

# A Neuropsychological Theory of Multiple Systems in Category Learning

F. Gregory Ashby, Leola A. Alfonso-Reese, And U. Turken, and Elliott M. Waldron  
University of California, Santa Barbara

A neuropsychological theory is proposed that assumes category learning is a competition between separate verbal and implicit (i.e., procedural-learning-based) categorization systems. The theory assumes that the caudate nucleus is an important component of the implicit system and that the anterior cingulate and prefrontal cortices are critical to the verbal system. In addition to making predictions for normal human adults, the theory makes specific predictions for children, elderly people, and patients suffering from Parkinson's disease, Huntington's disease, major depression, amnesia, or lesions of the prefrontal cortex. Two separate formal descriptions of the theory are also provided. One describes trial-by-trial learning, and the other describes global dynamics. The theory is tested on published neuropsychological data and on category learning data with normal adults.

Humans are remarkably adept at categorizing objects and events in their environment. In fact, it is now well established that humans can learn some extremely complex (i.e., nonlinear) categorization rules (e.g., Ashby & Maddox, 1992; McKinley & Nosofsky, 1995; Medin & Schwanenflugel, 1981). One characteristic of demanding categorization problems is that experts use rules that are often difficult or impossible to describe verbally. For example, it is difficult to verbalize the decision rules used by farmers to sex chicks, those used by wine tasters to determine that a certain wine is a Zinfandel or a Cabernet Sauvignon, or those used by artists to categorize unfamiliar paintings according to the Renaissance master who created them. On the other hand, in many cases, contrasting categories are separated perfectly (or nearly so) by some decision rule that can be described verbally. For example, a simple verbal rule separates triangles from rectangles, oranges from lemons, and evergreens from deciduous trees.

Current theories of category learning do not discriminate between these two kinds of tasks. Rather, they assume that all rules, whether verbal or nonverbal, are learned by using the same basic processes. Nevertheless, growing evidence indicates a qualitative difference in performance depending on whether the optimal decision rule—that is, the rule that maximizes categorization accuracy—can be described verbally. For example,

people can learn complex rules that are difficult or impossible to verbalize; when feedback about accuracy is withdrawn, however, people are restricted to rules that are easily verbalized (see the next section for details). Also, recent neuroimaging results indicate qualitatively different patterns of activation in verbal and nonverbal categorization tasks (E. E. Smith, Patalano, Jonides, & Koeppel, 1996). A complete theory of category learning must account for these qualitative differences.

This article proposes a formal neuropsychological theory that assumes people have available at least two separate categorization systems: a verbal system based on explicit reasoning that is under conscious control and a nonverbal implicit system that uses procedural learning. The theory generalizes and formalizes a number of recent results from the neurosciences. Principal among these are Squire's (1992) hypothesis that the striatum plays a key role in categorization and Posner and Petersen's (1990) model of executive attention that assigns key functions to the cingulate and prefrontal cortices.

The theory that we develop here assumes that category learning is a competition between separate verbal and implicit<sup>1</sup> categorization systems. As a consequence, we refer to this new theory by the acronym COVIS (competition between verbal and implicit systems). COVIS assumes that the verbal system initially dominates, presumably because it is controlled by consciousness. With training and experience, however, the potential of the implicit system for superior performance often overcomes the initial bias in favor of the verbal system. Nevertheless, COVIS predicts that, in most cases, both the verbal and the implicit systems remain active even after learning is complete,

---

F. Gregory Ashby, Leola A. Alfonso-Reese, And U. Turken, and Elliott M. Waldron, Department of Psychology, University of California, Santa Barbara.

This research was supported by Grants DBS92-09411 and SBR95-14427 from the National Science Foundation and by a Ford Foundation Dissertation Fellowship.

We thank Jerome Busemeyer, Aaron Ettenberg, Thomas Fikes, Vincent Filoteo, Alice Isen, Richard Ivry, Stanley Klein, Barbara Knowlton, Gordon Logan, Loy Lytle, Todd Maddox, Douglas Medin, Robert Nosofsky, Art Shimamura, and Michael Wich for their helpful comments and suggestions.

Correspondence concerning this article should be addressed to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, California 93106. Electronic mail may be sent to ashby@psych.ucsb.edu.

---

<sup>1</sup> Often, learning and knowledge are labeled *implicit* only if they occur without conscious awareness, that is, only if the participant is unaware of the knowledge that has been obtained and is even unaware that any learning has occurred (e.g., Reber, 1989; Shanks, Green, & Kolodny, 1994). For the implicit system we are postulating, the latter of these requirements will not generally be met. Because explicit feedback about categorization accuracy is commonly given, participants in categorization experiments will generally be explicitly aware that learning has occurred.

and each determines a significant proportion of categorization judgments.

Marr (1982) argued persuasively that, if a theory is to be complete, it should be formulated at several different levels. Specifically, he argued for the importance of descriptions at the implementational (i.e., hardware), algorithmic, and computational levels. Following Marr, we provide separate descriptions of COVIS at each of these levels. Together, the three resulting versions of COVIS provide a far broader theory of category learning than would be possible with a model described at only one of Marr's (1982) levels. At the implementational level, we describe the neural structures and pathways that COVIS assumes mediate category learning. To our knowledge, this is the first attempt to provide a detailed neuropsychological description of categorization. Not surprisingly, this version of the theory is necessarily speculative. Even so, we offer it here because it suggests a wide variety of new research directions that will likely lead to important insights into categorization processes, even if many of the details of the neuropsychological version of COVIS prove to be incorrect. For example, this version of COVIS makes specific predictions about category learning in patients suffering from lesions of the prefrontal cortex, from Parkinson's or Huntington's disease, and from depression or amnesia, as well as in children and elderly people. The algorithmic-level description of COVIS is a connectionist network that makes specific trial-by-trial category learning predictions. The

computational-level description is a formal model of the global dynamics of category learning that makes it easy to derive predictions about important experimental variables such as quality of feedback and salience of alternative verbal rule types.

This article begins with a review of the evidence supporting the notion of multiple categorization systems. The neuropsychological version of COVIS is then presented, and some of its predictions for various special populations are tested on published neuropsychological data. Next, the algorithmic-level description, which is based on standard connectionist architecture, is developed and tested. Finally, the computational version of COVIS is described, and the theory's global dynamics are explored. Several strong predictions of these formal versions of the model are derived and tested, both on previously published data and on data from a new experiment. We close by considering a prominent alternative to the multiple systems postulated by COVIS.

### Multiple Categorization Systems

The distinction between verbal and nonverbal rules is illustrated in Figure 1. Figure 1a shows the category prototypes from an experiment reported by Maddox and Ashby (1993) that used two different types of stimuli: rectangles that varied in height and width and circles that varied in size and in the orientation of a radial line. Figure 1b shows the category structure

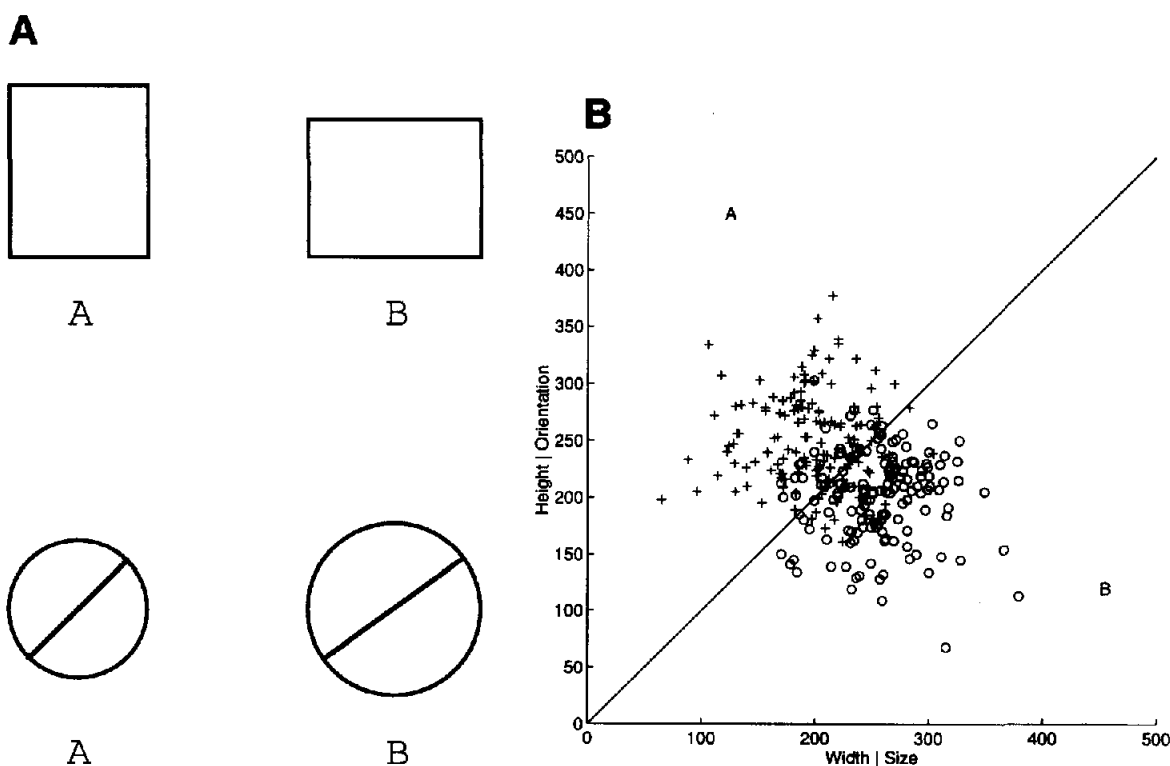


Figure 1. Stimuli and category structure of an experiment reported by Maddox and Ashby (1993). Panel (a): Category (A and B) prototypes for the rectangle condition and the circle condition. Panel (b): Common category structure used in both conditions. A plus sign indicates an exemplar from Category A, and a circle indicates an exemplar from Category B. The diagonal line is the decision bound that maximizes accuracy.

from the Maddox and Ashby (1993) experiment. The abscissa and ordinate correspond to width and height, respectively, in the case of the rectangles and to size and orientation, respectively, in the case of the circles. The plus signs indicate the coordinates of the exemplars of Category A, and the circles indicate the exemplars of Category B. The categories overlap, so perfect performance is impossible. Accuracy is maximized (at 82.5%) by the rule associated with the diagonal bound. Specifically, the optimal rule is to respond A to any stimulus falling above the bound and B to any stimulus falling below. On each trial of the experiment, a stimulus is sampled randomly from one of these two categories and shown to the observer (as a rectangle or a circle, depending on the experimental condition). The observer's task is to assign the stimulus to Category A or B (by pressing a button). Feedback indicating the correct category is given after every trial. This "randomization technique" was used effectively by Ashby and Gott (1988) and Ashby and Maddox (1990, 1992).

With the rectangles, the optimal rule in this task has the following simple verbal description: "Respond A if the stimulus rectangle is taller than it is wide; respond B if the rectangle is wider than it is tall." For the circles, however, the same rule has no simple verbal description. The best approximation is as follows: "Respond A if the orientation of the radial line is greater than the size of the circle; respond B if the size is greater than the orientation." However, because orientation and size are expressed in different units, this is like comparing apples and oranges. We have collected extensive amounts of data in tasks in which the optimal rule is difficult or impossible to verbalize. After the last experimental session, we typically ask participants for a verbal description of their response strategy. None of the participants in these "nonverbal" tasks have ever described the optimal rule. At best, they give a complex description that only roughly approximates their performance. Many participants, however, simply say that their responses were just a "gut reaction." It is as if learning of these complex rules is implicit or procedural (e.g., Anderson, 1976; N. J. Cohen & Squire, 1980; Schacter & Tulving, 1994b; Squire, 1984). For convenience, we refer to decision rules that cannot be verbalized as nonverbal rules.

