Chapter 23

Models of Categorization

Andy J. Wills

Abstract

This chapter reviews some of the main ways in which theories of categorization have been expressed in formal, mathematical terms. The focus of the models discussed is the categorization of abstract visual forms by adults in situations where prior knowledge is unlikely to contribute much to performance. Each of the main components of categorization models is discussed: input representations, attentional processes, intermediate representations (e.g., prototypes, exemplars), evidential mechanisms (e.g., similarity, rules), and decision mechanisms (e.g., the choice axiom; Luce, 1959). Models discussed include the Generalized Context Model (Nosofsky, 1986), ALCOVE (Kruschke, 1992), prototype models (e.g., Smith & Minda, 1998), clustering models (e.g., SUSTAIN; Love, Medin & Gureckis, 2004), and multiprocess models (e.g., COVIS; Ashby, Alfonso-Reese, Turken, & Waldron, 1998).

Key Words: categorization, models, formal, mathematical, connectionist, exemplar, prototype, GCM, ALCOVE, COVIS, SUSTAIN

Previous chapters in this volume outlined some of the phenomena and descriptive theories in the study of mental categories. The purpose of the current chapter is to describe some of the main ways in which theories of categorization have been expressed in formal, mathematical terms. Formal description of theories is important to the

development of cognitive psychology as a science—it encourages theorists to be explicit in their assumptions and to describe their theories in a way that permits independent objective verification of the theory's predictions. Formal description also, in principle, permits the possibility of unambiguous rejection of a theory.

The theories described in this chapter focus on adults categorizing abstract visual forms in situations where prior knowledge of real-world categories is unlikely to contribute much to performance. The level of experimental control this permits is often assumed to assist the development of formal theory. Of course, the study of categorization is much broader than the study of the classification of abstract forms, and this is reflected in the other chapters of Part IV.

Another way in which this chapter is narrower than the field it describes is that only models that might be loosely described as process models are discussed. Process models, at varying levels of abstraction, attempt to characterize the representations and information processing assumed to underlie categorization behavior. This tends to be done without much consideration of what adaptive purpose categorization might serve. A complementary approach is functional modeling, which considers the purposes categorization might serve, and then seeks to describe ways in which an optimal system (i.e., one with infinite time and resources) might best serve those purposes (Anderson, 1991; Pothos & Chater, 2002).

In this chapter, I discuss the components of formal process models of categorization. I do this in the order that information is assumed to pass through these components (at least in the first instance—models differ in the extent to which they assume information flows in both directions). This flow of information is represented

schematically in Figure 23.1. Categorization is seldom modeled from a retinal starting point—most modelers assume some form of higher level input representation of the presented stimulus. The information from this input representation is sometimes modulated by attentional mechanisms, usually with attention directed to maximize categorization accuracy. The attentionally modulated information from the input representations is sometimes assumed to activate one or more intermediate representations, which are defined in the coordinates of the input representation system. Among the types of intermediate representations assumed are exemplars, prototypes, clusters, and distributed representations.

<insert Figure 23.1 here>

Information from the intermediate representations is assumed to activate one or more category representations. The process by which a category representation or representations are activated is described, for the purposes of this chapter, as an evidential mechanism. Examples of evidential mechanisms include associative links, decision bounds, and rules.

Sometimes an evidential mechanism will activate more than one category representation. In the laboratory, and in everyday life, there is often a requirement to produce a categorical response (i.e., "it's a dog," rather than "it's quite similar to a dog, a bit similar to a cat, and not very similar at all to a bagel"). Therefore, there is a need for a decision mechanism that is able to turn graded information into a categorical response. This categorical decision ultimately results in an observable action, although the stages of information processing beyond the categorical decision are seldom considered by models of categorization.

This introductory section comprised a brief overview of the representations and processes generally assumed in models of categorization. In the sections that follow, some of the main approaches to modeling each of these components are discussed.

Input Representations

All formal models inevitably make assumptions about the nature of the information that is available to them at input. In the case of models of categorization, those assumptions generally take one of two forms: geometric models and featural models.

Featural (also known as elemental or microfeatural) models assume that any presented stimulus, even an apparently simple one such as a monochromatic light, is represented by a number of features. Two stimuli are similar to the extent that they have common features and the extent to which they do not have distinctive features. Some of the assumptions often found in a featural approach are that any given stimulus activates a relatively small subset of the features within the representational system (sparse coding; e.g., Granger, Ambros-Ingerson, & Lynch, 1989), that the subset of features activated by a given stimulus varies somewhat from one presentation to the next (stimulus sampling; Estes, 1950), and that features can have graded levels of activity (rather than simply being either “on” or “off”). Featural accounts have a long history (e.g., Estes, 1950), and they are also at the heart of some recent (Harris, 2006; McLaren & Mackintosh, 2000, 2002) and some very famous (McClelland & Rumelhart, 1985; Tversky, 1977) formal models.

Geometric models have a similarly impressive pedigree (Ashby & Gott, 1988; Nosofsky, 1986; Shepard, 1958) and, in recent times, have been more common in the modeling of adult categorization data than have featural models. This is in large part due

to the success of two geometric models: the Generalized Context Model (Nosofsky, 1986) and General Recognition Theory (Ashby & Gott, 1988). Geometric models assume that any presented stimulus can be represented as a point (or a distribution; Ashby & Gott, 1988) in a psychological similarity space. Two stimuli are similar to the extent that they are close to each other in this space. Figure 23.2A illustrates this. The Generalized Context Model (GCM) assumes that similarity is an exponential decay function of distance (see Fig. 23.2B; Shepard, 1958). Sometimes, a Gaussian function is used instead (Nosofsky, 1991; this approximates trial-to-trial variability in the perception of highly confusable stimuli in models that represent individual stimuli as points in space, rather than as distributions; Ennis, 1988).

