prior knowledge are remembered best. For example, you might be more successful at learning about joggers who own expensive running shoes than about joggers who do not own expensive running shoes. Second, prior knowledge may have a distortion effect; for example, a jogger without expensive running shoes might be misinterpreted as a jogger with expensive running shoes or even as a non-jogger. Although these additional processes seem plausible, the results of Heit (1994) could be explained without either of them, i.e., by the integration model alone.

Rule-based Models

An alternative scheme for developing models of categorization uses rule-based representations (e.g., Mooney, 1993; Nosofsky et al., 1994; Pazzani, 1991). These models assume that a decision whether some object X belongs to a particular category A depends on whether X satisfies the conditions of a rule defining category A. Using a complex data base of rules (e.g., Mooney, 1993; Pazzani, 1991) and probabilistic responding (e.g., Nosofsky et al., 1994) would allow for a rule-based models to account for a variety of interesting results in categorization. Furthermore, these rule-based models can readily be extended to address prior knowledge effects. Just as the integration model (which is an exemplar model) assumes that prior knowledge takes the form of prior examples, it would be natural for rule-based models to assume that prior knowledge takes the form of pre-existing rules. For example, Mooney's (1993) IOU model of categorization can learn the concept of cup after being presented with just a single example of a cup. This cup might be green, owned by Juliana, lightweight, with a flat bottom, and with a handle. Certain of these features, regarding weight, the cup's bottom, and the cup's handle, are critical to the cup

category. Mooney's model devotes special attention to these features because they are explainable in terms of pre-existing rules concerning liftability, stability, and graspability. This technique, known as explanation-based learning, is a quite powerful way to apply prior knowledge to new category members (see Mooney, 1993, and Wisniewski & Medin, 1994b, for more extensive reviews of explanation-based learning and for further applications to psychological data).

Pazzani (1991) also developed a rule-based model, known as the POST HOC model, that addresses prior knowledge effects. This model, like Mooney's IOU model, begins learning about a new category by accessing rules embodying prior knowledge. These rules may be incorporated into representations of a new category, and in addition, the POST HOC model selectively attends to features that seem especially relevant according to previous knowledge. For example, to account for Pazzani's (1991) experimental results on learning categories of balloons, the POST HOC model would assume that subjects access relevant rules, such as that stretched balloons are more elastic and thus easier to inflate. Then, to learn about the new category members, the model would assume that subjects pay greater attention to goal-relevant features, such as stretching, rather than irrelevant features such as the color of the balloon. This rule-based model successfully predicted Pazzani's results in term of relative difficulty of the various experimental conditions.

Connectionist Models

Finally, it is possible to extend connectionist, or neural network, models of categorization (e.g., Gluck & Bower, 1988; Kruschke, 1992; Shanks, 1991) to address the influences of previous knowledge. (In connectionist models, category learning entails learning a set of associations within a network of nodes. A categorization decision would be performed by assessing which output nodes would be activated after a pattern of inputs is presented to the network.) For example, Choi, McDaniel, & Busemeyer (1993) have explored connectionist models by assuming that at the beginning of learning, certain connections between inputs and outputs have positive or negative strengths. In effect, a connectionist network would have a head start towards learning, as if the network had already been trained on related stimuli. Choi et al. applied this idea to the result that people tend to learning disjunctively-defined concepts more readily than conjunctively-defined concepts (e.g., Salatas & Bourne, 1974). That is, it is generally easier to learn a concept defined in terms of (feature 1 or feature 2 or feature 3 ....) rather than a concept defined in terms of (feature 1 and feature 2 and feature 3 ....). Choi et al. assumed that people begin category learning tasks with initial hypotheses in mind, e.g., to favor disjunctive rules over conjunctive rules. In terms of

connectionist models, these hypotheses could be implemented with negative (or inhibitory) links between nodes corresponding to feature conjunctions and nodes corresponding to category labels. Choi et al. evaluated a few different variants of connectionist models, and were most successful in incorporating prior knowledge into Kruschke's (1992) ALCOVE model, which is a hybrid connectionist-exemplar model.

Also, Kruschke (1993) suggested that his ALCOVE model could account for prior knowledge effects by varying the attentional strengths on different dimensions at the beginning of learning. This suggestion would be an implementation of selective weighting. Note that this proposal differs from the method applied by Choi et al. (1993), which varied the initial connection strengths between nodes in a network rather than varying selective attention. It would be valuable to investigate Kruschke's suggestion further, because one of the strengths of the ALCOVE model is that it can vary attention dynamically over the course of learning. Dynamic attention would correspond to learners having initial hypotheses about which dimensions are relevant to categorization, then adjusting attention as category members are observed (see also Billman & Heit, 1988).