Maddox and Ashby (1993) included 6 participants in the experiment described in Figure 1. Three were shown rectangles, and 3 were shown circles. Each participant completed 400 trials per experimental session. Sessions were repeated over consecutive days until a criterion accuracy level had been reached (i.e., 75%). Table 1 summarizes performance. All participants in the rectangles condition exceeded the criterion accuracy level dur-

ing their first experimental session. Participants in the circles condition needed an average of two sessions to achieve criterion accuracy. In addition, the rectangles participants were more accurate during their first session than were the circles participants during their last session. Thus, performance in the two conditions was qualitatively different.

The Maddox and Ashby (1993) data also show that the distinction between verbal and nonverbal rules is not one simply of complexity. By any objective measure, complexity is identical in the circles and rectangles conditions; for example, optimal accuracy is the same, within- and between-category similarity is the same, and the optimal bound (i.e.,  $y = x$ ) is the same. Therefore, if the task is more difficult with circles than rectangles, then it cannot be because of any objective difference between the conditions. Rather, it must be due to a difference in how the human participants perceive the stimuli or solve the two categorization problems. Although the participants in the Maddox and Ashby (1993) experiment found the "verbal" (i.e., rectangles) condition to be easier than the "nonverbal" (i.e., circles) condition, it is important to note that we are not claiming that verbal categorization rules are always easier to learn than nonverbal rules. Clearly, some complex verbal rules are exceedingly difficult to learn. Our point only is that performance in tasks in which the optimal rule is easily verbalized is often qualitatively different from performance in tasks in which the optimal rule is difficult or impossible to verbalize.

There is other evidence for multiple systems of categorization. First and perhaps most impressive are the recent neuroimaging data of E. E. Smith et al. (1996). In this study, participants categorized artificial animals in separate verbal (i.e., rule) and nonverbal (i.e., similarity) conditions during position emission tomography (PET) scanning. Activation patterns were qualitatively different in the two conditions, and, on this basis, Smith et al. argued that different categorization mechanisms dominate the two conditions.

Second, category learning is qualitatively different depending on whether trial-by-trial feedback is provided. As mentioned earlier, with supervision (i.e., trial-by-trial feedback), people can learn complex, nonverbal rules (e.g., Ashby & Maddox, 1992; McKinley & Nosofsky, 1995; Medin & Schwanenflugel, 1981). In contrast, in unsupervised and free sorting tasks, in which no feedback is given as to the correct response, people almost always use simple verbal rules (Ahn & Medin, 1992; Ashby, Queller, & Berretty, 1997; Imai & Garner, 1965; Medin, Wattenmaker, & Hampson, 1987; Wattenmaker, 1992).

For example, in the Ashby et al., (1997) study, two widely separated categories were created from lines that varied in length and orientation. In two conditions a verbal (i.e., unidimensional) rule was optimal, and in two conditions the optimal rule was nonverbal. Participants used verbal rules in all four conditions. Interestingly, there was no evidence of learning in the nonverbal conditions, but there was strong evidence of learning in the verbal conditions. For example, in both verbal conditions, a number of participants initially tried a verbal rule of the incorrect type (i.e., a unidimensional rule on the wrong dimension), and then, in every case, they spontaneously switched to the correct verbal rule (i.e., to the correct dimension). In a follow-up nonverbal condition, participants were told explicitly that no unidimensional rule was optimal and,

Table 1  
Summary Statistics Describing Performance in the Experiment Shown in Figure 1 and Reported by Maddox and Ashby (1993)

Statistic	Rectangles	Circles
Days to criterion	1	2
Percentage correct	80.7	76.7
$d'$	1.74	1.48

therefore, that they had to use both stimulus dimensions when selecting a response. (In fact, the optimal rules were of the following type: "Respond *A* if length + orientation > some criterion; otherwise, respond *B*.") Despite these instructions, 8 of 10 participants used some form of unidimensional rule. The other 2 used a conjunctive rule of the form "Respond *A* if the length >  $x_1$  and the orientation >  $x_0$ ; otherwise, respond *B*, for some appropriately chosen constants  $x_1$  and  $x_0$ ." Thus, none of the participants responded optimally, and they all used rules that were easily verbalized. The 2 participants who used conjunctive rules are especially important because, in this experiment, a conjunctive rule is two dimensional but suboptimal. Thus, these data contradict the hypothesis that unidimensional rules are privileged in human category learning only because they are simpler (i.e., one dimensional) than nonverbal rules (which are two dimensional with the circle, rectangle, or line stimuli). If this were true, then the participants who expended the extra effort to use conjunctive rules should instead have responded optimally. The fact that they used conjunctive rules supports the hypothesis that humans have a separate category learning system that is constrained to use verbal rules.

The notion that there are multiple forms of reasoning is very old. Sloman (1996) traced the idea back to Aristotle and reviewed evidence that there are separate rule-based and associative reasoning systems. A similar distinction between explicit and implicit learning systems also evolved many years ago, originally to account for artificial grammar learning (e.g., Reber, 1967) and more recently to account for the many forms of human amnesia (e.g., N. J. Cohen & Squire, 1980). Even within the more specialized context of categorization, the idea of multiple systems is not new. For example, almost 20 years ago, Brooks (1978) distinguished between "deliberate, verbal, analytic control processes and implicit, intuitive, nonanalytic processes" (p. 207) that are active during category learning.

Although it seems clear that the verbal system, if it exists, would rely on explicit reasoning, the nature of the implicit system is less clear. An examination of the categorization and memory literatures points toward two prominent possibilities. Exemplar theory assumes that people assign objects to categories by accessing memory traces of exemplars from relevant categories (e.g., Brooks, 1978; Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). As such, exemplar theory hypothesizes that categorization depends on instance-based or exemplar-based memory. In contrast, decision bound theory assumes that people gradually learn to associate category labels (i.e., responses) with different regions of psychological space (Ashby, 1992; Ashby & Lee, 1991, 1992; Ashby & Maddox, 1990, 1992, 1993; Maddox & Ashby, 1993). The idea is that people learn a procedure for generating a response and do not access any exemplar memories during the categorization process. As such, the learning and memory involved are procedural (e.g., N. J. Cohen & Squire, 1980; Schacter & Tulving, 1994b; Squire, 1984). Both theories have been impressively successful at accounting for asymptotic categorization performance, and formal modeling attempts have been unable to distinguish empirically between the two (e.g., Maddox & Ashby, 1993; McKinley & Nosofsky, 1995).

Brooks (1978) postulated separate rule-based and exemplar-based categorization systems, and a formal model of this type

was recently proposed by Erickson and Kruschke (in press). In contrast, Squire (e.g., 1992) argued for separate exemplar-based and procedural learning systems. We believe that the neuropsychological data, reviewed later in this article, can be accounted for best by a model that assumes separate rule-based and procedural-learning-based categorization systems.<sup>2</sup> As mentioned earlier, we call the new theory based on this assumption COVIS. The rule-based, or theory-based, system is explicit and involves logical reasoning and semantic memory. Our operational definition is that this system operates on categorization rules that can be easily verbalized. The procedural-learning-based system is assumed to be implicit.

Categorization has not traditionally been considered to rely either on verbal-based reasoning or on procedural learning. As a consequence, before describing COVIS in detail, we briefly elaborate on the nature of the verbal system, and we consider the relation between category learning and more traditional procedural (i.e., motor) learning tasks.

### Verbal Rules in Categorization

Although previous theories of categorization have not explicitly considered verbal-based strategies, many theories have postulated categorization systems that are quite similar to the verbal system of COVIS. Historically, these systems have been of three types. The earliest notion was that categorization is a process of hypothesis testing (Bruner, Goodnow, & Austin, 1956). More recently, a number of theorists have argued for a rule-based category learning system (e.g., Busemeyer & Myung, 1992; Nosofsky, Palmeri, & McKinley, 1994) or a theory-based system (Murphy & Medin, 1985). The verbal system of COVIS should be considered as a particular instantiation of category learning systems of these types.

One difficulty with postulating a rule-based or theory-based category learning system is that the terms *rule* and *theory* are rarely defined rigorously. If the goal is to construct a computational model that makes specific quantitative predictions, then this is a serious problem. As a first step toward solving this problem, we operationally define *rule* or *theory* as any strategy that can be described verbally. This definition works well for the applications considered in this article, but we acknowledge that, under some conditions, explicit reasoning might occur that is not verbally based. As a consequence, the current version of the verbal system in COVIS should be considered tentative. A more complete account of this system awaits a detailed theory of explicit reasoning.

Simply defining a rule or theory as any strategy that can be verbalized does not solve all problems. In particular, any theory that postulates a separate categorization system that operates only on decision rules that can be verbalized must specify which rules can be verbalized and which cannot. This is also a difficult problem that we do not claim to have solved. Nevertheless, some inferences are possible. To begin, note that two operations are required to instantiate successfully a verbal rule such as the

<sup>2</sup> Even so, we suspect that a complete account of human category learning may require separate systems of all three types (i.e., rule, exemplar, and procedural learning).

following: "Respond *A* if the circle is large and *B* if it is small." First, of course, there must be a semantic label that corresponds to the particular property of the stimulus set on which the rule operates. In this case, the semantic label is "size." Second, the classifier must be able to attend selectively to this property of the stimulus. For example, it is possible to verbalize a rule such as "Respond *A* if the saturation of the color patch exceeds a criterion value; otherwise, respond *B*." Even so, it is generally thought that humans are incapable of attending to saturation in a way that is unaffected by variations in hue and brightness. A theory of verbal rules in categorization, therefore, must answer two questions. First, which rules can be verbalized? Second, to which components of a stimulus can the classifier selectively attend? We discuss each of these questions in turn.

An attempt to define precisely the complete set of verbalizable rules is beyond the scope of this article. However, it is straightforward to identify a large subset of verbalizable rules. In particular, most unidimensional rules are verbalizable. By a unidimensional rule, we mean a rule that uses a decision bound that is orthogonal to some stimulus dimension. Such a rule can always be written as follows: "Respond *A* if the value on dimension  $x > c$ ; respond *B* if the value on dimension  $x < c$ , for some appropriately chosen constant  $c$ ." In the categorization literature, unidimensional rules are associated with selective attention, but their relation to verbalization has been overlooked.

There are at least two arguments as to why unidimensional rules are verbalizable. First, one could argue that a major function of language is to denote psychologically salient aspects of the environment. Stimulus dimensions are those aspects of the stimulus that the experimenter chooses to manipulate. Because experimenters usually choose to manipulate psychologically salient aspects, stimulus dimensions will generally have an associated semantic label. The second argument works in the reverse direction. Because humans are verbal animals, experimenters are drawn to stimulus aspects that have semantic labels when selecting aspects of the stimulus to manipulate. Thus, unidimensional rules and verbal rules are highly confounded.

Of course, many rules that are not unidimensional can also be verbalized. For example, in the case of the circular stimuli shown in Figure 1, consider the following rule: "Respond *A* if the size exceeds the orientation; respond *B* if the orientation exceeds the size." Although such a rule can be verbalized, we have never had a participant describe his or her behavior in this fashion, even when his or her observable performance was well described by such a rule. Thus, we argue that rules of this type, although possible to verbalize, have very low saliency. As a result, the verbal categorization system will almost never spontaneously select such a rule. In fact, in this article, we assume that rules of this type are never selected by the verbal system.

Another important class of verbal rules comprises those involving a combination of unidimensional rules (among others, conjunctive, disjunctive, and exclusive-or rules). Whereas it is true that normal adults can learn these rules without excessive difficulty (e.g., Salatas & Bourne, 1974), it is quite clear that such rules are less salient than simple unidimensional rules. For example, in unsupervised categorization, unidimensional rules dominate (e.g., Ahn & Medin, 1992; Imai & Garner, 1965; Wattenmaker, 1992). In experiments of the type reported in this article, in which accuracy is maximized with a linear or qua-

dratic decision bound, it is extremely rare to encounter a data set that is better described by a conjunctive- or disjunctive-type rule than by a linear- or quadratic-bound-type rule.