<Figure 23.2 here>

Although in common usage "distance" implies Euclidean distance (Fig. 23.2C), other interpretations are possible, for example, "city-block" distance (Fig. 23.2D). The GCM typically employs Euclidean distance where stimulus dimensions are integral (e.g., hard to selectively attend, such as hue and saturation; Garner, 1978) and city-block distance where stimuli are separable (the antonym of integral).

There are well-known statistical methods for deriving a geometric representation of a stimulus set from data such as similarity ratings, or the extent to which two stimuli are confused in an identification task. The statistical method, known as multidimensional scaling (e.g., Kruskal, 1964), is akin to deriving the relative position of towns from the distances between them. To the extent that multidimensional scaling produces a good approximation to the psychological similarity data in a low-dimensional space (and it often does; Shepard, 1987), geometric models can provide an elegant and readily comprehensible representation of the information available to the categorization process.

The representational power of both geometric and featural systems is greater than might initially be apparent. For example, it is possible to approximate continuous dimensions with an arbitrary degree of precision using a featural representation (Restle, 1959; Shanks & Gluck, 1994). It is also possible to represent asymmetrical similarity (an ellipse is more similar to a circle than a circle is to an ellipse) in a geometric model, despite the fact that it is self-evidently true that the distance from A to B must be equal to the distance from B to A. Geometric models can account for asymmetric similarity by assuming stimulus-specific biases (a circle is more easily brought to mind than an ellipse). In fact, a geometry-plus-bias model is mathematically equivalent to certain featural models (Holman, 1979).

Mechanisms of Attention

Some models of categorization assume that the information provided by the input representations can be modulated by attentional processes, and that the function of this attentional modulation is to increase categorization accuracy. Attentional modulation is sometimes assumed to operate at the level of the dimensions of a geometric input representation (Nosofsky, 1986; see also Sutherland & Mackintosh, 1971) and sometimes assumed to operate in a more stimulus-specific or feature-specific manner (Kruschke, 2001; Mackintosh, 1975). At the level of dimensions, attentional modulation can be conceptualized as the stretching and compressing of a geometric input representation along one or more of its dimensions. Figures 23.3A and 23.3B illustrate this form of attentional modulation; note that the effect of the modulation is that the within-category similarities are increased and the between-category similarities are decreased. This will make it easier for the model to correctly categorize the presented stimuli, and it is this

kind of process that underlies the success of certain models in capturing the relative difficulty people have in acquiring different category structures (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994). For example, the category structure in Figure 23.3C is harder for people to learn than the structure in Figure 23.3A (Kruschke, 1993), and one appealing explanation for this is that selective attention to one dimension facilitates the learning of the latter, but not the former, problem.

<insert Figure 23.3 here>

Similar processes of attentional modulation can also be applied to models with featural input representations. In these kinds of models (e.g., Mackintosh, 1975; Kruschke, 2001), attention to a feature is assumed to increase to the extent that it is a better predictor of the category label than other features that are simultaneously present. Symmetrically, attention to a feature is assumed to decrease if it is a worse predictor of the category label than other present features. The consequence of this attentional allocation will be to increase categorization accuracy, and the existence of such a process in humans is supported by behavioral (Le Pelley & McLaren, 2003; Lochmann & Wills, 2003), eye-tracking (Kruschke, Kappenman, & Hetrick, 2005), and electrophysiological (Wills, Lavric, Croft, & Hodgson, 2007) data. Eye-tracking data are also consistent with dimensional allocation of attention (Rehder & Hoffman, 2005).

In addition to these processes of selective attention, some models assume that overall differentiation of the input representation is possible—typically as a result of extended exposure to the stimuli. Within geometric input representations, this can be conceptualized as an overall expansion of psychological similarity space (Fig. 23.3D; Nosofsky, 1986). With featural input representations, differentiation can be represented by a reduction in the activity levels of features that stimuli have in common (McLaren &

Mackintosh, 2000, 2002). One notable aspect of a featural account of stimulus differentiation is that it predicts when exposure to stimuli will aid the subsequent categorization of those stimuli, and when exposure will impair subsequent classification (Wills & McLaren, 1998: Wills, Suret, & McLaren, 2004). Another phenomenon that is naturally conceptualized within a featural representation is that of unitization. There is evidence that, with extended experience, the components of multiattribute stimuli become "psychologically fused" into a more unitary representation (Goldstone, 2000). The formation of associations between featural input representations is one way to conceptualize the process of unitization (McLaren & Mackintosh, 2000, 2002).

Intermediate Representations

Some models of categorization assume the presence of intermediate representations that mediate between the (sometimes attentionally modulated) input representation and the evidential process. Intermediate representations are generally expressed in the same mode of representations as the input representation. In other words, if a geometric input representation is assumed, then the intermediate representations are also expressed in that geometric space. If the input representation is featural, then so is the intermediate representation. For illustrative convenience, this section will describe most intermediate representations geometrically, but they can also be expressed in featural terms.

The assumed nature of the intermediate representations varies greatly between different accounts of categorization. In this section, I will outline some of the types of intermediate representation that have been assumed.

Cluster Representations

Cluster representations (e.g., Anderson, 1991; Love et al., 2004; Vanpaemel & Storms, 2008) are activated by a region of the input representation. Clusters are usually constrained to represent coherent regions of the input representation, and they are generally assumed to be some form of average of the stimuli that activate the cluster representation. For example, Figure 23.4 shows one possible way in which the 16 stimuli shown might result in four cluster representations (marked "C"). Note that, as shown in this example, there is no necessity for a cluster representation to correspond exactly with any of the stimuli the participant has experienced.