Conclusions from Modeling Efforts

Despite the differences between these exemplar-based, rule-based, and connectionist approaches to modeling the effects of knowledge on concept learning, several themes emerge clearly. Even though the representational details of the models differ, each modeling effort includes two basic kinds of processes. First, in what may be called an integration or anchor-and-adjust process, the model begins with an initial representation for a new category, then revises this representation as additional information is observed. For example, the connectionist model of Choi et al. (1993) begins a learning task with certain network connections already set with negative or positive values. Then these connections are updated during learning. Second, in a selective weighting process, the model is directed to pay attention to certain observations or features of observations that seem especially relevant to the task. For example, Pazzani's (1991) rule-based model allocated more resources to learning about whether or not a balloon was stretched compared to whether the balloon was yellow or purple.

Can these categorization models (Choi et al, 1993; Heit, 1994; Mooney, 1993; Pazzani, 1991) address the other effects of prior knowledge, besides integration and weighting effects? These models can also address knowledge facilitation effects, in which it is easier to learn about a

new category to the extent that it fits with previous beliefs (e.g., Murphy & Allopenna, 1994; Wattenmaker et al., 1986; Wattenmaker, 1995). For example, Murphy and Allopenna found that it was easier for people to learn about new categories of vehicles than to learn categories defined in terms of unrelated or conflicting characteristics (e.g., has thick walls, keeps fish as pets, made in Africa, and has a barbed tail). It makes sense that people learning about new vehicles could use previous knowledge about vehicles as a starting point (an integration process) as well as more easily focus on relevant information (a selective weighting process). In contrast, these processes would not help in learning about nonsensical or completely unfamiliar categories. More generally, integration and selective weighting processes are two possible underlying explanations for why people might show knowledge-related facilitation in learning about categories (for additional possible explanations, see Murphy & Allopenna, 1994).

Therefore, categorization models with these integration and selective weighting processing assumptions can address three of the basic effects of specific knowledge on learning: integration effects, selective weighting effects, and facilitation effects. That is, when the models are provided with suitable information about what specific facts or prior knowledge would influence the learning of a particular new category, the models can reproduce the general patterns of human performance in category learning. This is a significant feat, considering that most formal models of categorization, without assumptions about integration and weighting, do not address the influences of prior knowledge at all. However, so far these models are incomplete in that the relevant prior knowledge must be specified by the modeler. That is, the models address the processes by which prior knowledge and new observations would be combined, but they do not address the processes by which a learner would determine which prior knowledge is relevant. Such processes might be called knowledge selection processes.

For example, Heit (1994) assumed that when subjects learned about joggers in a new city, their prior knowledge consisted of prior examples of joggers from other places. This assumption may be straightforward in the context of a simple laboratory experiment, but knowledge selection processes would necessarily be more complicated in the real world. Imagine that you meet a group of people who are all either British, American, or Belgian, with various occupations and hobbies. What sorts of prior examples or prior knowledge would you use to guide learning about this group? The possibilities seem endless. As another example, imagine that you are learning about a new kind of device that cleans up roadside trash with a suction hose, and you have no previous experience with this sort of device (Wisniewski, 1995). What prior knowledge would be used here? Note that

finding the relevant prior knowledge would be critical for both integration and selective weighting. It appears that assembling the knowledge that is relevant to learning a new concept may require rather sophisticated reasoning processes, in addition to simply retrieving observations from memory. These reasoning processes might include conceptual combination (Hampton, this volume; Murphy, 1993; Rips, 1995) as well as mechanisms for imagining or imaging possible category members (Ward, 1994). A further complexity is that the use of background knowledge and observations might alternate, so that initial beliefs might guide early category learning, which would then lead to the retrieval and perhaps even revision of additional background knowledge. In the terminology of Wisniewski and Medin (1991, 1994b), knowledge and learning would be tightly coupled (see Heit, 1994, for additional evidence). In principle, these additional processes could be implemented in an even more complete model of categorization, but for the most part this work has not yet been performed.

The final effect of specific knowledge described in this chapter is feature interpretation effects, in which the very features that are used to represent category members are themselves learned (e.g., Lesgold et al., 1988; Wisniewski & Medin, 1994b). As pointed out by Murphy (1993) and Wisniewski and Medin (1994b), one current limitation of most current models of categorization is that they operate with a fixed, pre-specified representational system. In principle, however, feature learning might be treated as another form of concept learning. Indeed, developing techniques for learning features has been an active area of research in artificial intelligence research (e.g., Matheus & Rendell, 1989; see Wisniewski & Medin, 1994a for a review). Likewise, researchers who develop connectionist models of learning have been concerned with how a model might form internal representations (e.g., Sejnowski & Rosenberg, 1986) or develop feature detectors (e.g., Rumelhart & Zipser, 1986). Thus, there is good reason to hope that further progress on this issue will be made in the near future.