Alfonso-Reese (1996) attempted to measure the salience of alternative categorization rules. She presented 47 participants with lines that varied uniformly across trials in length and orientation. Participants were asked to divide the stimuli into two categories according to any criterion they desired. This task was repeated for each participant five different times, so, in effect, each participant was asked to choose five different categorization strategies. Participants overwhelmingly preferred unidimensional rules. In fact, only 1 participant used a conjunctive rule, and this occurred on that participant's fifth and final sort. For these reasons, in the modeling that follows, we make the simplifying assumption that the verbal system does not routinely formulate rules of the conjunctive-disjunctive type.

The issue of which stimulus aspects can receive selective attention is more straightforward, at least theoretically. The critical distinction is whether the stimulus dimensions are *separable* or *integral*. With separable dimensions, it is straightforward to attend to one and ignore the other. With integral dimensions, however, it is difficult or impossible to attend selectively to a single dimension (Ashby & Maddox, 1994; Ashby & Townsend, 1986; Garner, 1974; Lockhead, 1966; Maddox, 1992; Shepard, 1964). Prototypical separable dimensions are hue and shape, and prototypical integral dimensions are saturation and brightness (for a review of this literature, see, e.g., Maddox, 1992). Therefore, when stimuli are constructed from separable dimensions, unidimensional rules should be fairly simple to form and to implement. With integral dimensions, however, unidimensional rules should be difficult or impossible to form. For this reason, COVIS predicts that, in category learning tasks in which the stimuli are constructed from integral dimensions, the implicit system should dominate the verbal system.

### Procedural Learning in Categorization

As mentioned earlier, procedural memory is the memory of procedures or skills. It is closely related to what some theorists call associative memory (e.g., Oakley, 1981), and it is thought to be distinct from episodic memory (e.g., Tulving, 1985). The prototypical task that is thought to depend critically on procedural learning and memory is the learning and retention of motor skills (e.g., tennis playing). Logan (1988) proposed an influential theory assuming that a skill becomes automatic when "it relies on the retrieval of stored instances" (p. 492). As such, Logan's theory could be interpreted as an instance-based account of tasks that have traditionally been thought to rely on procedural learning and memory. Similarly, COVIS could be viewed as a procedural (and semantic) memory account of a task that has traditionally been thought to rely on instance-based memory (as a result of the popularity of exemplar theory).

Consider a tennis player as his or her opponent initiates a ground stroke. The ball traveling over the net toward the player defines a complex stimulus that is essentially unique on every trial. The player's task is to assign the stimulus to one of a number of categories and then to initiate the motor program associated with that category (e.g., cross court topspin forehand). The process by which a novice learns to become a profi-

cient tennis player is a classic example of procedural learning. Usually the motor component of the tennis playing task is emphasized, but a complex categorization component is also included. We believe that standard perceptual categorization involves a similar procedural learning process. For example, participants in the Maddox and Ashby (1993) experiment saw a unique stimulus on every trial, assigned it to one of a number of categories (i.e., two), and then initiated the motor program associated with that category (by pressing the left or right response key). It is true that the motor component of the standard categorization task is far simpler than the motor component associated with tennis playing, but apart from this complexity difference, the logical structure of the two tasks is similar.

There is now substantial evidence that the striatum, a region in the basal ganglia consisting of the putamen and the caudate nucleus, is necessary for procedural learning (e.g., Jahanshahi, Brown, & Marsden, 1992; Mishkin, Malamut, & Bachevalier, 1984; Saint-Cyr, Taylor, & Lang, 1988; Willingham, Nissen, & Bullemer, 1989). For example, procedural learning is impaired in patients with diseases of the basal ganglia (e.g., Huntington's and Parkinson's disease) but not in cortical diseases, such as Alzheimer's (Gabrieli, 1995; Heindel, Butters, & Salmon, 1983; Heindel, Salmon, Shults, Walicke, & Butters, 1989; Knopman & Nissen, 1991).

Much of the recent work on procedural learning has involved Nissen and Bullemer's (1987) *serial reaction time* (SRT) task, in which participants are required to produce a sequence of button presses. The greatest improvements in reaction time occur when the sequence has some repeating structure, even when participants display no explicit knowledge of the sequential structure (a defining characteristic of procedural learning) and even when they suffer from amnesia (Nissen & Bullemer, 1987; P. J. Reber & Squire, 1994). A number of studies have found evidence for striatal involvement in the SRT task. For example, patients with Huntington's disease and patients with Parkinson's disease both show reduced implicit learning in the SRT task (Jackson, Jackson, Harrison, Henderson, & Kennard, 1995; Knopman & Nissen, 1991; Willingham & Koroshetz, 1993). In addition, a recent neuroimaging study found striatal (i.e., in putamen) activation in an implicit learning condition in which attention was drawn away from the sequence because of a tone-counting secondary task (Grafton, Hazeltine, & Ivry, 1995). These results reinforce the argument that the striatum is a critical component of the procedural learning system.<sup>3</sup>

Within the striatum, the putamen is most closely associated with motor behavior, and the caudate nucleus is most closely associated with cognitive behavior (e.g., Heimer, 1995). Therefore, if the striatum participates in category learning, one's focus should be on the caudate. In fact, there is evidence from nonhuman animal studies that the caudate may mediate perceptual decision processes of the type that should be important in category learning (e.g., McDonald & White, 1993, 1994; Packard, Hirsch, & White, 1989; Packard & McGaugh, 1992; Packard & White, 1991). The most popular of the relevant experimental designs uses a maze of some sort (e.g., radial arm or water maze) and two experimental conditions. In one, a rat must return to a certain spatial location to receive a reward, whereas, in the other, a specific perceptual cue signals reward. For example, Packard and McGaugh (1992) used a water maze with two

floating rubber balls. One was attached to a hidden platform that the animal could mount, thereby escaping the water. The other was attached to a thin anchor that could not be mounted. At the same time, one of the balls was marked with vertical stripes and the other was marked with horizontal stripes. In the spatial condition, the safe ball (i.e., the one attached to the platform) was always in the same spatial location within the maze but was not consistently associated with either the horizontal or vertical stripes. In the visual discrimination condition, the visual pattern on the safe ball was always the same, and the spatial location varied from trial to trial. In all of these studies, rats with hippocampal lesions (i.e., to the fornix) were impaired in the spatial condition but not the visual discrimination condition, whereas rats with caudate lesions were impaired in the visual discrimination condition but not in the spatial condition.

In the spatial condition of the Packard and McGaugh (1992) study, virtually all cues associated with the safe platform were the same on every trial, except for the horizontal and vertical stripes. Thus, the rat could obtain reward by recognizing that the conditions associated with the safe platform were familiar.

<sup>3</sup> On the other hand, it is now clear that there is also a significant cortical component to SRT learning. For example, several neuroimaging studies have reported learning-related changes in motor cortex (Grafton et al., 1995; Pascual-Leon, Grafman, & Hallett, 1994), and Grafton et al. (1995) also found activation in the supplementary motor cortex and the inferior parietal cortex in their implicit learning condition. These findings have caused some researchers to question the significance of the striatal contribution to SRT learning (e.g., Curran, 1995; Keele & Curran, in press). There are a number of reasons to believe, however, that the striatum may play a more significant role in implicit category learning than in SRT learning. First, much learning in the SRT task is explicit. In fact, it has been questioned whether any significant component of SRT learning is implicit (Perruchet & Amorim, 1992; Shanks & St. John, 1994). Second, the SRT task is very different from traditional perceptual categorization, because it requires learning long response sequences. In contrast, perceptual categorization requires a simple motor response but involves complex pattern recognition. It may be that the caudate nucleus is more important for the pattern recognition process than for the chaining required to link a sequence of motor responses. Some evidence supports this hypothesis. In particular, lesions in rat striatum impair visual pattern recognition but have no effect on spatial exemplar-based memory, even when the pattern recognition task and the spatial memory task require the same motor response (Packard, Hirsch, & White, 1989; Packard & White, 1991). Also, patients with Parkinson's disease are impaired in category learning but not in recognition memory (Knowlton, Mangels, & Squire, 1996). Third, evidence from the SRT task suggests that the striatum may be able to learn simple stimulus-response mappings but not higher order associations (see, e.g., Curran, 1995, for a review). For example, A. Cohen, Ivry, and Keele (1990) showed that in the presence of a secondary attention demanding task, which was thought to block explicit learning, SRT learning occurred only when the sequence contained unique pairwise associations. In traditional perceptual categorization tasks, the learning of unique stimulus-response mappings is all that is required to maximize response accuracy (e.g., Ashby, 1992; Ashby & Gott, 1988). This is true even in so-called probabilistic categorization tasks, in which some stimuli have a nonzero probability of belonging to two or more categories (e.g., as in the Maddox & Ashby, 1993, experiments described earlier). Furthermore, the evidence is good that humans do learn deterministic stimulus-response mappings in perceptual categorization tasks (Ashby & Maddox, 1998), although this position is not without controversy (e.g., Estes, 1995).

For this reason, the spatial condition is similar to a recognition memory task. In contrast, in the visual discrimination condition, only one cue remained constant, namely the visual pattern on the safe ball. The rat had to therefore learn to associate this single cue with reward. This condition is similar to a category learning task. In fact, in both the Packard and McGaugh (1992) study and the Maddox and Ashby (1993) circle condition, an important component of the relevant stimuli was line orientation. The major difference between these studies was that Packard and McGaugh's (1992) "safe" category contained only a single exemplar.

Although most studies of this type have used rats, similar results have been obtained with monkeys. For example, in a series of experiments, Gaffan and colleagues successively lesioned all major output pathways from visual association (i.e., inferotemporal) cortex, except for the projection to the caudate nucleus (e.g., projections to the hippocampus and amygdala were lesioned by Gaffan & Harrison, 1987; projections to the prefrontal cortex were lesioned by Eacott & Gaffan, 1992, and Gaffan & Eacott, 1995). None of these lesions affected visual discrimination learning, which prompted Gaffan and Eacott (1995) to conclude that the caudate is a critical component of such learning. Thus, a large number of lesion studies with rats and monkeys all support the hypothesis that the caudate nucleus mediates a form of perceptual discrimination learning that is independent of the instance-based learning and memory thought to be mediated by hippocampal structures.

### The Neuropsychological Version of COVIS

COVIS assumes that there are separate verbal and implicit (i.e., procedural-learning-based) category learning systems and that the striatum is a key structure in the implicit system. Of course, many cortical areas must also be involved. For example, the verbal system must depend heavily on frontal and temporal language areas (and possibly on cingulate cortex). With visual stimuli, the visual cortex is required for representation and early processing, and the posterior parietal cortex is probably required to direct visual attention. How can the striatum, a region in the basal ganglia, coordinate with such a wide variety of cortical areas?

A number of parallel circuits have now been identified, all with the same basic structure. In each case, a pathway has been identified that projects from an area of cortex to striatum, from striatum to the globus pallidus (the output portion of the basal ganglia), from globus pallidus to thalamus, and, finally, from thalamus back to cortex (Alexander, DeLong, & Strick, 1986). It has been proposed that such loops play a major role in motor performance (e.g., Strick, Dum, & Picard, 1995), working memory (e.g., Gabrieli, 1995; Goldman-Rakic, 1995b), selective attention (Miller & Wickens, 1991; Posner & Petersen, 1990), and skill learning (Gabrieli, 1995). There is even evidence that such a loop mediates verbal rules of categorization (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991). We propose that cortical-striatal-pallidal-thalamic loops mediate category learning in functionally separate verbal and implicit systems.

A very rough sketch of the more important neuropsychologi-

cal underpinnings of COVIS is shown in Figure 2 (obviously, many other structures are also involved). The model assumes a hierarchical structure consisting of three primary levels: cortex, thalamus, and the basal ganglia. Communication among levels is mediated by parallel cortical-striatal-pallidal-thalamic circuits. Two such circuits are shown in Figure 2. The posterior circuit, originating in extrastriate visual areas (e.g., inferotemporal cortex [IT]), defines the implicit system. The anterior circuit, which originates in anterior cingulate gyrus and prefrontal cortex, defines the verbal system. Projections from each of these cortical areas (i.e., inferotemporal cortex and other high-level visual areas, cingulate gyrus, and prefrontal cortex) to the caudate nucleus are known to exist (e.g., Selemon & Goldman-Rakic, 1985; Yeterian & Pandya, 1995).