<insert Figure 23.4 here>

Prototype representations (e.g., Reed, 1972; Smith & Minda, 1998) and exemplar representations (Medin & Schaffer, 1978; Generalized Context Model, Nosofsky, 1986) are special types of cluster representation. In prototype-based representation, each category label is assumed to result in exactly one cluster representation (marked "P" in Fig. 23.4). In exemplar-based representation each experimenter-defined stimulus is assumed to correspond to exactly one cluster (the 16 circles in Fig. 23.4). Some of the better-known debates in categorization research have centered on the question of whether exemplar-based or prototype-based representation is the better basis for models of categorization. One possible answer (Love et al., 2004) is that cluster size is a function of the stimuli participants are presented with and the level of experience participants have with those stimuli (Smith & Minda, 1998).

Other evidence that clusters less specific than exemplars are formed includes the effect of partially reversing a category structure. For example, train participants on 16 A stimuli and 16 B stimuli. Now train a partial reversal of those category assignments—for example, 6 As are now labeled B, and 6 Bs are now labeled A—and do not present the

remaining 20 stimuli during this partial reversal. Under some circumstances (e.g., Wills, Noury, Moberly, & Newport, 2006) participants seem to assume that these remaining stimuli have also reversed their category membership. One straightforward explanation of this result is that the partial reversal leads participants to reverse the mapping between representations of the categories and representations of their labels. For this explanation to work, the representations have to be less specific than exemplar representations. One possible solution to this problem for exemplar-based models is to assume another, categorical, representation layer in addition to an exemplar representation layer (Kruschke, 1996).

Distributed Hidden-Layer Representations

Some of the best known models of certain cognitive processes assume that distributed representations mediate between input and output representations (for example, Seidenberg & McClelland's, 1989, account of word naming). Distributed hidden-layer representations are most naturally conceived in featural terms, although geometric interpretations are also possible. Each input representation produces a set of activities across the features of the distributed hidden-layer representation. This pattern of hidden-layer activities arises through the, initially random, connections from the input representation. The hidden-layer representations are assumed to develop over time by a process of error attribution and reduction, typically via the back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986; Werbos, 1974).

Distributed hidden-layer representations are relatively rare in models of categorization, but they have been used with some success in models that attempt to

characterize the effects of hippocampal damage on the ability to categorize (Gluck & Myers, 1997).

Evidential Mechanisms

Most models of categorization convert the information from the intermediate representations (or input representations) into a set of evidence magnitudes, generating one evidence magnitude for each category under consideration. So, for example, a particular pattern of activity in the intermediate representations might give the evidence magnitudes of 0.87 and 0.42 for categories A and B, respectively. It is these numbers that eventually give rise to a decision about the category membership of the presented item, via a categorical decision process. In the next three sections, some of the more common ways of calculating evidence terms are discussed.

Summed Similarity

Probably the most common form of evidence magnitude is a summed similarity (e.g., Nosofsky, 1986; Smith & Minda, 1998). When a stimulus X is presented, the evidence that it belongs to category Y is assumed to be the sum of the similarity of stimulus X to all relevant intermediate representations associated with category Y. For example, if the intermediate representation is exemplar based, then the evidence magnitude for category Y is the sum of the similarities of stimulus X to all stored exemplars known to belong to category Y. Similarity is calculated in the manner described in the "Input Representations" section of this chapter. For example, in a geometric exemplar model, similarity is related to distance in a psychological space by an exponential or Gaussian decay function (see Fig. 23.2). In featural model, similarity is an increasing function of the number of features shared, and a decreasing function of the number of features that

are different. Stewart and Brown (2005) have recently proposed a geometric exemplar-based model that employs both similarity to category Y exemplars, and dissimilarity to members of other categories, in the calculation of the evidence term for category Y.

Decision Bounds

Decision bounds are most naturally conceptualized as working directly on a geometric input representation. The input representation is often expressed in terms of physical measurements of the stimuli (e.g., size), rather than a psychological space derived from similarity or confusability data. A decision bound is a line through that space that separates one category from another (see Fig. 23.5A). Sometimes that line is assumed to be straight (Ashby & Gott, 1988); sometimes it is assumed to be quadratic (Ashby & Maddox, 1992). In some applications of a decision-bound evidential process, the decision bound is assumed to be placed optimally (Ashby & Gott, 1988); in other words, its form and location is such that categorization accuracy is maximized.

The position of the presented stimulus relative to that line determines its category membership. As should be apparent, there are certain category structures that cannot be captured by a single linear or quadratic decision bound. An example is given in Figure 23.5B; a potential solution is to assume that more than one decision bound is used.

Note that, unlike a summed similarity mechanism, on any given presentation of a stimulus the evidence magnitudes for all categories except one are zero, and for the remaining category the evidence is maximal. In Figure 23.5A, the presentation of any stimulus in category 1 results in maximal evidence magnitude for category 1 and zero evidence magnitude for category 2. However, decision-bound models typically assume perceptual noise (Ashby & Gott, 1988), so a given stimulus will not always be

represented in exactly the same location in the input representation. Some decision-bound models also assume decisional noise—in other words, that the decision bound varies somewhat from decision to decision.

One striking aspect of a decision-bound evidential mechanism is that it is distance from the decision boundary, rather than distance from the known examples of a category, that determines categorization accuracy. This is illustrated in Figure 23.5C, where the novel stimuli marked "2" are predicted to be classified at least as accurately as the novel stimuli marked "1," and more accurately if one assumes substantial perceptual and decisional noise. Hence, a decision-bound evidence mechanism can be said to extrapolate from the known members of category. Extrapolation is observed in categorization experiments, and at least some extrapolation phenomena seem difficult to explain without positing a decision-bound evidence mechanism (Denton, Kruschke, & Erickson, 2008).