In contrast to this favorable picture of how current models of categorization can and might address influences of specific knowledge, the day that such models will address effects of more general knowledge seems further off. Consider the sophisticated knowledge representations and processes that must be involved in the taxonomic constraint (Markman, 1989), the shape bias (Landau, 1994), or psychological essentialism (Medin & Ortony, 1989). The knowledge that is relevant to these issues would seem to consist of a richly-connected set of abstract beliefs about categories in general, for example beliefs about relations between the shape of an object and its internal parts. It seems plausible that the simple processes used in explaining effects of specific

knowledge (integration and weighting processes) would have some role in explaining the influences of more general knowledge. For example, the shape bias involves selectively paying attention to the contour of an object. However, such simple processes are only part of the story to be told. It remains an open question how much further development will be required to address the effects of more general knowledge with computational models. An optimistic conjecture might be that categorization models will be able to address influences of general knowledge in the same manner as influences of specific knowledge, once representational issues for describing general and specific knowledge are solved. However, even these representational issues are not easy problems.

To return to the point at the beginning of this chapter, it is clear that knowledge about categories is complex and "deeply intertwingled." It is important to keep in mind that although categorization models can presently explain some of the basic phenomena regarding influences of knowledge on concept learning, this is a complex problem that is not going to be solved entirely anytime soon. Yet, these initial, and certainly incomplete, models of knowledge effects on categorization still serve some of the important purposes of computational modeling. That is, these models are explicit implementations of accounts of how background knowledge shapes category learning, allowing these accounts to be compared and applied to psychological data.

Relations to Inductive Reasoning

Now that the influences of prior knowledge on category learning have been described in some detail, the next two sections will describe research on knowledge effects in two areas of cognitive psychology that are related to category learning: reasoning and memory. After a person has learned about some category, it is natural to ask what this person will do with the category. One important function that categorization serves is to allow inductive inferences or predictions about additional features (J. R. Anderson, 1991; Billman & Heit, 1988; Estes, 1994; Heit, 1992; Ross & Murphy, in press). For example, once you know that someone belongs to the category salesperson, you may predict that this person will try to sell you something.

Inductive reasoning is typically studied in the laboratory by presenting subjects with inductive arguments to be evaluated, such as:

```
Robins are susceptible to a certain disease
-----------------------------------------------------
How likely is it that
ostriches are susceptible to this disease?
```

Research by Rips (1975) and Osherson, Smith, Wilkie, Lopez, and Shafir (1990) has shown that two kinds of information are critical to inductive reasoning. First, inferences will be stronger to the extent that the premise category (e.g., robin) and the conclusion category (e.g., ostrich) are similar. Inferences between similar categories (e.g., robins and sparrows) are stronger than inferences between less similar categories (e.g., robins and ostriches). Secondly, general knowledge about relations to other categories also has influences. One such influence is that inferences will be stronger to the extent that the premise category is typical of its superordinate category (Rips, 1975; Osherson et al., 1990). For example, the knowledge that robins are typical members of the bird category lends strength to inferences from robins to ostriches. On the other hand, if subjects were asked "Given that ostriches are susceptible to a certain disease, how likely is it that robins are susceptible to this disease?", inferences would be relatively weak, because the premise category, ostrich, is not typical of the bird category. (Also see Shipley, 1993, for a further analysis of these phenomena and a discussion of their relation to Goodman's, 1955, work on overhypotheses.)

Another kind of knowledge about categories that affects inductive reasoning is knowledge about variability. Nisbett, Krantz, Jepson, and Kunda (1983) tested subjects on inductive statements of the following form: Given that you observe that one member of category A has property P, what percentage of the members of category A have property P? Nisbett et al. found that the strength of inferences was affected by knowledge of how variable this property would be in the category. For example, given that one member of a certain tribe of people is obese, adults subjects estimated that less than 40% of the members of the tribe are obese. But given that one tribe member has a certain color of skin, subjects concluded that over 90% of the other tribe members would have the same property. Nisbett et al. showed that people make stronger inferences about less variable properties (e.g., skin color) than about more variable properties (e.g., obesity) for a particular category.

Selective weighting effects, due to background knowledge, are also evident in inductive reasoning. Heit and Rubinstein (1994) have found that when people evaluate inductive arguments, they tend to focus on certain features of the categories, depending on what property is being considered in the argument (see also, Medin, Goldstone, & Gentner, 1993). For example, consider the argument:

```
Sparrows travel shorter distances in extreme heat
--------------------------------------------------
How likely is it that
bats travel shorter distances in extreme heat?
```

The behavioral property being considered, traveling shorter distances in extreme heat, would lead subjects to compare sparrows and bats in terms of other behavioral features. Because sparrows and bats are similar in terms of flying, this argument was considered fairly strong. On the other hand, consider the argument:

```
Sparrows have livers with two chambers
--------------------------------------------------
How likely is it that
bats have livers with two chambers?
```

Here, the anatomical property being considered, having a two-chambered liver, would lead subjects to focus on other anatomical properties. Because of the anatomical dissimilarities between sparrows and bats (e.g., one is a bird and one is a mammal), this argument was considered relatively weak. In addition to these results from Heit and Rubinstein, evidence for selective weighting effects in inductive reasoning has been provided by Coley (in press), Gelman and Markman (1986), and Springer (1992). For additional evidence of the influences of knowledge about properties on induction, see Sloman (1994).