### Implicit System

A high-level visual representation of the stimulus is computed in the inferotemporal cortex and other extrastriate visual areas (e.g., Damasio, 1985; Mishkin, Ungerleider, & Mako, 1983; Tanaka, 1993). These areas are known to project to the tail of the caudate, which, according to COVIS, learns to associate a category response with each stimulus. At this point, there is little in the literature that allows one to speculate on the details of this associative learning. There are two prominent possibilities. First, the striatum might learn a decision bound through some reward-mediated procedural learning process. On each trial, the resulting striatal network would then determine on which side of the bound the stimulus representation lies and subsequently recommend the appropriate response. Second, the striatum might naively learn to associate a category response with different subregions of the perceptual space (Ashby & Maddox, 1989). In this model, the decision bound is an emergent property that separates the subregions assigned to the different category responses.<sup>4</sup> This latter model, which requires little or no computation by the striatum, might be more consistent with known neurophysiology (e.g., Wickens, 1993), but we know of no behavioral data that distinguish between these two possibilities.

Regardless of how the association between stimulus and response is learned (or computed), we assume that, for every categorization judgment, the striatum activates a unit in prefrontal cortex associated with one of the alternative category responses (via a pathway through the globus pallidus and the thalamus) and that the strength of this activation is a measure of the confidence that the implicit system has responded correctly. The category units in prefrontal cortex are assumed to be abstract representations that project via (glutamatergic) cortico-cortical projections to appropriate motor units in premotor or motor cortex. Frith, Friston, Liddle, and Frackowiak (1991a, 1991b; Friston, Frith, Liddle, & Frackowiak, 1991) reported PET scanning data that support this model of prefrontal cortex. In fact, they argued that their data pointed toward a more general model in which prefrontal cortex modulates activity in a variety of remote, but task-relevant, cortical areas (e.g., temporal lan-

<sup>4</sup> It is important to note that this second possibility is not an exemplar-based model, because a subregion of perceptual space is not an exemplar.

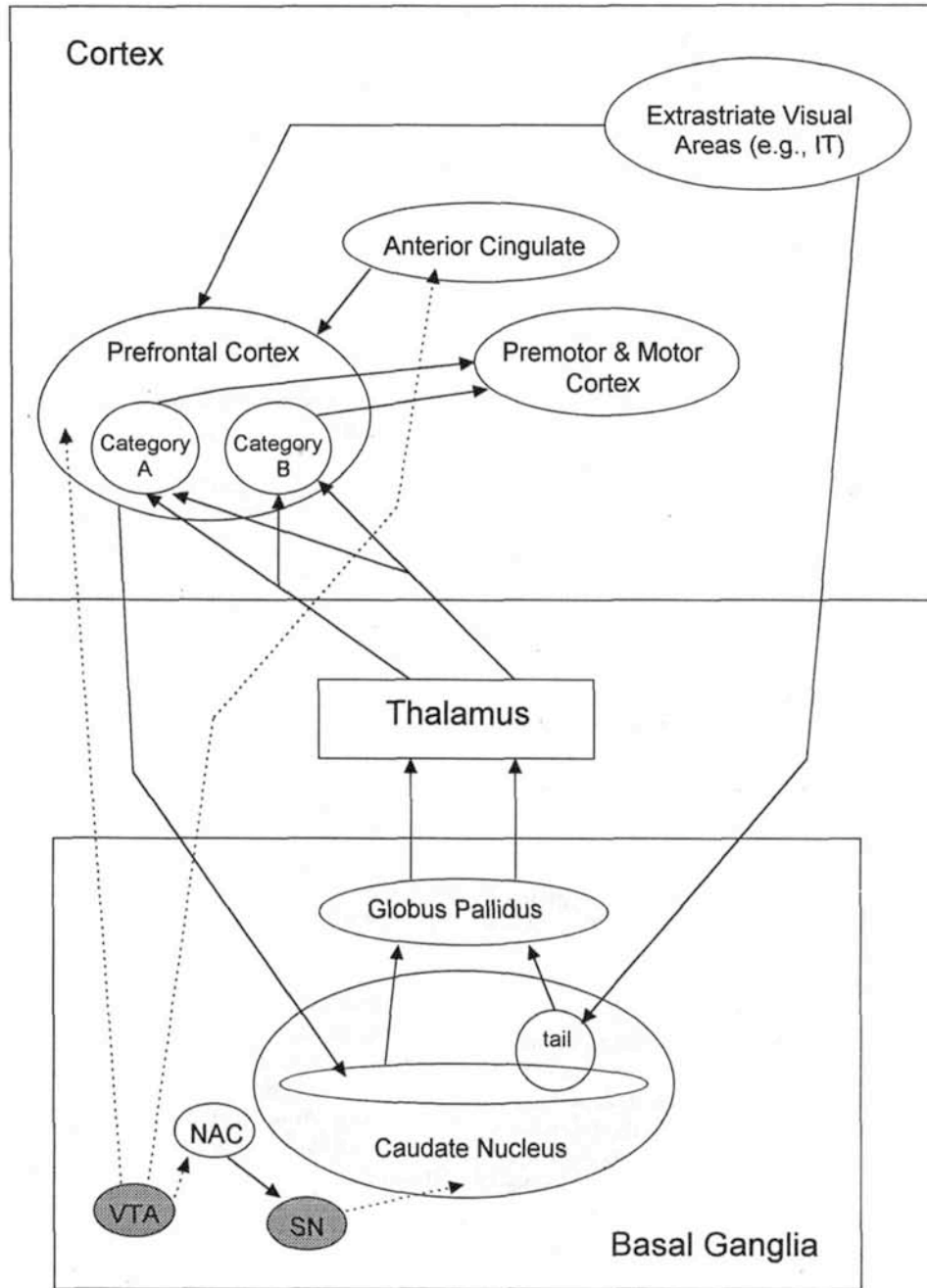


Figure 2. A schematic depicting the neuropsychological underpinnings of COVIS (competition between verbal and implicit systems). The dotted lines denote dopamine projections. VTA = ventral tegmental area; SN = substantia nigra; NAC = nucleus accumbens; IT = inferotemporal cortex.

guage areas) in many tasks requiring “willed action.” One important advantage of this model of prefrontal cortex in the present application is that it allows rapid transfer in categorization tasks in which the motor response is changed, but not the category structure (e.g., rather than button pressing with index fingers, participants are asked to depress a foot pedal).

Categorization by the implicit system can be modeled in a straightforward fashion by decision bound theory (Ashby, 1992; Ashby & Lee, 1991, 1992; Ashby & Maddox, 1990, 1992, 1993;

Ashby & Perrin, 1988; Ashby & Townsend, 1986; Maddox & Ashby, 1993). Let the vector  $\mathbf{x}$  denote the representation of the stimulus as computed, for example, in IT. With the circular stimuli shown in Figure 1,  $\mathbf{x}$  would be the two-dimensional vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$



where  $x_1$  is the perceived size of the circle and  $x_2$  is the perceived orientation of the radial line. A decision bound can be described as the set of all perceived sizes and orientations that satisfy the condition  $h_i(\mathbf{x}) = 0$  for some function  $h_i$ . For example, for the linear bound shown in Figure 1,  $h_i(\mathbf{x}) = x_2 - x_1 = 0$ . If a perceptual representation does not fall exactly on the decision bound, then either  $h_i(\mathbf{x}) > 0$  or  $h_i(\mathbf{x}) < 0$ . In fact, it turns out that all points on the same side of the bound generate the same sign in  $h_i(\mathbf{x})$ . For example, in Figure 1,  $h_i(\mathbf{x}) > 0$  for all points above the bound, and  $h_i(\mathbf{x}) < 0$  for all points below the bound. As a consequence, the decision rule "Respond *A* to all points on one side of the bound and *B* to all points on the other side" is equivalent to the following rule: "Respond *A* if  $h_i(\mathbf{x}) > 0$  and respond *B* if  $h_i(\mathbf{x}) < 0$ ." Because the function  $h_i(\mathbf{x})$  discriminates between stimuli on either side of the decision bound, it is often called a *discriminant function*.

The absolute value of the discriminant function provides a measure of response confidence, because  $|h_i(\mathbf{x})|$  tends to increase with the distance between  $\mathbf{x}$  and the decision bound. Stimuli near the bound could belong to either category, and so response confidence is low, and  $|h_i(\mathbf{x})|$  is small. Stimuli far from the bound obviously belong to only one of the two categories, and so confidence is high, and  $|h_i(\mathbf{x})|$  is large. COVIS assumes that the output of the striatum is a discriminant value, but, for now at least, we make no assumptions about how the discriminant value is computed (i.e., we do not choose between either of the two alternatives just described).

According to this model, animals with caudate lesions should be impaired in visual discrimination tasks, not because of any visual deficits but because of impaired decision processes. As mentioned previously, many studies support this prediction (e.g., Eacott & Gaffan, 1992; Gaffan & Eacott, 1995; Gaffan & Harrison, 1987; McDonald & White, 1993, 1994; Packard et al., 1989; Packard & McGaugh, 1992; Packard & White, 1991). For example, in one study, rats with caudate lesions failed to learn an orientation discrimination (vertical lines vs. horizontal lines; Packard & McGaugh, 1992), and, in another, rats with a similar lesion failed to learn a flicker-rate discrimination (high flicker rate vs. low flicker rate; Reading, Dunnett, & Robbins, 1991). In both cases, the stimuli varied on a single dimension, so the optimal decision rule was of the form "Respond *A* if  $x_1 + c > 0$ , and respond *B* if  $x_1 + c < 0$ ," where  $x_1$  is either perceived orientation or perceived flicker rate (depending on the application). Because the animals had intact visual systems, COVIS predicts that their perception of orientation and flicker rate was normal and that their deficits occurred because of an inability to learn or implement this discriminant function.<sup>5</sup>

There are also projections from auditory cortex to caudate and from somatosensory cortex to caudate (e.g., Arnauld, Jeantet, Arsaut, & Demotes-Mainard, 1996; Chudler, Sugiyama, & Dong, 1995), so it is possible that implicit category learning of the type we are postulating for visual stimuli might also occur for auditory and haptic stimuli. At this point, however, much less is known about these projections, and there is also a corresponding dearth of behavioral data on auditory and haptic category learning. Therefore, although the generalization to other sensory modalities appears promising, at this point COVIS should be considered a theory of category learning with visual stimuli.

### Verbal System

Before the verbal system can recommend a categorization response, it must select among the alternative types of verbal rules and access the appropriate visual and semantic information. Among these tasks, perhaps the most critical is selection of the appropriate verbal rule type. For example, in experiments in which two alternative categories are composed of the Figure 1 circles, there are two especially simple types of verbal rules. One is to set a criterion on size, and the other is to set a criterion on orientation. The first type of verbal rule is of the form "Respond *A* if the circle is large and *B* if it is small." A verbal rule of the second type is "Respond *A* if the radial line is tilted up and *B* if it is tilted down." In a task in which the optimal rule is of the first type, a participant who selects a verbal rule of the second type will perform poorly.

Posner and Petersen (1990) postulated that a similar kind of selection process is carried out by an anterior attentional network that includes the anterior cingulate and prefrontal cortices. Their arguments were based partly on neuroimaging studies suggesting that the cingulate may be heavily involved in rule selection. In these studies, cingulate activation is found in tasks in which the participant must select among various interpretations or aspects of the stimulus. In conditions in which such selection is not required, the cingulate is not activated. For example, Corbetta et al. (1991) found cingulate activation in a same-different task in which two visual stimuli could differ in any one of three components but not when the stimuli could differ only in one component. The former condition requires selection of the appropriate dimension, whereas the latter condition does not. Similarly, the cingulate is not activated when participants are simply required to read a stimulus word, but it is activated when the word is a color name printed in the ink of a different color (i.e., the classic Stroop task; Bench et al., 1993) or when participants are required to name a verb related to the stimulus word (Peterson, Fox, Posner, Mintan, & Raichle, 1988). Reading a word does not require one to select a meaning, but naming a related verb does, and Stroop tasks require one to select between conflicting semantic and perceptual cues.