<insert Figure 23.5 here>

Verbalizable Rules

The idea that people make decisions on the basis of "rules" is a pervasive concept in psychology and in everyday life, but what does it mean to say performance is "rule-based"? Decision-bound theories seem in some ways to be quite rule-like, for example, in their ability to predict extrapolation. On the other hand, some would argue (Ashby et al., 1998) that the decision bound in Figure 23.5A is unlikely to represent a rule, because it is not easily verbalizable. A verbal representation of this decision bound would have to be of the form "It's category 1 if it's more obtuse than it is large, and category 2 if it's less obtuse than it is large." Among the problems with formulating and applying a rule of this type is that the things being compared have different units. In other words, what does it mean to say something is more obtuse than it is large? In contrast, the rule "category A is

composed of small obtuse items" is readily formulated verbally and can be applied without having to compare things measured in different units.

Hence, what some theorists mean when they talk of a categorization model being rule based is a decision bound that can be readily expressed in verbal terms. Limiting the concept of "rules" to things that are verbalizable is not an idea that is universally accepted in psychology, but it does currently have currency in the modeling of categorization. This view of rules as verbalizable naturally leads to the assumption that easy-to-verbalize rules (e.g., "Category A is blue") are more likely to be employed than hard-to-verbalize rules (e.g., "Category A is small and blue, or large and red"). This assumption is instantiated in some rule-based models of categorization such as RULEX (Nosofsky, Palmeri, & McKinley, 1994) and COVIS (Ashby et al., 1998). The extent to which participants spontaneously prefer simple rules over complex rules depends on the time available for a decision (e.g., Milton, Longmore, & Wills, 2008), with greater preference for complex rules where time permits. Hence, while rules are sometimes simple, simplicity should not be considered a defining property of a rule (although see Pothos, 2005, for a contrasting perspective). Finally, it is worth noting that there are some categories that seem to be rule based but not describable in terms of a decision bound, for example, the category of prime numbers.

Decisional Mechanisms

In some models, the evidential mechanisms themselves result in a categorical decision (e.g., certain decision-bound models; see earlier). However, in most models, the output of the evidential mechanism is a set of non-zero evidence magnitude terms. For example, the evidence terms for a presented item being a cat, a dog, or a bagel might be 0.83, 0.41,

and 0.05, respectively. In many situations, a decision is required—should I call this thing a cat, a dog, or a bagel?

The seemingly obvious answer to this question is that I should call it a cat, because that is the category for which the evidence is greatest. This strategy of "pick the biggest" is not, however, what the vast majority of models of categorization do. Instead, they engage in a form of probability matching. Applying the simplest form of probability matching to the earlier example, the probability of the model responding "cat" would be 0.83/(0.83 + 0.41 + 0.05) = 0.64. So despite "cat" being the most likely answer, that answer is only produced on 64% of occasions. This decision mechanism is generally known as the Luce choice axiom (Luce, 1959).

Although it is well known that organisms probability match (e.g., Herrnstein, 1961), it is generally accepted that the level of probability matching seen in studies of categorization is much lower than a simple application of the Luce choice axiom would predict (Ashby & Gott, 1988; McKinley & Nosofsky, 1995). One solution to this problem is to transform the evidence terms (v) in some way—for example, by using v^k or e^{kv}. Increasing the value of k reduces the level of probability matching predicted by the model—large values of k approximate a "pick the biggest" strategy. Applying v^{10} to our earlier example, the probability of responding "cat" exceeds 0.999.

The fact that the Luce choice axiom can approximate the behavior of a pick-the-biggest strategy does not imply that the two formulations are equivalent. In fact, Yellott (1977) demonstrated mathematically that noisy pick-the-biggest (in other words, picking the biggest under conditions where the evidence terms for a given item vary) is not equivalent to the Luce choice axiom in virtually all situations that involve three or more

response options. In cases where researchers have investigated the nature of the decision mechanism in three-choice situations, the empirical evidence favors Thurstonian (i.e., noisy pick-the-biggest) choice over the Luce choice axiom (Wills, Reimers, Stewart, Suret, & McLaren, 2000).

Multiprocess Models

Over the last decade, it has become increasingly common to propose that there are multiple categorization processes at work. For example, ATRIUM (Erickson & Kruschke, 1998) assumes the presence of both a decision-bound process and an exemplar-based process. Each of these processes provides a set of evidence terms for the possible category responses, and hence part of the decision process involves the integration of this information. In ATRIUM, this process involves keeping track of the past success of each process in providing the correct answer for the stimulus presented. The COVIS model (Ashby et al., 1998) also assumes a rule-like and exemplar-like process, although the details are different.

The Time Course of Categorization

To summarize what has been said so far, models of categorization assume the passing of information from attentionally modulated input representations, through intermediate representations, to evidential and decisional processes. A categorical decision is the result. In this context, the time course of categorization can be considered in two ways. First, one can consider the time course within a single decision—how is information accumulated over the time course from stimulus presentation to decision, and what additional phenomena can be captured by modeling this accumulation of information? Second, one can consider the time course across multiple decisions—what are the

mechanisms that allow the system's ability to categorize to improve with experience? These two questions are considered in the following sections.

Single-Decision Time Course

There are two main ways information is assumed to accumulate over the course of a single decision: at the level of input representations, and at the level of categorical decisions. At the level of input representations, models such as the Extended Generalized Context Model (Lamberts, 1995) assume that the dimensions of the input representation become available at different intervals after the presentation of the stimulus, with the average interval being a function of both the perceptual salience of that dimension and its usefulness in determining category membership of the item. Evidence in support of this form of information accumulation includes the fact that time pressure can systematically change the category into which a stimulus is placed. For example, under time pressure the stimulus might be systematically considered to be in category A, while in the absence of time pressure, it was systematically considered to be in category B. Such "cross-over" effects (Lamberts & Freeman, 1999) can be explained by assuming that, under time pressure, not all of the dimensions of the stimulus representation are available to later components of the categorization process (Lamberts, 1995; Milton et al., 2008).