Models of Inductive Reasoning

The category-based induction (CBI) model (Osherson et al., 1990; Osherson, Stern, Wilkie, Stob, & Smith, 1991) is a computational model of induction that addresses some of the

influences of categorical knowledge. This model may be applied to complex inductive arguments with multiple premises, such as:

```
Category A1 has property P
Category A2 has property P
Category A3 has property P
-------------------------------------
```

How likely is it that

```
Category B has property P?
```

According to the CBI model, two factors influence how people evaluate the inductive soundness of such inferences. First, inferences will be stronger to the extent that the premise categories (A1, A2, ...) are similar to the conclusion category (B). The second factor in the CBI model is the coverage of the premise, that is the similarity between the category or categories in the premise and members of the superordinate category that encompasses the categories in the premise and conclusion. A few examples should make the idea of coverage clear. Consider again an inductive inference from robin to ostrich. The most specific superordinate category that includes robins and ostriches is bird. Now, robin is fairly similar to other members of the category bird. Thus, if robins have some property P, it is plausible that all birds, including ostriches, have property P. In the CBI model, the two sources of evaluating inferences, similarity and coverage, are just added together. Category members that are atypical do not contribute much to coverage, for example, ostrich as a premise category would provide little coverage for the superordinate category bird. The CBI model also provides an elegant way to evaluate the coverage of arguments with multiple premises. For example, given the premises that both robins and penguins have property P, it seems likely that all birds have property P, because robins and penguins are quite diverse members of the superordinate, birds. On the other hand, the premises that robins and sparrows have some property does not lend as much support to the belief that all birds have the property, because robins and sparrows do not cover the superordinate category birds much better than just robins alone.

The CBI model provides a successful account of several influences of categorical knowledge on inductive reasoning, especially how knowledge about superordinate categories affects reasoning (see Osherson et al., 1990, 1991, for reviews). However, the CBI model does not address the other knowledge effects described here, such as selective weighting effects (e.g., Gelman & Markman, 1986; Heit & Rubinstein, 1994) or effects of knowledge about variability (Nisbett et al., 1993). In principle, it would be possible to add a selective weighting component to

the CBI model, just as it is possible to add selective weighting to categorization models (e.g., Pazzani, 1991). That is, it would be possible to have the CBI model focus on different category features depending on which property is being inferred, so that it could begin to address the results indicating selective weighting. However, it might well take a complex reasoning process to figure out which features are relevant to inferring various properties, e.g., which features are relevant to inferring whether an animal travels shorter distances in extreme heat. As mentioned earlier, a challenge for computational models of categorization is to determine which prior knowledge is relevant to a particular situation. Likewise, future computational models of induction will be faced with the challenge of assembling the prior knowledge that is relevant to guiding an inference.

Relations to Memory

There is a strong affinity between research on categorization and research on memory, because categorization and memory are highly interdependent (or intertwingled) facets of cognition. Two parallels between categorization research and memory research will be drawn here. First, studies of the influences of prior knowledge on category learning are closely related to research on the impact of schemas and stereotypes on memory. Second, there are close connections between categorization models and memory models, suggesting that the task of developing categorization models that address knowledge effects is part of a larger enterprise in cognitive modeling.

Influences of Knowledge on Memory

Research on memory has largely followed two traditions. In the tradition of Ebbinghaus (1885/1964), researchers have focused on precise quantitative relations among various factors that affect memory and various memory tasks (e.g., the effect of amount of study on free recall performance, Underwood, 1970). This research tradition has typically used simple verbal stimuli (e.g., nonsense syllables or concrete nouns) with the intent of isolating certain aspects of memory and minimizing the influences of the subject's prior knowledge. Second, in the tradition of Bartlett (1932), researchers have focused on the richness of human knowledge and the interesting influences of knowledge on new learning (see Johnson & Sherman, 1990, for a review). (Note the similarity to the description of two traditions of research in categorization by Murphy, 1993.) To some extent, there may be a trade-off between working in the first tradition and working in the second

tradition, but there is plenty of research that draws from both (e.g., Collins & Quillian, 1969; Graesser, 1981; Smith & Zarate, 1992).

As an illustration of work in the second tradition, consider the classic example from Carmichael, Hogan, & Walter (1932) in Figure 2. When subjects were shown the drawing in Figure 2a, their memories of this picture were influenced by their background knowledge. If the picture was originally labeled as eyeglasses, then subjects tended to recall something like Figure 2b: Their knowledge of eyeglasses influenced their specific memories of the picture. If the picture was originally labeled as a barbell, then subjects tended to recall something like Figure 2c. Note that this result is quite like the feature interpretation phenomena for category learning described by Wisniewski and Medin (1991, 1994b), in terms of ambiguous figures being influenced by labeling. Another classic example of the influence of schemas, or general knowledge structures, on memory was provided by Bransford and Johnson (1972). In this study, subjects read a rather abstract paragraph concerning a procedure for arranging items into different groups, going to the proper facilities, etc. Their later recall memory for this passage was poor, unless they had also been told that the passage describes washing clothes. In other words, the subjects' general knowledge about doing laundry facilitated memory for this text. Note the resemblance between this result and the knowledge facilitation results in category learning (e.g., Murphy & Allopenna, 1994; Wattenmaker et al., 1986; Wattenmaker, 1995).