Following Posner and Petersen (1990), we hypothesize that the anterior cingulate selects among the alternative types of verbal rules in categorization tasks. Specifically, we hypothesize that a cingulate-prefrontal cortex network is responsible for learning which type of verbal rule is most effective in a particular task and for selecting one for use on each trial.

In an experiment in which the stimuli are the Figure 1 circles, let  $x_1$  denote perceived size and let  $x_2$  denote perceived orientation. Then the verbal rules on size and orientation are formally equivalent to "Respond *A* if  $h_v(\mathbf{x}) = x_1 + c_1 > 0$  and *B* if  $h_v(\mathbf{x}) = x_1 + c_1 < 0$ " and "Respond *A* if  $h_v(\mathbf{x}) = x_2 + c_2 > 0$  and *B* if  $h_v(\mathbf{x}) = x_2 + c_2 < 0$ ," respectively, for some constants  $c_1$  and  $c_2$ . In addition to learning which type of verbal rule is

<sup>5</sup> COVIS predicts that, in humans, such unidimensional rules would be learned more quickly by the verbal system. Even so, the implicit system in COVIS could also learn the unidimensional rules used in these rat studies. Because, presumably, category learning in rats is dominated by the implicit system, the only way rats can learn these unidimensional rules is via their analogue of the implicit system.

most accurate, the verbal system must also learn the optimal values of the response criteria,  $c_1$  and  $c_2$ . There is little in the literature to suggest where such learning might occur. Perhaps most relevant are the animal studies discussed earlier in which lesions of the caudate nucleus disrupted the learning of a simple (unidimensional) visual discrimination (e.g., Packard & McGaugh, 1992; Reading et al., 1991). Because of these data, we tentatively assume that criterion learning in the verbal system is also mediated by the striatum. This is not a critical assumption of COVIS. For example, none of the COVIS predictions derived and tested in this article depend in any way on this assumption.

### *Selection and Switching in the Verbal System*

The Wisconsin Card Sorting Test (Berg, 1948) is a widely used neuropsychological instrument in which participants complete a series of categorization tasks (e.g., Kolb & Whishaw, 1990). Stimuli are geometric patterns that vary in color, shape, and symbol numerosity, and in all cases the correct categorization rule is easily verbalized (i.e., the correct rule is always unidimensional). Owen et al. (1993) argued that perseverative errors on the Wisconsin Card Sorting Test could occur for two different reasons. One is a failure to select the appropriate verbal rule, and the other is a failure to switch attention from an inappropriate to an appropriate rule. To test the hypothesis that two different kinds of errors can be made on the Wisconsin Card Sorting Test, Owen et al. (1993) designed two new tasks in which the different errors would each be observable. Using this clever design, Owen et al. (1993) found that patients with lesions of the prefrontal cortex were impaired in the switching task but not in the selection task, whereas (unmedicated) Parkinson's patients were impaired in both tasks. These results support the hypothesis that selection and switching are separate operations.

More recently, Ashby, Isen, and Turken (1998) argued that the anterior cingulate mediates the selection operation and the basal ganglia mediate the switching operation via cortical–striatal–pallidal–thalamic loops. Much of the evidence implicating the anterior cingulate in the selection operation was reviewed earlier when we introduced the Posner and Petersen (1990) anterior attentional network. In addition, the Ashby et al. hypothesis is consistent with the Owen et al. (1993) data because frontal patients presumably have no anterior cingulate damage, but Parkinson's patients do (this is discussed in more detail later).

The evidence that the basal ganglia mediate the switching operation is as follows. First, injections of a glutamate agonist directly into the striatum increase the frequency with which cats switch from one motor activity to another in a task in which food rewards are delivered for such switching behaviors (Jaspers, de Vries, & Cools, 1990a, 1990b). Second, lesioning the dopamine fibers that project from the ventral tegmental area (VTA) into the prefrontal cortex actually improves the performance of monkeys in an analogue of the Wisconsin Card Sorting Test (Roberts et al., 1994). If switching occurs in the prefrontal cortex, then such lesions should impair switching performance. In fact, this result is strong evidence that the basal ganglia mediate switching. This is because such lesions tend to increase dopamine levels in the basal ganglia<sup>6</sup> (Roberts et al., 1994). If the basal

ganglia are responsible for switching and switching is enhanced by dopamine, and if lesioning the dopamine fibers that enter the prefrontal cortex increases dopamine levels in the basal ganglia, then lesioning dopamine fibers in prefrontal cortex should improve switching. Cortical–striatal–pallidal–thalamic loops provide a perfect mechanism through which the striatum could contribute to rule-switching behavior. For example, separate loops could be established for each alternative rule type from prefrontal cortex down to the caudate and back up to prefrontal cortex. The caudate could participate in rule switching by inhibiting all but one of these loops.

If this hypothesis is correct, then, according to COVIS, it should be difficult to find special populations that are impaired in category learning tasks in which the optimal rule is nonverbal but not in tasks in which the optimal rule is verbal.<sup>7</sup> This is because the caudate is a critical structure for both systems. As such, any population with widespread striatal dysfunction should be impaired in all types of category learning tasks. On the other hand, populations with dysfunction limited to frontal areas (i.e., either the prefrontal cortex or the anterior cingulate) should be impaired in tasks in which the optimal rule is verbal but not in tasks in which the optimal rule is nonverbal. As shown shortly, the neuropsychological literature generally supports this predicted dissociation.

There are several major dopamine pathways in the mammalian brain (Beatty, 1995). In the nigrostriatal system, dopamine-producing cells in the substantia nigra project into striatum. In the mesocorticolimbic system, dopamine-producing cells in the VTA project to a number of different areas (e.g., Scheel-Krüger & Wilner, 1991), but with respect to COVIS, the most important projections are into prefrontal cortex and anterior cingulate (i.e., see Figure 2). Ashby et al. (1998) also hypothesized that the dopamine projections from VTA into anterior cingulate facilitate rule selection. Evidence supporting this hypothesis can be found in the neuropsychological literature. First, dopamine antagonists (i.e., haloperidol) increase the number of perseverative errors in a simplified version of the Wisconsin Card Sorting Test (Berger et al., 1989). Second, amphetamines, which stimulate dopamine release, increase the number of alternation responses in a two-choice guessing task (Ridley, Baker, Frith, Dowdy, & Crow, 1988). Third, schizophrenics with negative symptoms (e.g., flattened affect and motivational difficulties), who are thought to have decreased dopamine levels in frontal areas, perseverate in the Wisconsin Card Sorting Test (e.g., Malmo, 1974). Fourth, schizophrenics with positive symptoms (e.g., delusions and hallucinations), who are thought to have increased dopamine levels in frontal areas (e.g., Crow, 1980, 1982), have difficulty maintaining cognitive set (e.g., see the *Diagnostic and Statistical Manual of Mental Disorders* [4th

<sup>6</sup> Dopamine levels in basal ganglia are apparently increased because the prefrontal cortex tonically inhibits the VTA. Lesioning the dopamine fibers into prefrontal cortex releases this inhibition, which effectively stimulates the VTA. There are a number of scenarios in which increased VTA activation could lead to increased dopamine levels in the striatum. Perhaps the most likely is the VTA–nucleus accumbens–substantia nigra–striatum pathway.

<sup>7</sup> According to COVIS, the most likely exception would be a patient group with striatal damage restricted to the tail of the caudate.

edition; *DSM-IV*; American Psychiatric Association, 1994]). Fifth, Parkinson's patients show difficulty in rule switching (e.g., R. G. Brown & Marsden, 1988; Cools, van den Bercken, Horstink, van Spaendonck, & Berger, 1984). Although the most widely known effect of Parkinson's disease is damage to the nigrostriatal system, there is also concomitant damage to the mesocorticolimbic system. In particular, dopamine production in VTA is substantially diminished as the disease progresses (Javoy-Agid & Agid, 1980; Scatton, Rouquier, Javoy-Agid, & Agid, 1982). Finally, positive affect, which is thought to increase dopamine levels (Ashby et al., 1998), has been shown to facilitate cognitive set switching (i.e., to reduce anchoring) in medical diagnoses (Estrada, Isen, & Young, 1997).

For these reasons, COVIS predicts that any condition that decreases dopamine levels in anterior cingulate below normal levels should cause a selective impairment in category learning tasks in which the optimal rule is verbal (e.g., as in Parkinson's disease). Slight increases in dopamine, such as might occur during periods of mild transient happiness, could improve performance in verbal conditions. This latter prediction, however, must be considered tentative because there are many reasons, unrelated to those discussed here, why it could fail (see Ashby et al., 1998, for more details). Finally, large increases in frontal dopamine levels, such as might occur with positive schizophrenia or under the influence of certain drugs (e.g., cocaine), should also impair performance in verbal categorization tasks, because the excess dopamine should increase the probability that inappropriate verbal rules are selected.

### *Competition Among the Systems, Learning, and the Role of Dopamine*

The output of the verbal and implicit striatal units is a discriminant value. In information-processing terms, the sign of this value signals whether the category response should be *A* or *B*, and the magnitude of the discriminant value increases with response confidence. Thus, COVIS assumes that a stimulus on or near the category decision bound elicits a weak striatal response, because it does not strongly favor either category. However, a stimulus far from the bound elicits a strong response, because it strongly favors one category over the other.

Next, the verbal and implicit units must compete with each other, with the result that the system producing the strongest response wins out. There is little in the literature on how such competition might be resolved. In principle, it could occur at many different levels (e.g., striatal, pallidal, thalamic, and cortical). With respect to COVIS, however, the only important detail is that the competition is resolved somewhere (and somehow). One intriguing possibility is that the competition between the two systems is resolved via lateral inhibition within the striatum. Competition of this type is thought to be common in the striatum<sup>8</sup> (i.e., mediated by the acetylcholine spiny matrix; see, e.g., Alexander & Wickens, 1993; Miller & Wickens, 1991). There are also suggestions that dopamine may enhance this inhibition, perhaps through activation of D<sub>2</sub> receptors on cholinergic interneurons (e.g., Miller & Wickens, 1991; Wickens, 1993; Wickens, Alexander, & Miller, 1991). In Parkinson's disease, the dopamine-producing cells in the substantia nigra die, and, as a result, the striatum loses its dopamine input. Loss of dopamine

in the striatum is thought to produce muscle rigidity, which is characteristic of Parkinson's disease, in the following way (Hayashi et al., 1988; Wickens, 1993). Separate units in the striatum (i.e., in the putamen) are associated with different muscle groups. When a movement is initiated, the unit associated with the appropriate muscle group inhibits opposing muscle groups, and fluid movement results. In Parkinson's disease, the loss of dopamine greatly reduces such inhibition, and, as a result, opposing muscle groups simultaneously fire, which produces the muscle rigidity that is characteristic of the disease. It would be interesting if the competition between verbal and implicit categorization systems is resolved in a similar fashion.

Another possible means of resolving the competition was suggested by Berns and Sejnowski (1996), who argued that the anatomy of the output portion of the basal ganglia is ideally suited for selecting among a broad class of competing cognitive, motor, and sensory states. In the Berns and Sejnowski model, the competition is resolved in the internal segment of the globus pallidus (i.e., the GP<sub>i</sub>, which is the output structure of the basal ganglia), and a critical role is played by the so-called indirect pathway, which passes from the striatum to the GP<sub>i</sub> via the external segment of the globus pallidus and the subthalamic nucleus.

As learning progresses, COVIS assumes that the more successful of the two systems begins to dominate the other. For example, in a task in which the optimal rule is nonverbal, the model predicts that people will initially tend to use verbal rules and then gradually rely more heavily on nonverbal rules. Even highly experienced classifiers, however, are predicted to use verbal rules on some trials.