Some models, such as the Exemplar-based Random Walk model (Nosofsky & Palmeri, 1997), the EGCM-RT model (Lamberts, 2000), and the winner-take-all model (Wills & McLaren, 1997), also assume that information accumulates over time at the level of categorical decisions. Making this assumption allows these models to predict the time taken to make a decision, in addition to the more standard prediction of the probability with which a particular category will be chosen.

Multiple-Decision Time Course

The discussion of the components of categorization models in this chapter assumed the presence of a lot of information. Models are assumed to have information about which dimensions and/or features to attend to in order to maximize accuracy. They are assumed to have information about how the structure of the stimuli can best be represented within the model's chosen intermediate representations (e.g., prototypes). Models are also assumed to know which intermediate representations correspond to which category labels and/or which decision bound to use to maximize categorization accuracy. How is this information acquired?

Perhaps surprisingly, many models of categorization have no specific mechanisms by which they can acquire the information they are assumed to have; those that do mainly rely on the concept of reduction of prediction error through, for example, changing the structure of associative connections between representations (Gluck & Bower, 1988; Rescorla & Wagner, 1972; Widrow & Hoff, 1960).

This principle is illustrated in the very simple featural model shown in Figure 23.6. The model is too simplistic to be a convincing account of categorization, but it serves to illustrate the principle of the reduction of prediction error. In an experiment where participants have to predict which fictitious disease a patient has on the basis of his symptoms (e.g., Gluck & Bower, 1988), it is presumably the case that participants will not start the experiment with the information required to solve it. This state of affairs is illustrated in Figure 23.6 by assuming a series of associations of arbitrary strength between the input and category representations. Now assume that the participant sees a patient with symptoms 1 and 2. On the basis of the connections shown in Figure 23.6,

there is more evidence (within this initially arbitrary knowledge) that the patient has disease 1 than that he has disease 2. Assuming a take-the-best decision mechanism, the model predicts that the patient has disease 1 but is told that the patient actually has disease 2. The model has therefore made a prediction error. To reduce the likelihood of a future error, the model increases the strength of the connection between symptom 1 and disease 2 (S1-D2), increases S2-D2, decreases S1-D1, and decreases S2-D1. In this particular instantiation of prediction-error-driven learning, connections from absent symptoms (symptom 3) do not change. The process just described is approximately that undertaken by the Widrow-Hoff rule (1960), and it is closely related to the Rescorla-Wagner theory (1972) and the LMS rule (Gluck & Bower, 1988).

<insert Figure 23.6 here>

The acquisition of other information within the categorization process is also commonly assumed to be driven by prediction error. For example, in the ALCOVE model (Kruschke, 1992), selective attention to the stimulus dimensions that produce prediction errors is reduced, and selective attention to the stimulus dimensions that reduce error is increased. The McLaren and Mackintosh model (2000, 2002) accounts for unitization and differentiation by assuming that the system attempts to predict the co-occurrence of stimulus features, and that associative links are modified such that prediction errors are reduced. The SUSTAIN model (Love et al., 2004) assumes that a new cluster representation is formed when the existing clusters fail to predict the category membership of the presented stimulus

Role of Feedback

The concept of prediction error, discussed earlier, could be taken to imply that category learning is only possible where some external agent (e.g., a teacher) or the environment

provides specific information about the category membership of presented stimuli. This is not the case; category learning can and does occur in the absence of feedback (e.g., Homa & Cultice, 1984; Wills & McLaren, 1998). Relatively few models of categorization can account for this phenomenon. Those that can—such as Adaptive Resonance Theory (Grossberg, 1976), the Rumelhart and Zipser model (1986), and SUSTAIN (Love et al., 2004)—do so by assuming that the categorical decision produced by the model is correct, and adjusting associative strengths and other parameters in the same way as if "correct" feedback had been received. Where the stimuli presented to the model are well structured (as in, for example, Fig. 23.5A), such models are able to produce categorical representations that capture much of that structure.

Conclusions and Future Directions

Over the last 35 years, there has been a great deal of progress in the modeling of categorization. Since Hull's work (Hull, 1920) and, perhaps more famously, the work of Bruner and colleagues (Bruner, Goodnow, & Austin, 1956), the field had been steadily accumulating empirical phenomena, but it was not until the early 1970s (e.g., Reed, 1972) that theories of these phenomena started routinely taking mathematical form. The late 1970s to the mid-1990s saw the introduction and development of a number of still highly influential single-process models, such as the Generalized Context Model (Nosofsky, 1986) and General Recognition Theory (Ashby & Gott, 1988). Toward the end of the 1990s, there was increasing recognition that categorization may involve more than one competing process, and the development of multiprocess theories such as COVIS (Ashby et al., 1998) and ATRIUM (Erickson & Kruschke, 1998), and also the idea that intermediate representations may develop and change in specificity over time

(e.g., Smith & Minda, 1998). COVIS is also notable in that it is a theory of categorization specified in both formal mathematical terms and related to the assumed underlying neuroscience (another example of this combined approach is the Gluck-Myers model; e.g., Gluck & Myers, 1997).

The first decade of the 21st century saw a rapidly increasing data set on the neuroscience of categorization and the development of existing theories to account for these data (e.g., Ashby, Ennis & Spiering, 2007). It also saw attempts to expand the range of behavioral phenomena to be modeled to include, for example, classification in the absence of feedback, classification and feature inference (e.g., Love et al., 2004), and the effects of background knowledge (e.g., Rehder & Murphy, 2003). The broadening of data to both more behavioral phenomena and to neuroscientific data is very welcome, as both sets of information should serve to constrain and reduce the number of formal models that remain plausible accounts of the known phenomena.