Insert Figure 2 About Here

Researchers in social psychology have also been concerned with influences of knowledge on learning, in particular the influences of social stereotypes on what is remembered about individual persons. For example, in a study of the effects of developing gender stereotypes on memory, Stangor and Ruble (1989) showed children television commercials that were either congruent with their stereotypes (e.g., girls playing with toy dolls) or incongruent with their stereotypes (e.g., girls playing with toy trucks). Stangor and Ruble found that the congruent commercials were recalled better. More generally, it appears that what we remember about the persons we meet depends on much more than just our direct observations of these persons; the influences of social group stereotypes are widespread (see Srull & Wyer, 1989, and Stangor & McMillan, 1992, for reviews).

Given these similarities between memory phenomena and categorization phenomena, future research on the influences of knowledge on concept learning may be well-informed by considering the related work in memory. For example, the processes proposed in this chapter as influencing category learning have been discussed extensively by theorists in the area of memory as well. Selective weighting influences of knowledge on memory have been emphasized by Alba and Hasher (1983), who discussed how schematic knowledge would operate as a filter either at encoding or retrieval. Similarly, Smith and Zarate (1992) have discussed how a person's goals, recent experiences, and immediate environment would affect selective attention to different social dimensions such as gender, age, ethnicity, or race. In addition to the classic work by Asch (1946) on processes of interpretation and distortion, Taylor and Crocker (1978) have discussed how general knowledge may be used to fill in missing featural information. Finally, integration processes in person memory have been proposed by N. H. Anderson (1991) and Brewer and Nakamura (1984). Work on these topics by memory researchers can certainly guide research on the corresponding issues in categorization. Likewise, categorization research can influence work on memory and social cognition. For example, Rothbart and Taylor (1992) discuss how conceptual knowledge about psychological essentialism and mutual exclusivity might apply to stereotypes and social categories.

Memory Models and Categorization Models

Models of categorization, ideally, will not be isolated accounts of a particular task or experiment but instead will dovetail with other theoretical accounts of cognitive activities such as memory and reasoning. One example of the potential synergy between categorization models and computational models of memory is the compatibility between exemplar models of categorization and multiple-trace models of memory (Gillund and Shiffrin, 1984; Hintzman, 1986, 1988). Multiple-trace models assume that a memory judgment, such as a recognition decision, depends on evaluating the total similarity of a test item to memory traces of particular stimuli (see Jones & Heit, 1993 for a review). Likewise, exemplar models assume that a decision whether to place a test item in one category or another depends on evaluating the similarity of the test item to memory traces for members of each category. Heit (1993) applied the exemplar models of categorization in Heit (1994) to a set of experiments on stereotype effects on recognition memory (Stangor & McMillan, 1992). The simulations in Heit (1993, 1994) provided converging evidence that integration processes can explain a variety of results concerning the influences of prior knowledge on memory

as well as categorization. (For a related example of applying exemplar models to categorization and memory tasks, see Smith and Zarate, 1992.) Note that such a synergy between categorization models and memory models need not be limited to the common framework of exemplar models and multiple-trace models. For example, connectionist modeling provides another framework for developing general models of categorization and memory.

Conclusion

In sum, the interrelated cognitive abilities of category learning, inductive reasoning, and memory are significantly guided by people's background knowledge, including both specific knowledge and more general principles. To an encouraging extent, these influences can be captured by computational models. Yet at the same time these modeling efforts highlight their own incompleteness, in terms of what needs to be explained even further.

The variety of influences of knowledge reviewed in this chapter are if anything an underestimate of the intertwingled nature of knowledge and concepts. Theoretical accounts of categorization, whether or not they are in the form of computational models, face a significant challenge in accounting for these influences. Although it has been traditional (e.g., Smith & Medin, 1981) to describe accounts of categorization in terms of pure representational formats (exemplar models, prototype models, rule-based models, etc.), it appears that more complex conceptions of representation may be required. These basic forms of representation may well serve as a starting point for future work. In the future, it seems likely that an important question in categorization research will be what sort of complex, multimodal representational scheme can be used to describe the rich body of conceptual knowledge that is critical to learning. Such a scheme would need to account for knowledge of relations among categories, and knowledge at multiple levels of abstraction. People's knowledge about categories might well include many forms of information such as exemplars, images, and rules and other abstractions (Barsalou, 1993; Graesser, Langston, & Baggett, 1993; Malt, 1993). The problem of developing more sophisticated forms of conceptual representation may eventually overshadow comparisons between pure forms of representations, such as experiments intended to address whether exemplar models are better than prototype models.