Within the striatum, the verbal system must learn the appropriate response criteria, and the implicit system must learn the appropriate decision bound. Because such learning must occur quickly (i.e., over the course of an hour or 2), neural plasticity is required. The striatum is an excellent candidate for such learning because it contains a large number of NMDA binding sites (Monaghan & Cotman, 1985), which have been implicated in the Hebbianlike learning associated with long-term potentiation (Bliss & Lømo, 1973). In addition, learning in the striatum is facilitated by a dopamine projection from the substantia nigra (see Figure 2). Dopamine may facilitate category learning in two different ways. First, if the verbal and implicit units inhibit each other in the striatum and dopamine enhances this inhibition, then this should allow the more accurate of the systems to quickly dominate the less accurate. Second, dopamine input to

<sup>8</sup> This assumption seems to require that regions of prefrontal cortex and IT project to nearby areas in the caudate. As already mentioned, extrastriate visual areas project most heavily to the tail of the caudate (e.g., Yeterian & Pandya, 1995). In contrast, the prefrontal cortex apparently projects to a strip that runs longitudinally from the head of the caudate to the tail (Selemon & Goldman-Rakic, 1985). Thus, the caudate projections of prefrontal cortex and extrastriate visual areas overlap. In particular, both areas, together with the anterior cingulate, project to the ventrocaudal striatum, which includes the tail of the caudate and the adjacent part of the ventral putamen (e.g., Brown, Desimone, & Mishkin, 1995). Because this region in striatum has been implicated in visual discrimination learning, it is a possible candidate for the hypothesized inhibiting effects.

the striatum is thought to function as the reward signal in reward-mediated learning (e.g., Beninger, 1983; Wise, 1982; Wise & Rompré, 1989). In one plausible scenario (e.g., Wickens, 1993), glutamate projections from cortex and dopamine projections from the substantia nigra both synapse on the dendritic spines of striatal neurons (DiFiglia, Pasik, & Pasik, 1978; Freund, Powell, & Smith, 1984). A cortical signal causes an influx of free  $\text{Ca}^{2+}$  into the spines (e.g., through NMDA receptors), which initiates a cascade of reactions that modify the synapse (Lynch, Larson, Kelso, Barrionuero, & Schottler, 1983; Wickens, 1993). Because the spines are somewhat separated from the bulk of the intracellular medium, free  $\text{Ca}^{2+}$  persists until the arrival of the dopamine reward signal (Gamble & Koch, 1987; MacDermott, Mayer, Westbrook, Smith, & Barker, 1986). Dopamine (acting through  $\text{D}_1$  receptors) leads to the inhibition of a process that reverses the calcium-activated synaptic modification (Hemmings, Walaas, Ouimet, & Greengard, 1987; Pessin et al., 1994; Wickens, 1990, 1993). Thus, the presence of dopamine strengthens the synapses that were active on a trial when reward was delivered (e.g., Huang & Kandel, 1995).

The separation of function in visual cortex and striatum is an important advantage of a corticostriatal categorization system. The perceptual representation is computed in the visual cortex, and the categorization decision is computed in the striatum. In this way, learning a rule for separating two arbitrary categories of the type encountered in a typical psychological experiment does not fundamentally alter the way a person perceives the world, as it might if representation and decision were computed in the same neural tissue. In addition, such separation allows perceptual learning to occur independently of what is typically called category (i.e., decision) learning. The striatum receives a dopamine projection, whereas the visual cortex apparently does not (e.g., Lidow, Goldman-Rakic, Gallager, & Rakic, 1991). Therefore, according to COVIS, category learning is facilitated by a reward signal, whereas perceptual learning is not. As such, although categorization may affect perceptual learning (e.g., Goldstone, 1994; Schyns & Rodet, 1997), we predict that these two types of learning are qualitatively different. For example, category learning should be severely disrupted by the withdrawal of feedback, whereas at least some forms of perceptual learning should be unaffected by a feedback signal.

It is important to note that the neuropsychological version of COVIS is a theory of category learning. It specifically is not a theory of categorization behavior in experts. It seems quite likely that the development of expertise is a process by which the subcortical pathway through the basal ganglia is gradually replaced by a more direct path from perceptual to motor areas. One intriguing possibility is that cortico-cortical projections develop from extrastriate visual areas to the relevant motor units in premotor or motor cortex (or perhaps to the abstract category units in prefrontal cortex). Such projections, which are prominent in primate cortex (Heimer, 1995), would have a distinct speed advantage over the longer pathway through the striatum, globus pallidus, and thalamus postulated by COVIS. Why, then, is the longer, subcortical pathway needed at all? One obvious possibility is that because of the prominent dopamine projection into the striatum, the subcortical pathway enjoys a higher level of neural plasticity than the cortical pathway. The idea here

is that the dopamine projection into the striatum allows the subcortical pathway to learn new category structures quickly. During this time, the cortical pathway is slowly learning the same thing, perhaps because the subcortical pathway is activating the appropriate cortical motor units. Eventually, the subcortical pathway is no longer needed, and expertise has developed.

This model of expertise is speculative and is not considered further here. However, before moving on, we briefly mention three pieces of evidence that support a model of this type. First, several studies have reported learning-related changes in motor cortex, which are consistent with this model, in more traditional procedural learning tasks (i.e., the SRT task; e.g., Grafton et al., 1995; Pascual-Leon, Grafman, & Hallett, 1994). Second, neuropsychological patients with damage to the striatum (e.g., those with Parkinson's or Huntington's disease) are not impaired in terms of assigning objects to well-learned categories (e.g., they can still determine whether a novel object is a chair or table). Third, it is widely known that Parkinson's patients can execute some highly complex, well-learned motor behaviors flawlessly, even though they have extreme difficulty with novel, willed actions (e.g., Langston & Palfreman, 1995). For example, one of us is familiar with the case of a Parkinson's patient who played tennis throughout his life. After the disease developed, he experienced the prototypical Parkinson's movement difficulties, except when playing tennis. Within the current model, the subcortical pathway, which is damaged by Parkinson's disease, is necessary for learning and executing novel actions. Highly learned actions, however, can be executed via the cortico-cortical projections, which are relatively unaffected by the disease.

### Neuropsychological Tests of COVIS

Several recent neuroimaging studies provide general support for the neuropsychological version of COVIS. Seger, Poldrack, Prabhakaran, and Gabrieli (1997) performed functional magnetic resonance imaging (fMRI) on participants as they were learning two categories created from low-level distortions of two texture patterns. A verbal rule allowed an accuracy level of about 95% correct, so COVIS predicts that the verbal system will be active throughout category learning. In support of this prediction, Seger et al. (1997) reported that categorization elicited significantly more activation in the anterior cingulate than did a baseline task. Poldrack et al. (1997) reported the results of a similar study in which each of 14 highly distinct visual patterns was probabilistically associated with each of two categories in such a way that the optimal rule was nonverbal. In this case, COVIS predicts that the implicit system will eventually dominate and that the activity of the verbal system will decrease with experience. In support of these predictions, Poldrack et al. (1997) found that categorization was associated with significantly more activation in the caudate nucleus than the baseline task and that there were learning-related decreases in activation in the (left) dorsolateral prefrontal cortex (i.e., a prominent structure in the COVIS verbal system). These results represent promising preliminary support for the COVIS network illustrated in Figure 2. With the recent dramatic proliferation of neuroimaging studies, other direct tests of COVIS are likely.

Another less direct method of testing COVIS is to examine category learning in certain special neuropsychological popula-

tions. COVIS makes specific predictions about the categorization abilities of a number of such special populations. In general, these predictions have not been completely tested. Even so, in many cases there are results that partially test the COVIS predictions. In this section, we develop some of these predictions and review the available evidence.

### *Parkinson's Disease*

Parkinson's disease causes the death of dopamine-producing cells, most notably in the substantia nigra but also in the VTA (e.g., Kolb & Whishaw, 1990). As a result, patients with Parkinson's disease have decreased dopamine levels in the substantia nigra and VTA projection sites. This includes a number of structures that play a critical role in COVIS, including the striatum, the prefrontal cortex, and the anterior cingulate. COVIS makes the following predictions about the effects of decreased dopamine levels on category learning.

1. Decreased dopamine levels in the tail of the caudate should cause impaired category learning in tasks in which the optimal rule is nonverbal.

2. Decreased dopamine levels throughout the caudate should cause deficits in tasks that require switching executive attention. This includes many verbal category learning tasks.

3. Decreased dopamine levels in frontal areas (i.e., prefrontal cortex and anterior cingulate) should cause deficits in rule selection (Ashby et al., 1998) and in working memory (e.g., Goldman-Rakic, 1995a; Roberts et al., 1994), which are important components of explicit reasoning. Such deficits should impair category learning in most verbal rule tasks.

In Parkinson's disease, there are large individual differences in the amount of damage sustained by the tail of the caudate, the head of the caudate, and frontal cortical areas (e.g., van Domburg & ten Donkelaar, 1991). For example, some patients have very little cortical damage until late in the disease. Some Parkinson's patients, however, will have significant damage in each of these areas, so COVIS predicts that within the population of Parkinson's patients, there will be widespread deficits in all kinds of category learning tasks. At the individual patient level, however, significant individual differences should be expected. Within the striatum, there is evidence that damage in the head of the caudate is usually more severe than in the tail (van Domburg & ten Donkelaar, 1991). As a consequence, COVIS predicts that although Parkinson's patients, as a group, should be impaired in verbal as well as nonverbal category learning, deficits in verbal tasks should be more common or severe, or both.

When testing these predictions, it is important to consider the effects of medication (e.g., L-DOPA), especially because there is a possibility that standard dopamine replacement therapies correct striatal dysfunction more successfully than frontal dysfunction. In particular, the well-known improvements in motor activity associated with medication indicate at least some reduction in striatal dysfunction. On the other hand, a recent large meta-analytic study found that the standard Parkinson's medications produced little, if any, reduction in the cognitive impairments associated with the frontal dysfunction common to the disease (Delis & Massman, 1992). On the basis of these results,

COVIS predicts that medication might improve Parkinson's patients more in nonverbal than in verbal category learning tasks.

The prediction that Parkinson's patients should be impaired in category learning tasks in which the optimal rule is nonverbal was supported recently by Knowlton, Mangels, and Squire (1996). In their study, stimuli varied on four binary-valued dimensions. Each of 14 different stimuli was associated probabilistically with two categories in such a way that no simple verbal rule was optimal. All participants completed 150 trials of category learning, and feedback was provided after every response. Medicated Parkinson's patients were impaired in this task relative to age-matched controls. In addition to their overall poorer performance, there was a strong negative correlation ( $r = -.55$ ) between severity of the Parkinsonism and performance on the category learning task; that is, patients with more severe symptoms of Parkinson's disease (as measured by the Hoehn and Yahr Scale; Hoehn & Yahr, 1967) performed substantially worse in the category learning task.

There is a substantial literature showing that Parkinson's patients are impaired in category learning tasks in which the optimal rule is easily verbalized. For example, it is well known that Parkinson's patients are impaired in terms of their performance on the Wisconsin Card Sorting Test (Brown & Marsden, 1988; Cools et al., 1984).

### *Elderly People*

During the course of normal aging, dopamine levels in the human brain decrease by 7% or 8% during each decade of life (e.g., Gabrieli, 1995; van Domburg & ten Donkelaar, 1991). It is thought that there are no obvious behavioral effects of this decrease until overall dopamine levels are reduced by about 70%. At this point, elderly people begin to experience motor and cognitive deficits similar to those seen in the early stages of Parkinson's disease. In fact, it has been hypothesized that Parkinson's disease occurs when the normal dopamine depletion process is accelerated for one reason or another (e.g., Calne & Langston, 1983). As a consequence, COVIS predicts that healthy elderly people should show categorization deficits of the same type, albeit not as severe, as patients with mild cases of Parkinson's disease.

There is abundant evidence that elderly people are less adept at rule selection than normal younger adults. For example, elderly individuals show more perseverative errors on the Wisconsin Card Sorting Test than young adults (e.g., Collins & Tellier, 1994; Kramer, Humphrey, Larish, & Logan, 1994; Parkin & Lawrence, 1994). On the other hand, we know of no rigorous tests of the hypothesis that elderly people are impaired, relative to young adults, on categorization tasks in which the optimal rule is nonverbal. Even so, a pilot study using the randomization technique found that elderly people were unable to learn a quadratic decision bound of the type readily learned by younger adults (W. T. Maddox & J. V. Filoteo, personal communication, November 12, 1995).