In conclusion, formal modeling of psychological phenomena is a complex and time-consuming task. What indicates that it is a worthwhile enterprise? Probably the single biggest advantage of formal modeling over more informal forms of theorizing is that the ability of a formal theory to encompass an empirical phenomenon is unambiguously determinable. This can sometimes lead to surprising conclusions. For example, asymmetry in similarity relations (an ellipse is more similar to a circle than a circle is to an ellipse) seems, informally, to be incompatible with the idea that similarity relations can be represented in a geometric space. Yet categorization models employing a geometric space can unambiguously be demonstrated to be able to capture certain asymmetric similarity relations (Nosofsky, 1991). Another example is that some

philosophers have argued the concept of similarity is ill defined because any two objects have an arbitrary number of things in common (e.g., An ostrich and an aircraft carrier? Both weigh more than one gram, neither can fly, both can be found on the Earth, both can carry people, etc.). This is known as Goodman's paradox (Goodman, 1972). Models of categorization may help us work through Goodman's paradox by specifying, for example, the ways in which attention to features is directed through experience.

The unambiguous specification of theory that formal modeling brings should also, in principle, confer two further advantages. First, it should be possible to compare different theories in an unambiguous manner and come to a consensual conclusion about which theory encompasses more of the known empirical phenomena. Second, the formal specification of theories of psychological processes should bring with it the possibility of re-creating those processes in artificial systems (automated cognition). However, the potential of both these aspects currently remains largely unfulfilled in the formal modeling of categorization. The relative adequacy of different models of categorization is seldom systematically compared, and never across a broad range of empirical phenomena. Perhaps as a result of this, the field continues to have a large range of competing formal theories with no consensus over their relative adequacy. The lack of such consensus may in part explain the relative lack of successes in the application of formal categorization theory to the development of automated cognition. The resolution of these issues is the single biggest challenge that the formal modeling of categorization must address in the coming years.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*, 632–656.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 50–71.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Denton, S. E., Kruschke, J. K., & Erickson, M. A. (2008). Rule-based extrapolation: A continuing challenge for exemplar models. *Psychonomic Bulletin and Review*, *15*(4), 780–786.

Ennis, D.M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, *117*, 408–411.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal Of Experimental Psychology: General*, *127*(2), 107–140.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*, 94–107.

Garner, W. R. (1978). Aspects of a stimulus: Features, dimensions and configurations. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 99–133). Hillsdale, NJ: Erlbaum.

Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, *27*, 166–195.

Gluck, M. A., & Myers, C. E. (1997). Psychobiological models of hippocampal function in learning and memory. *Annual Review of Psychology*, *48*, 481–514.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 86–112.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–447). New York: Bobbs-Merrill.

Granger, R., Ambros-Ingerson, J., & Lynch, G. (1989). Derivation of encoding characteristics of layer II cerebral cortex. *Journal of Cognitive Neuroscience*, *1*, 61–87.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, *23*, 121–134.

Harris, J. A. (2006). Elemental representations of stimuli in associative learning. *Psychological Review*, *113*, 584–605.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267–272.

Holman, E. W. (1979). Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology*, *20*, 1–15.

Homa, D., & Cultice, J. C. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*(1), 83–94.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts: An experimental study. *Psychological Monographs*, *28*(1), No. 123.

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

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, *5*(1), 3–36.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*(2), 225–247.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology-Learning Memory and Cognition*, *31*(5), 830–845.

Kruskal, J. (1964). Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–28.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*(2), 161–180.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260.

Lamberts, K., & Freeman, R. P. J. (1999). Building object representations from parts: Tests of a stochastic sampling model. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 904–926.

Le Pelley, M. E., & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *Quarterly Journal of Experimental Psychology*, *56B*, 68–79.

Lochmann, T., & Wills, A.J. (2003). Predictive history in an allergy prediction task. In F. Schmalhofer, R. M. Young, & G. Katz (Eds.), *Proceedings of EuroCogSci 03: The European Cognitive Science Conference* (pp. 217–222). Mahwah, NJ: Erlbaum.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*(2), 159–188.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision-bound models in large-size, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 128–148.

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning and Behavior*, *28*, 211–246.

McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning and Behavior*, *30*, 177–200.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.

Milton, F., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 676–692.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, *23*, 94–140.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and

extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition*, *22*(3), 352–369.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences*, *28*, 1–49.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, *26*, 303–343.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*, 382–407.

Rehder, B., & Hoffman, A.B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.

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

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research* (pp. 64–99). New York: Appleton-Century-Crofts.

Restle, F. (1959). A metric and an ordering on sets. *Psychometrika*, *24*(3), 207–220.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland

(Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 151-193). Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Shanks, D. R., & Gluck, M. A. (1994). Tests of an adaptive network model for the identification and categorization of continuous-dimension stimuli. *Connection Science*, *6*(1), 59–89.

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

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*, 509–523.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1411–1436.

Stewart, N., & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, *49*, 403–409.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin and Review*, *15*(4), 732–749.

Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. Unpublished Ph.D. dissertation, Harvard University, Boston.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*. Paper presented at the IRE WESCON Convention.

Wills, A. J., Lavric, A., Croft, G. S., & Hodgson, T. L. (2007). Predictive learning, prediction errors, and attention: Evidence from event-related potentials and eye tracking. *Journal of Cognitive Neuroscience*, *19*(5), 843–854.

Wills, A. J., & McLaren, I. P. L. (1997). Generalization in human category learning: A connectionist explanation of differences in gradient after discriminative and non-discriminative training. *Quarterly Journal of Experimental Psychology*, *50A*(3), 607–630.

Wills, A. J., & McLaren, I. P. L. (1998). Perceptual learning and free classification. *Quarterly Journal of Experimental Psychology*, *51B*(3), 235–270.