Although theorists such as Anderson (1978) and Barsalou (1990) have noted that models of cognition must address both representation and processing, in categorization research representational issues have perhaps received more emphasis than processing issues (e.g., see

reviews by Komatsu, 1992; Medin & Heit, in press; Smith & Medin, 1981). In addressing the topic of how previous knowledge guides the learning of new concepts, as well as performing category-based inductive inferences, processing issues are fundamental. The critical questions concern what are the processes by which people assemble relevant knowledge, form the initial representations for new categories, selectively attend to important information, and interpret the category members they observe in light of prior knowledge. It is notable that categorization models with three different representational frameworks (exemplar, rule-based, and connectionist models) are each able to make progress towards addressing knowledge effects by adopting similar sets of processing assumptions. Indeed, these processing assumptions appear more important for fitting various experimental results than the particular representational assumptions of each model.

Another way of going beyond issues of representation to distill highly general principles is to consider various cognitive activities at the computational level (Marr, 1982), that is, to consider what computational problems are being solved and at an abstract level how they are being solved. One framework for describing computational-level problems and solutions is provided by Bayesian statistical theory (Edwards et al, 1963; Raiffa & Schlaifer, 1961). It is assumed in Bayesian theory that to learn about some new part of the environment (e.g., a novel category or novel property), one begins with an initial estimate based on previous knowledge, then revises this information as new information is encountered. At a very general level, this description can be applied to influences of prior knowledge on concept learning, induction, and memory. We seem to assume initially that new categories will be like old categories, novel properties in inductive arguments will be like familiar properties, and new experiences to be stored in memory will resemble our previous memories. Perhaps future accounts of categorization, inductive reasoning, and memory will receive further guidance from Bayesian statistics (for some examples of Bayesian accounts of cognitive activities, see J. R. Anderson, 1990, 1991; Oaksford & Chater, 1994).

Although many other issues in categorization will, of course, continue to be important, it is easy to be optimistic about future research on knowledge and concept learning. There have been a large number of recent empirical discoveries and the development of formal models is also beginning to take off. In addition, the importance of prior knowledge in related areas of cognitive psychology, reasoning and memory, is suggestive of the centrality of this issue. It is not possible to know where this line of categorization research will lead, but it appears that it is heading in a promising direction.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? Psychological Bulletin, 93, 203-231.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. Psychological Review, 85, 249-277.

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

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

Anderson, N. H. (1991). Stereotype theory. In N. H. Anderson (Ed.), Contributions to information integration theory, Volume II: Social. (pp. 183-240). Hillsdale, NJ: Erlbaum.

Asch, S. E. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology, 41, 258-290.

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in memory representation. In T. K. Srull & R. S. Wyer (Eds.), Advances in social cognition (pp. 61-88). Hillsdale, NJ: Erlbaum.

Barsalou, L. W. (1993). Structure, flexibility, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual systems. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), Theories of memory. London: Erlbaum.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge, England: Cambridge University Press.

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

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11, 717-726.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer & T. K. Srull (Eds.), Handbook of social cognition (pp. 119-160). Hillsdale, NJ: Erlbaum.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), Linguistic theory and psychological reality Cambridge, MA: MIT Press.

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

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. Journal of Experimental Psychology, 15, 73-86.

Choi, S., McDaniel, M. A., & Busemeyer, J. R. (1993). Incorporating prior biases in network models of conceptual rule learning. Memory & Cognition, 21, 413-423.

Clark, E. V. (1983). Meanings and concepts. In P. H. Mussen (Ed.), Handbook of Child Psychology New York: Wiley.

Coley, J. D. (1995). Emerging differentiation of folkbiology and folkpsychology. Child Development, 66, 1856-1874.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Ebbinghaus, H. (1885/1964). A contribution to experimental psychology. New York: Dover.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. Psychological Review, 70, 193-242.

Estes, W. K. (1986). Array models for category learning. Cognitive Psychology, 18, 500-549.

Estes, W. K. (1994). Classification and cognition. New York: Oxford University Press.

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

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

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. Journal of Experimental Psychology: General, 117, 227-247.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General, 123, 178-200.

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

Graesser, A. C. (1981). Prose comprehension beyond the word. New York: Springer-Verlag.

Graesser, A. C., Langston, M. C., & Baggett, W. B. (1993). Exploring information about concepts by asking questions. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), The Psychology of Learning and Motivation: Categorization by humans and machines (pp. 411-436). San Diego: Academic Press.

Hampton, J. A. (1993). Prototype models of concept representation. In Van Mechlen, I., Hampton, J., Michalski, R., & Theuns, P. (Eds.), Categories and Concepts: Theoretical Views and Inductive Data Analysis (pp. 67-88). San Diego: Academic Press.

Hayes, B. K., & Taplin, J. E. (1992). Developmental changes in categorization processes: Knowledge and similarity-based models of categorization. Journal of Experimental Child Psychology, 54, 188-212.

Hayes, B. K., & Taplin, J. E. (1995). Similarity-based and knowledge-based process in category learning. European Journal of Cognitive Psychology, 7, 383-410.

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

Heit, E. (1993). Modeling the effects of expectations on recognition memory. Psychological Science, 4, 244-252.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 1264-1282.