### *Huntington's Disease*

Huntington's disease attacks and kills the output cells in the caudate nucleus of the striatum (e.g., Kolb & Whishaw, 1990).

According to COVIS, these are the cells that signal the response of the implicit system, so Huntington's patients should be impaired in category learning tasks in which the optimal rule is nonverbal. On the other hand, the dopamine input to the striatum is intact in Huntington's patients, so for those output cells not affected by the disease, learning should be normal. Thus, although Huntington's and Parkinson's patients both suffer from diseases of the basal ganglia, COVIS predicts a subtle difference in their category learning deficits. In Parkinson's disease, there should be partial but impaired learning in all striatal output cells, whereas, in Huntington's disease, striatal output cells should be reduced in number but show normal learning.<sup>9</sup> Without a more detailed model of exactly how the striatum associates category responses and percepts, however, it is difficult to predict how such a difference might manifest itself in behavioral data.

The COVIS prediction that Huntington's patients should be impaired in category learning tasks in which the optimal rule is nonverbal was supported recently by Knowlton, Squire, et al. (1996). This study used the same task as the Knowlton, Mangels, and Squire (1996) Parkinson's experiment. In particular, each of 14 different stimuli varying on four binary-valued dimensions was probabilistically associated with two categories in such a way that no simple verbal rule was optimal. All participants completed 150 trials of category learning, and feedback was provided after every response. Patients with Huntington's disease were significantly impaired in this task relative to age-matched controls. In fact, according to Knowlton, Squire, et al. (1996), "the patients with HD [Huntington's disease] exhibited no demonstrable learning across 150 trials of training" (p. 541).

On the other hand, Knowlton, Squire, et al. (1996) reported that the same Huntington's patients showed no deficits in artificial grammar learning. In this task, a set of 23 grammatical letter strings was generated from a finite-state artificial grammar rule system. Participants studied these training items and then were shown 46 new letter strings, one at a time. Half of these strings were generated from the same artificial grammar, and half were not. The participants' task was to determine whether each new string was grammatical or nongrammatical. The Huntington's patients were not significantly worse than the age-matched controls during this testing phase.

Technically, artificial grammar learning is a categorization task because there are two categories (grammatical and nongrammatical), and each contains many exemplars. Also, there is clearly no simple verbal rule that is optimal. As one might therefore expect, people can respond at above-chance accuracy levels without explicit awareness of the optimal rule (e.g., Knowlton & Squire, 1994, 1996; Reber, 1976). Indeed, the artificial grammar learning literature has discussed in considerable detail many of the issues considered in this article (e.g., implicit vs. explicit learning, the roles of consciousness and verbalization, and whether implicit learning is instance based; e.g., Durlan, Carlson, & Dewey, 1985; Mathews et al., 1989; Reber, 1993; Vokey & Brooks, 1992; Whittlesea & Dorken, 1993). On the other hand, artificial grammar learning differs in several important ways from the more traditional category learning task in which the Huntington's patients were impaired. In particular, in the Knowlton, Squire, et al. (1996) artificial grammar task, the letter strings used during training and test were different. Thus, any associative learning between specific stimuli and re-

sponses that occurred during training could not be used directly during testing. In addition, the relevant categories were not perceptual, in the sense that category membership was not defined by perceptual similarity. If the striatum learns to associate responses with regions in perceptual space, then it could not participate in artificial grammar learning, because there was no relation between location of the test stimuli in the perceptual space and category membership.

Amnesic patients also perform normally in artificial grammar learning (Knowlton, Ramus, & Squire, 1992; Knowlton & Squire, 1994), so Knowlton, Squire, et al. (1996) concluded that "it appears that this type of learning is independent of both the medial temporal lobe and diencephalic structures damaged in amnesia as well as the basal ganglia" (p. 544). Knowlton, Squire, et al. (1996) argued that success in artificial grammar learning depends, at least partly, on the perceptual priming of specific letter bigrams and trigrams. Apparently, then, a complete theory of category learning also may need to include a mechanism that learns through perceptual priming.

Unlike in Parkinson's disease, the prefrontal cortex and cingulate gyrus are not directly affected in Huntington's disease (at least not in the early stages). Thus, Huntington's disease should provide a test of whether the striatum contributes to rule switching. If it does, then Huntington's patients should be impaired in category learning tasks in which the optimal rule is verbal. However, if rule selection and switching are performed exclusively within the cingulate-prefrontal cortex, then Huntington's patients should not be significantly impaired in categorization tasks in which the optimal rule is verbal (although they may be impaired in learning the optimal value of the criterion). This issue is far from resolved. Huntington's patients do perseverate in rule-switching tasks such as the Wisconsin Card Sorting Test, but there are reports that this is true only during later stages of the disease (Josiassen, Curry, & Mancall, 1983), when cortical impairment is likely. Specifically, Josiassen et al. (1983) found that patients in the early stages of Huntington's disease did not make any more perseverative errors on the Wisconsin Card Sorting Test than a control group.

In general, however, Huntington's patients may not provide strong tests of COVIS predictions because of the concomitant dementia that is a direct consequence of the disease. According to the *DSM-IV* (American Psychiatric Association, 1994), "the onset of Huntington's disease is often heralded by insidious changes in behavior and personality, including depression, irritability, and anxiety" (p. 149).

#### *Patients With Lesions in the Prefrontal Cortex*

According to COVIS, patients with lesions restricted to the prefrontal cortex (e.g., from stroke) should have an impaired

<sup>9</sup> Early in the disease, the greatest damage is to the striatal output cells that project to the external segment of the globus pallidus (e.g., Strange, 1992). This is the first stage in the so-called indirect pathway that continues through the subthalamic nucleus and into the GP<sub>1</sub>. The direct pathway is defined by cells that project directly from the striatum to the GP<sub>1</sub>. As mentioned earlier, Berns and Sejnowski (1996) postulated that the indirect pathway is vital for resolving competition among competing cognitive, motor, and sensory states. If this hypothesis is correct, then Huntington's patients should have difficulty resolving the competition between the verbal and implicit category learning systems.

working memory and an impaired ability to switch from one verbal rule type to another. Although we tentatively hypothesized that the basal ganglia initiates switching, it should be recalled that the hypothesized switching was between cortical–striatal loops. As such, by disrupting these loops, cortical (i.e., prefrontal cortex) lesions should impair switching. Therefore, frontal patients should be impaired in category learning tasks in which the optimal rule is easily verbalized. On the other hand, because the implicit system is assumed to not involve frontal areas, such patients should not be impaired in tasks in which the optimal rule is not easily verbalized. A huge amount of data supports the first prediction. In fact, perseverative responding on the Wisconsin Card Sorting Test is perhaps the classic symptom of a frontal lesion (e.g., Kolb & Whishaw, 1990; Robinson, Heaton, Lehman, & Stilson, 1980). There is considerably less known about the performance of frontal patients in nonverbal category learning tasks. Even so, Knowlton, Mangels, and Squire (1996) reported that frontal patients performed normally in the same nonverbal category learning task in which Parkinson's patients were impaired.

### *Amnesia*

Amnesia is another neuropsychological condition in which category learning has been studied. In most of these studies, the focus has been on medial temporal lobe amnesia, but Korsakoff's patients are sometimes also included. Still another strong prediction of COVIS is that neuropsychological patients with damage to structures not shown in Figure 2 should show relatively normal category learning. Patients with medial temporal lobe amnesia have damage to hippocampal structures (sometimes including the parahippocampal gyrus), but in classical cases, they have intact basal ganglia and frontal cortices. As a result, COVIS predicts that amnesic patients should show relatively normal category learning. This prediction has been supported in a number of different studies (Knowlton et al., 1992; Knowlton & Squire, 1993; Kolodny, 1994). On the other hand, Knowlton, Squire, and Gluck (1994) found that their amnesic patients performed as well as controls during the first 50 trials but thereafter showed a deficit. This study used 14 highly distinct stimuli, so it is possible that the amnesic deficit occurred because the control participants began memorizing some of the responses. This hypothesis is supported by the results of a study that used randomly configured dot patterns as stimuli (Kolodny, 1994). With confusable stimuli of this type, memorization is a more difficult strategy. In the Kolodny (1994) study, amnesic and control participants each classified several hundred dot patterns, yet there was no accuracy difference between the two groups, even during the last test block.

The one special case in which COVIS predicts that amnesic patients may be impaired in category learning is when the task is difficult (so that learning is slow) and the optimal rule is verbal. In such cases, an important component of success is to remember which verbal rules have already been tested and rejected. Depending on the duration of learning, amnesic patients may show deficits in this type of memory.

### *Depression*

COVIS also predicts that there may be selective categorization deficits in people suffering from depression. Depression may

appear in a number of psychiatric disorders, including schizophrenia and bipolar disorder (American Psychiatric Association, 1994). In these latter cases, the etiology may be different than with major depression, which is thought to be due to abnormal functioning in monoamine pathways (e.g., Hyman & Nestler, 1993). Specifically, current research has focused on serotonin, norepinephrine, and, to a lesser extent, dopamine. Because monoamine pathways project to a wide variety of neural sites, there is potential for depression-linked deficits in a wide variety of cognitive and motor tasks. Of particular relevance to COVIS are monoamine projections into the basal ganglia and prefrontal cortex.

Depressed patients often show psychomotor agitation or psychomotor retardation (American Psychiatric Association, 1994), both of which are consistent with abnormal functioning in the basal ganglia. In addition, they are impaired on the Wisconsin Card Sorting Test and other neuropsychological tests that are thought to measure prefrontal cortex functioning (e.g., Berman, Doran, Pickard, & Weinberger, 1993; Franke et al., 1993). In particular, depressed patients perseverate in rule-switching tasks, which is indicative of prefrontal cortex hypofunctioning. On the other hand, PET studies of monoamine metabolism are not entirely consistent with these behavioral data. Specifically, Ågren et al. (1993) detected abnormalities in the prefrontal cortex of depressed patients but found no differences between depressed and control patients with respect to monoamine metabolism in the basal ganglia.

On the basis of these results, COVIS predicts that people with depression should be impaired, relative to age-matched controls, on categorization tasks in which the optimal rule is verbal. The data on monoamine metabolism suggest that depression should not affect tasks in which the optimal rule is nonverbal, but the presence of characteristic psychomotor abnormalities makes this prediction more tentative.

J. D. Smith, Tracy, and Murray (1993) reported data that can be used to test these predictions. Depressed and nondepressed people (as measured by the Beck Depression Inventory; Beck, 1967) participated in two categorization tasks. In both tasks, two categories were constructed from stimuli (either letter strings or schematic faces) that varied on four dimensions. In one task, the optimal rule was verbal (i.e., unidimensional), and, in the other task, it was nonverbal. Call the four dimensions  $x_1$ ,  $x_2$ ,  $x_3$ , and  $x_4$ . Then, in the verbal task, the optimal rule was of the following form: "Respond A if the value on dimension  $x_3 < c$ ; respond B if the value on dimension  $x_3 > c$ , for some appropriately chosen constant  $c$ ." In the nonverbal task, perfect performance could be achieved by using the following rule: "Respond A if  $x_1 + x_2 + x_3 + x_4 < c$ ; respond B if  $x_1 + x_2 + x_3 + x_4 > c$ ." This rule is nonverbal because the dimensions are expressed in different units.

In two separate experiments, one with university students and one with psychiatric and medical patients, there was no difference in the performance of the depressed and control groups on the nonverbal task. However, in both experiments, the depressed participants were impaired on the verbal task relative to the nondepressed controls.<sup>10</sup> Specifically, on the verbal task, the

<sup>10</sup> In the study with psychiatric and medical patients, a third group of moderately depressed participants performed the same as the nondepressed control group.

depressed participants required significantly more training to achieve a criterion level of accuracy. Interestingly, however, both before and after criterion accuracy was reached, depressed and nondepressed participants performed equivalently (see Figure 6a for the shifted learning curves of the two groups). The only difference between the two groups involved the number of training blocks required to achieve criterion accuracy. According to COVIS, these results are expected from anyone with a selective deficit in the rule selection and switching architecture of prefrontal cortex and cingulate gyrus.