Wills, A. J., Noury, M., Moberly, N. J., & Newport, M. (2006). Formation of category representations. *Memory and Cognition*, *34*(1), 17–27.

Wills, A. J., Reimers, S., Stewart, N., Suret, M., & McLaren, I. P. L. (2000). Tests of the ratio rule in categorization. *Quarterly Journal of Experimental Psychology*, *53A*(4), 983–1011.

Wills, A. J., Suret, M., & McLaren, I. P. L. (2004). The role of category structure in determining the effects of stimulus preexposure on categorization accuracy. *Quarterly Journal of Experimental Psychology*, *57B*(1), 79–88.

Yellott, J. I., Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*, 109–144.

Further Reading

Kruschke, J. K. (2008). Models of categorization. In: R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology*, pp. 267–301. New York: Cambridge University Press.

Pothos, E.M. & Wills, A.J. (2011). *Formal approaches in categorization*. Cambridge University Press.

Wills, A. J. (2009). Prediction errors and attention in the presence and absence of feedback. *Current Directions in Psychological Science*, 18(2), 95–100.

Wills, A.J. & Pothos, E.M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin, 138,* 102-125.

Figure 23.1

Components of a model of categorization.

Figure 23.2

(*A*) Representing the similarity structure of stimuli 1, 2, and 3 in a two-dimensional geometric space; in this example the dimensions of this space are readily interpretable as size and angle. (*B*) An exponential decay relationship between similarity and distance in

psychological space. (*C*) Euclidean distance ($distance^2 = x^2 + y^2$). (*D*) City-block distance (distance = x + y).

Figure 23.3

(*A*) Geometric representation of two categories, each of four stimuli (category membership denoted by type of dot). (*B*) Stretching along the x-axis and compression along the y-axis, thereby increasing within-category similarity and decreasing between-category similarity. (*C*) A category structure for which selective attention to the x-axis would be less helpful than in Figure 23.3A. (*D*) Overall expansion of psychological similarity space.

Figure 23.4

Geometric representation of two categories, each of eight items; category membership denoted by type of circle. C, cluster; P, prototype.

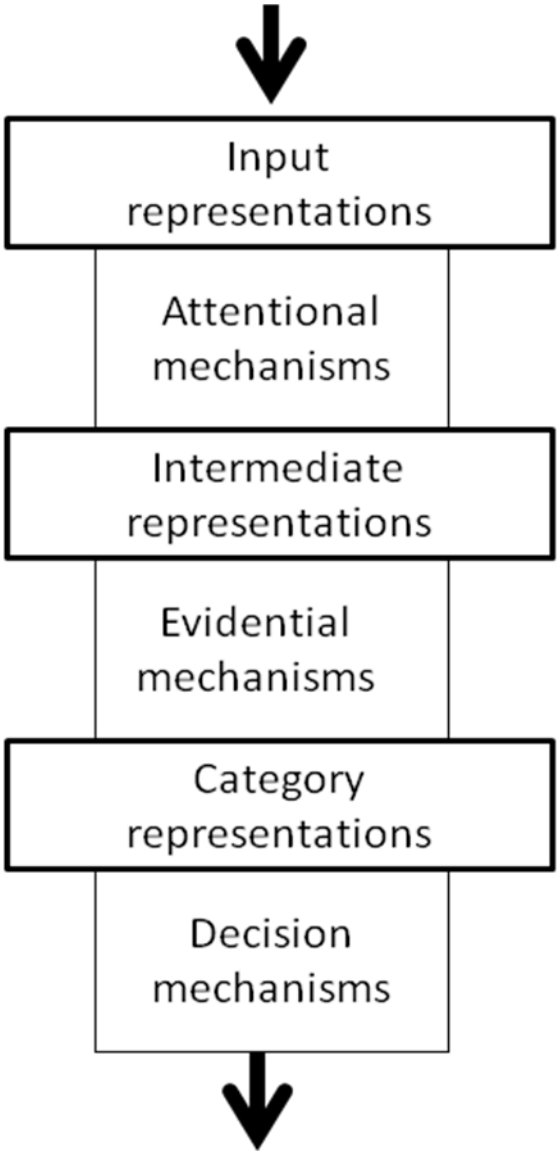
Figure 23.5

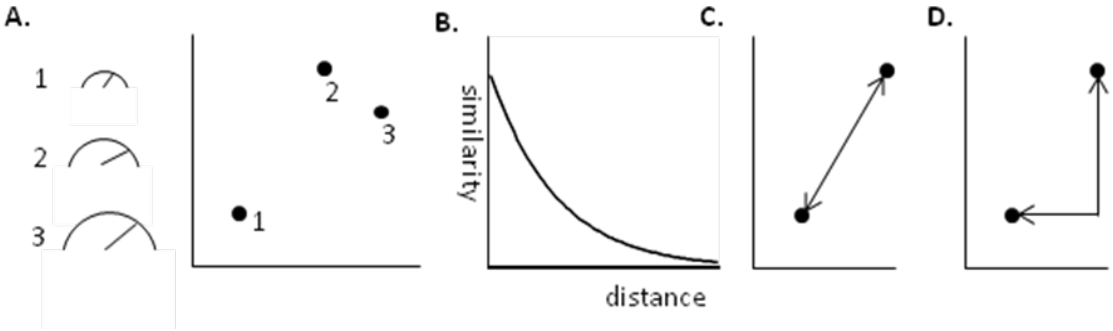
(*A*) Geometric representation of two categories, and a linear decision bound separating them; category membership denoted by type of circle. (*B*) A category structure that cannot be represented by a single linear or quadratic decision bound. (*C*) In a decision-bound account, novel stimuli marked “2” will be responded to at least as accurately as novel stimuli marked “1.”