Heit, E. (1995). Belief revision in models of category learning. In Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, Pittsburgh. 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.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review, 95, 528-551.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock (pp. 39-56). Hillsdale, NJ: Erlbaum.

Hirschfeld, L. A. (1995). Do children have a theory of race? Cognition, 54, 209-252.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. Psychological Bulletin, 114, 3-28.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins & R. M. Sorrentino (Eds.), Handbook of motivation and social cognition: Foundations of social behavior (pp. 482-526). New York: Guilford Press.

Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 799-812.

Keil, F. C. (1989). Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.

Keleman, D., & Bloom, P. (1994). Domain-specific knowledge in simple categorization tasks. Psychonomic Bulletin & Review, 1, 390-395.

Komatsu, L. K. (1992). Recent views of conceptual structure. Psychological Bulletin, 112, 500-526.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. Psychological Review, 99, 22-44.

Kruschke, J. K. (1993). Three principles for models of category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), The Psychology of Learning and Motivation: Categorization by humans and machines (pp. 283-326). San Diego: Academic Press.

Landau, B. (1994). Object shape, object name, and object kind: Representation and development. In D. L. Medin (Ed.), The Psychology of Learning and Motivation (Vol. 31, pp. 253-304). San Diego: Academic Press.

Landau, B., Smith, L., & Jones, S. (1988). The importance of shape in early lexical learning. Cognitive Development, 3, 299-321.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), The Nature of Expertise (pp. 311-342). Hillsdale, NJ: Erlbaum.

Malt, B. C. (1993). Concept structure and category boundaries. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), The Psychology of Learning and Motivation: Categorization by humans and machines (pp. 363-390). San Diego: Academic Press.

Markman, E. M. (1989). Categorization and naming in children. Cambridge, MA: MIT Press.

Markman, E. M. (1990). Constraints children place on word meanings. Cognitive Science, 14, 57-77.

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

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. Psychological Review, 100, 254-278.

Medin, D. L., & Heit, E. (in press). Categorization. In D. E. Rumelhart & B. O. Martin (Eds.), Handbook of Cognition and Perception. San Diego: Academic Press.

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

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. Psychological Review, 85, 207-238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. Journal of Experimental Psychology: Human Learning and Memory, 7, 355-368.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. Cognitive Psychology, 19, 242-279.

Mooney, R. J. (1993). Integrating theory and data in category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), The Psychology of Learning and Motivation: Categorization by humans and machines (Vol. 29). San Diego: Academic Press, Inc.

Mumma, G. H. (1993). Categorization and rule induction in clinical diagnosis and assessment. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), The Psychology of Learning and Motivation: Categorization by humans and machines (pp. 283-326). San Diego: Academic Press.

Murphy, G. L. (1993). Theories and concept formation. In I. V. Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), Categories and concepts: Theoretical views and inductive data analysis (pp. 173-200). London: Academic Press.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 904-919.

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

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), Advances in Cognitive Science (pp. 23-45). Chichester: Ellis Horwood.

Nakamura, G. V. (1985). Knowledge-based classification of ill-defined categories. Memory & Cognition, 13, 377-384.

Nelson, T. H. (1987). Computer lib; Dream machines. Redmond, Washington: Microsoft Press.

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

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 700-708.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. Psychological Review, 101, 53-79.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. Psychological Review, 101, 608-631.

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

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

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 416-432.

Peirce, C. S. (1931-1935). Collected papers of Charles Sanders Peirce. Cambridge: Harvard University Press.

Quine, W. V. O. (1960). Word and object. Cambridge, MA: MIT Press.

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

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

Rips, L. (1995). The current status of research on conceptual combination. Mind and Language, 10, 72-104.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 21-59). New York: Cambridge University Press.

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. (in press). Category-based predictions: The influence of uncertainty and feature associations. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Rothbart, M. & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. Semin & K. Fielder (Eds.), Language, interaction and social cognition (pp. 11-36). Sage Publications.

Rumelhart, D. E. & Zipser, D. (1985). Feature discovery by competitive learning. Cognitive Science, 19, 75-112.

Salatas, H., & Bourne, L. E. (1974). Learning conceptual rules: III. Processes contributing to rule difficulty. Memory & Cognition, 2, 549-553.

Sejnowski, T. J., & Rosenberg, C. R. (1986). NETtalk: A parallel network that learns to read aloud (Technical Report No. JHU/EECS-86/01). Johns Hopkins University, Department of Electrical Engineering and Computer Science.

Shanks, D. R. (1991). Categorization by a connectionist network. Journal of Experimental Psychology: Learning, Memory, & Cognition, 17, 433-443.

Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.), The Psychology of Learning and Motivation (Vol. 30, pp. 265-301). Orlando, FL: Academic Press.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), The Psychology of Learning and Motivation (Vol. 31, pp. 305-349). San Diego: Academic Press.

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

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge: Harvard University Press.

Smith, E. R., & Zaraté, M. A. (1992). Exemplar-based models of social judgment. Psychological Review, 99, 3-21.

Springer, K. (1992). Children's awareness of the biological implications of kinship. Child Development, 63, 950-959.