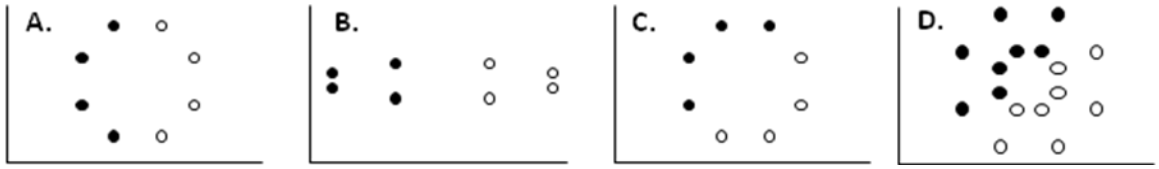
Figure 23.6

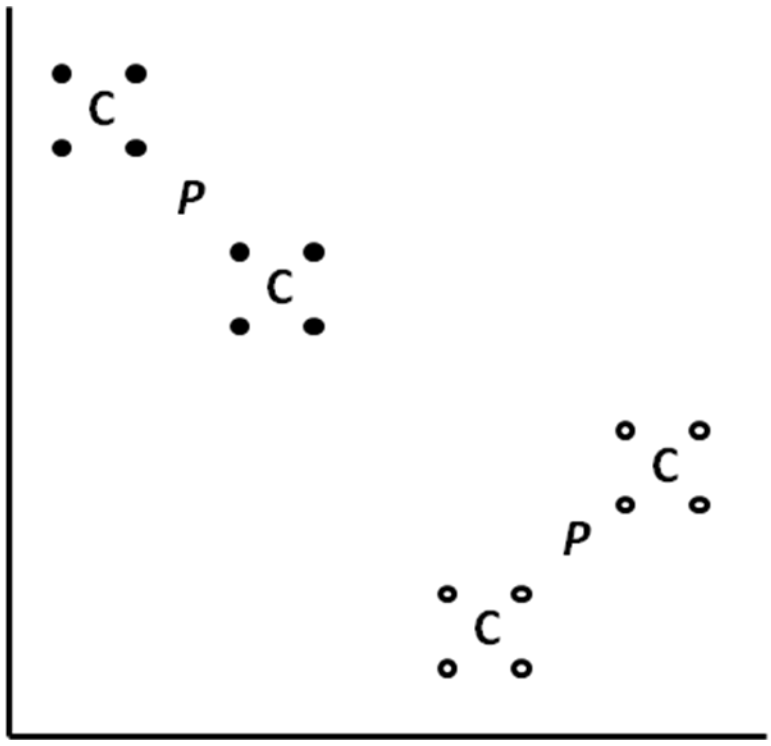
A simple associative learning model of a disease prediction task. Units representing symptoms have connections of variable strength to units representing disease. Representational units are shown as circles; associative connections are shown as arrows.

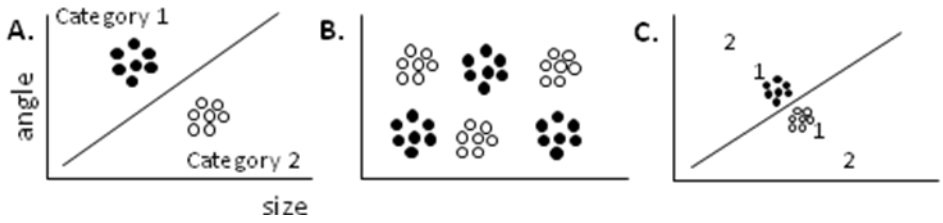
The thickness of the arrows denotes the relative strength of the connections in this example.

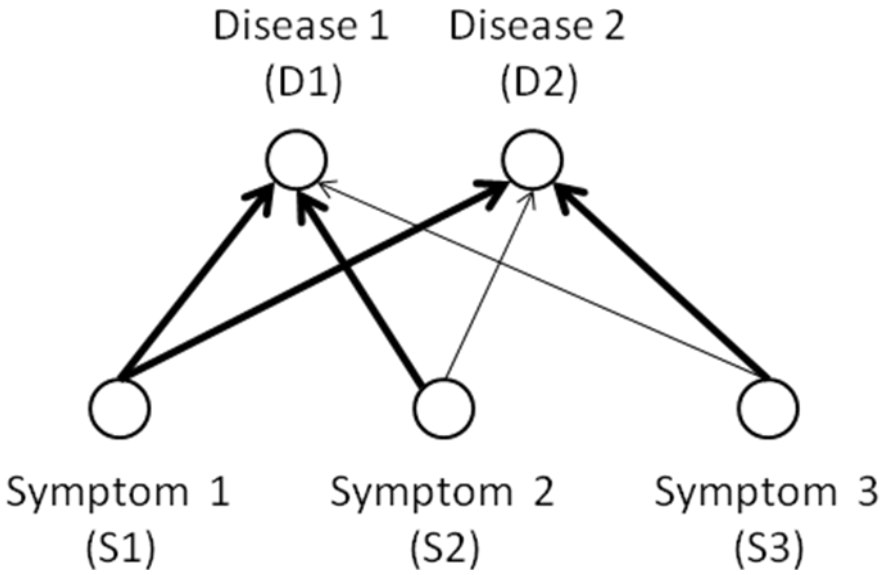

Input representations
Attentional mechanisms
Intermediate representations
Evidential mechanisms
Category representations
Decision mechanisms


A.
1
2
3
1
2
3
B.
similarity
distance
C.
D.


A.
B.
C.
D.


C
P
C
C
P
C


A.
Category 1
angle
Category 2
size
B.
C.
2
1
1
2


Disease 1
(D1)
Disease 2
(D2)
Symptom 1
(S1)
Symptom 2
(S2)
Symptom 3
(S3)