Springer, K., & Belk, A. (1994). The role of physical context and association in early contamination sensitivity. Developmental Psychology, 30, 864-868.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. Psychological Review, 96, 58-83.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. Psychological Bulletin, 111, 42-61.

Stangor, C., & Ruble, D. N. (1989). Differential influences of gender schemata and gender constancy on children's information processing and behavior. Social Cognition, 7, 353-372.

Taylor, S. E., & Crocker, J. (1978). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), Social cognition: The Ontario symposium (pp. 89-134). Hillsdale, NJ: Erlbaum.

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

Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. Journal of Verbal Learning and Verbal Behavior, 9, 573-580.

Ward, T. B. (1993). Processing biases, knowledge, and context in category formation. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), The Psychology of Learning and Motivation: Categorization by humans and machines (pp. 257-282). San Diego: Academic Press.

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. Cognitive Psychology, 27, 1-40.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. Cognitive Psychology, 28, 274-328.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. Cognitive Psychology, 18, 158-194.

Waxman, S. R., Shipley, E. F., & Shepperson, B. (1991). Establishing new subcategories: The role of category labels. Child Development, 62, 127-138.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 449-468.

Wisniewski, E. J., & Medin, D. L. (1991). Harpoons and long sticks: The interaction of theory and similarity in rule induction. In D. H. Fisher, M. J. Pazzani, & P. Langley (Eds.), Concept

formation: Knowledge and experience in unsupervised learning. San Mateo, CA: Morgan Kaufmann.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. Cognitive Science, 18, 221-282.

Wisniewski, E. J., & Medin, D. L. (1994). The fiction and nonfiction of features. In R. S. Michalski & G. Tecuci (Eds.), Machine Learning. San Mateo, CA: Morgan Kaufmann.

Acknowledgments

Please address correspondence to Evan Heit, Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom. I am grateful to John Coley, Koen Lamberts, Douglas Medin, Gregory Murphy, and David Shanks for comments on this chapter.

Figure Captions

Figure 1. Results of Heit (1994), Experiment 2, indicated as data points, and predictions of the integration model, indicated as lines. Reprinted by permission.

Figure 2. Illustration of schematic effects on memory, adapted from Carmichael, Hogan, and Walter (1932).

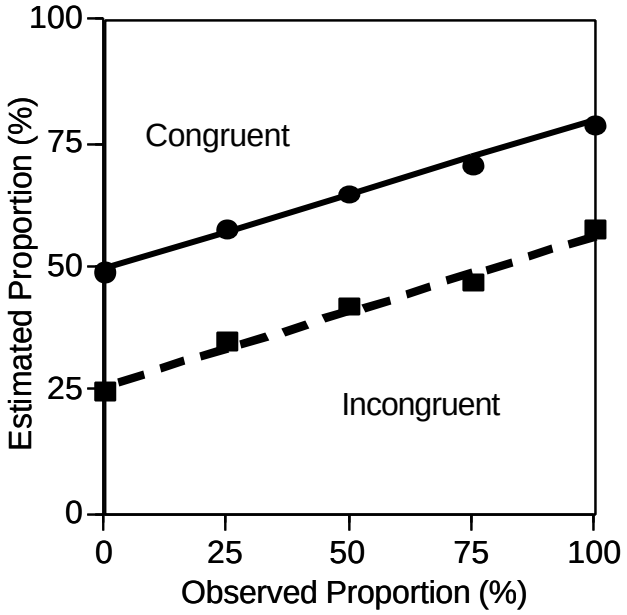
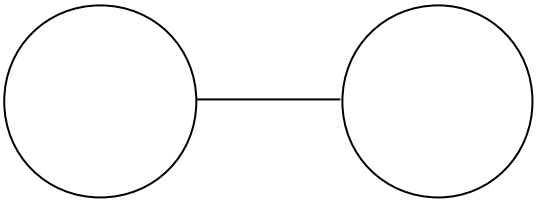
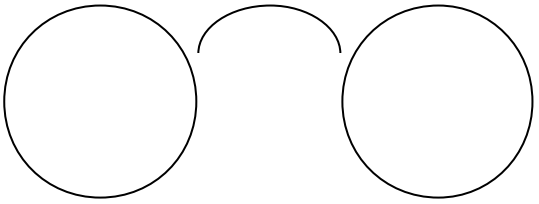

100
75
50
25
0
Estimated Proportion (%)
Congruent
Incongruent
0
25
50
75
100
Observed Proportion (%)

Figure 1.

A.



B.



C.

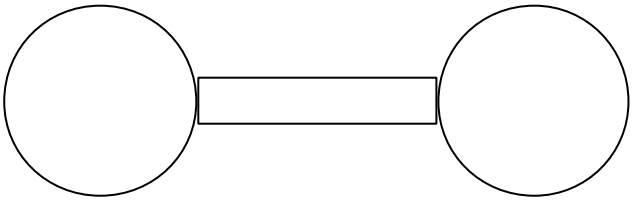


Figure 2.