### Children

The prefrontal cortex is one of the last brain areas to fully develop during normal growth (Diamond & Goldman-Rakic, 1989). As a result, COVIS predicts that young children (e.g., under the age of 5 years) will have difficulty in rule switching and so will be impaired in category learning tasks in which the optimal rule is verbal. A number of results in the literature support this prediction. First, in tasks in which the optimal rule is verbal, young children perseverate on incorrect rules, even in the face of overwhelming disconfirmatory evidence (Aschkenasy & Odom, 1982; Gholson, Levine, & Phillips, 1972; Kemler, 1978).

A second set of relevant results comes from restricted classification tasks, in which participants are presented with three stimuli and asked to decide which two "go together best" or "belong in the same group." Call the stimuli *A*, *B*, and *C*. In the most popular design, stimuli *A* and *B* have an identical value on one stimulus dimension but differ substantially on a second dimension. In contrast, stimuli *B* and *C* differ by a small amount on two dimensions. This is an unsupervised task; no feedback is given (because there is no objectively correct response). Participants who sort by overall similarity (a nonverbal rule) will place stimuli *B* and *C* in the same category, whereas participants who use a logical criterion (a verbal rule) will place stimuli *A* and *B* together. There is little or no chance for procedural learning to occur in this task (i.e., because no feedback is given), so COVIS predicts that normal adults will tend to use unidimensional rules. This finding is well established (e.g., J. D. Smith, 1989; J. D. Smith & Kemler, 1977, 1978; Ward, 1983). For many years, it was thought that children sort by overall similarity. However, recent results and more sophisticated reanalyses of the earlier data indicate that children also use unidimensional rules (Cook & Odom, 1988, 1992; Thompson, 1994; Wilkening & Lange, 1987). Even so, in accord with COVIS predictions, children tend to use these rules inconsistently, and they often fixate on a single stimulus dimension (Thompson, 1994).

It is interesting to note, however, that adults have been reported to switch from the verbal to the nonverbal rule in this restricted classification task when encouraged to respond quickly (J. D. Smith & Kemler-Nelson, 1984; Ward, 1983) or impulsively (i.e., on the basis of their first impression; Foard & Kemler-Nelson, 1984; J. D. Smith & Kemler-Nelson, 1984). These results are easily predicted by COVIS if the reasonable assumption is added that processing time for the verbal system is greater than processing time for the implicit system. According to COVIS, another manipulation that should induce adults to switch from the verbal to the nonverbal rule in re-

stricted classification is to require them to perform simultaneously a secondary verbal task. J. D. Smith and Kemler-Nelson (1984, Experiment 6) required adults to count backward orally by 17s while simultaneously performing the restricted classification task. As predicted by COVIS, the concurrent verbal task significantly increased the frequency of overall similarity classifications (i.e., use of the nonverbal rule).

### Nonhuman Animals

All vertebrates have a brain structure analogous to the striatum, but they do not all have a prefrontal cortex or even a neocortex. For example, in rats, the prefrontal cortex is almost nonexistent (Uylings & van Eden, 1990), and pigeons have no neocortex (Patterson & Rose, 1992). Instead, the striatum plays the role of the cortex in the pigeon brain. In these animals, the analogue of the verbal system is absent. Some animals (e.g., nonhuman primates) may have explicit reasoning systems that, although primitive relative to those of humans, allow them to solve certain problems through "insight" (e.g., Köhler, 1925). Therefore, the complexity of the "verbal" system may vary continuously across the phylogenetic scale. Even so, COVIS predicts that most nonhuman animals must learn to categorize via some sort of striatal-based implicit system. One might even argue that, because of common ancestors, the implicit systems of vertebrates are all similar in some rudimentary fashion (e.g., Ashby & Lee, 1993).

These hypotheses lead to some straightforward predictions about the differences between nonhuman and human category learning. In particular, humans should show the greatest advantage over other animals in category learning tasks in which the optimal rule is verbal. Although such a predicted interaction is interesting, note that COVIS makes the obvious prediction that, under most conditions, humans will outperform other animals in all types of category learning tasks. Even so, COVIS predicts that, under certain special conditions, animals may outperform humans. In particular, consider a task in which the optimal rule is nonverbal but in which there is a verbal rule that performs reasonably well. In this case, pigeons or rats may learn to perform optimally, but, as a result of the initial bias favoring the verbal system, humans may be misled into using the suboptimal verbal rule. Such a demonstration would be strong evidence in favor of the multiple categorization systems proposed by COVIS.

In fact, a demonstration very similar to this was reported by Jitsumori (1993). The performance of pigeons and humans was compared in the same category learning task. Two categories comprised visual stimuli that varied on the dimensions of symbol color, background color, symbol shape, and symbol numerosity. Denote these four dimensions by  $x_1$ ,  $x_2$ ,  $x_3$ , and  $x_4$ , respectively. Perfect performance could be achieved by the following nonverbal rule: "Respond *A* if  $x_1 + x_2 + x_3 > c$ ; respond *B* if  $x_1 + x_2 + x_3 < c$ ," where  $c$  is the appropriate response criterion. The pigeons were reinforced for key pecks to all category *A* exemplars. No reinforcement was given to any key pecks made in the presence of category *B* exemplars.

All participants were trained to a criterion level of accuracy on a subset of the category exemplars. They were then shown a set of 56 new stimuli, some of which were further removed



from the optimal bound than any of the training stimuli. All pigeons showed a high level of transfer to the new stimuli. In contrast, the humans performed poorly on the transfer stimuli. For example, only 1 of 10 humans exhibited better-than-chance performance on the transfer stimuli that were furthest from the optimal bound. When queried, the humans reported using complex verbal rules that "were all in conditional or 'if, then' form" (Jitsumori, 1993, p. 249). Thus, these data support the conclusion that the humans constructed elaborate verbal rules during training. Apparently, these rules worked well for the training stimuli, but they did not perfectly separate the categories and so failed on the transfer stimuli.

A fundamental assumption of COVIS is that in the learning of some new category structure, there is a large initial bias in favor of the verbal system. The implicit system can overcome this bias only if it is reinforced at a higher rate than the verbal system over some reasonably long block of trials. In the Jitsumori (1993) experiment, the humans found a verbal rule that succeeded on the training stimuli. Thus, during training, the verbal system was reinforced at least as frequently as the implicit system. As a result, COVIS predicts that the initial bias in favor of the verbal system will be preserved throughout training and, therefore, that humans will tend to use verbal rules during the transfer condition. In contrast, because pigeons lack a verbal system, their only recourse was to rely on procedural learning, which, in this case, was the optimal strategy.

### Conclusions

The predictions of COVIS are strong and clear cut. Any abnormality in the tail of the caudate nucleus should cause deficits in visual category learning tasks in which the optimal rule is nonverbal, whereas widespread abnormalities throughout the caudate should cause deficits in tasks in which the optimal rule is verbal. Although more research is needed, these predictions are generally supported by results from elderly people and from patients with Parkinson's or Huntington's disease. Abnormalities or immaturity in the cingulate–prefrontal cortex should lead to deficits in categorization tasks in which the optimal rule is verbal. Furthermore, assuming that the striatum is fully developed and functioning normally, performance should be normal in categorization tasks in which the optimal rule is nonverbal. Evidence from patients with frontal lesions, depressed adults, children, and nonhuman animals generally supports this prediction. Finally, neuropsychological patients with damage to structures that are not part of the COVIS network should show relatively normal category learning. Evidence from patients with amnesia generally supports this prediction.

### A Connectionist Implementation of COVIS

Figure 2 describes the neuropsychological "hardware" mediating the separate verbal and implicit categorization systems hypothesized by COVIS. This section develops a connectionist network that provides an algorithmic-level description of COVIS and makes quantitative predictions about trial-by-trial learning. Later we describe the global-level dynamics of this network. The resulting computational-level version of COVIS

is useful for deriving specific global-level predictions about category learning.

It is straightforward to construct a connectionist network with the same global architecture as the neuropsychological version of COVIS shown in Figure 2. When applied to the circle condition from the Maddox and Ashby (1993; i.e., see Figure 1) experiment, the resulting model takes the form shown in Figure 3. Each component in the network is constructed from standard connectionist architecture. The unique feature of the model, therefore, is that it assumes separate (and parallel) verbal and implicit systems that compete throughout category learning. Recently, Erickson and Kruschke (in press) proposed a similar connectionist model of category learning. In the Erickson and Kruschke model, however, the competition is between rule-based (i.e., unidimensional) and exemplar-based categorization systems.

In the Figure 3 model, the implicit system is a perceptronlike network that implements the decision bound model of categorization (of the type described by Maddox & Ashby, 1993). Let  $\mathbf{x}_n = [x_1(n), x_2(n)]'$  denote the psychological representation—as computed in IT, for example—of the  $n$ th stimulus in the training sequence. It is well documented that, in experiments such as the one illustrated in Figure 1 (i.e., in which the optimal decision bound is linear), the category *A* and category *B* responses of individual participants are accurately separated by a linear bound (Ashby & Gott, 1988; Ashby & Maddox, 1990; Maddox & Ashby, 1993). Therefore, in experiments in which the optimal bound is linear, we assume that the striatal output of the implicit system can be modeled by the linear discriminant function

$$h_i(\mathbf{x}_n) = a_1(n)x_1(n) + a_2(n)x_2(n) + b_i(n), \quad (1)$$

where  $a_1(n)$ ,  $a_2(n)$ , and  $b_i(n)$  are weights that the network updates after each trial. The decision bound implemented by the implicit system on trial  $n$  is the set of all  $\mathbf{x}$  satisfying  $h_i(\mathbf{x}_n) = a_1(n)x_1(n) + a_2(n)x_2(n) + b_i(n) = 0$ . It is straightforward to generalize this model to allow more complex decision bounds (e.g., quadratic), but, for data from the Figure 1 experiments, linear bounds are sufficient. It is important to realize, however, that although we are modeling the striatal output of the implicit system by the discriminant function shown in Equation 1, we make no assumptions about the manner in which this discriminant function is computed. As mentioned before, the two prominent possibilities are that the striatum learns the weights  $a_1(n)$ ,  $a_2(n)$ , and  $b_i(n)$  directly or that the striatum simply learns to associate responses with regions of perceptual space (e.g., via an algorithm like the one proposed by Ashby & Maddox, 1989). In this latter model, the linear decision bound is the partition between the regions of perceptual space that are associated with each of the two responses.

The verbal system has two parts. The rule selection mechanism selects the rule type that will be active during the upcoming trial. Once a rule type is selected, the verbal system computes its own discriminant function. We now describe the rule selection algorithm. Denote the set of all salient verbal rule types by  $\mathcal{R} = \{R_1, R_2, \dots, R_m\}$ , and suppose rule  $R_i$  is used on trial  $n$  (with the circle stimuli  $m = 2$ ). If the response on trial  $n$  is correct, then rule  $R_i$  is used again on trial  $n + 1$ . If the response

on trial  $n$  is incorrect, then the active rule on trial  $n + 1$  is selected via the following three steps.

Step 1: Choose a rule at random from  $\mathcal{R}$ . Call this rule  $R_j$ .

Step 2: Define a weight,  $Y_r(n)$ , for each rule as follows:  $Y_i(n) = Z_i(n) + \gamma$ , for the active rule  $R_i$ ;  $Y_j(n) = Z_j(n) + X$ , for the rule  $R_j$  chosen in Step 1; and  $Y_k(n) = Z_k(n)$ , for all rules  $R_k \neq R_i$  or  $R_j$ . The constant  $Z_r(n)$  measures the current salience of rule  $R_r$ . The initial salience of rule  $R_r$  is  $Z_r(0)$ , but this value is updated each time rule  $R_r$  is used. Specifically, if rule  $R_r$  is used on trial  $n - 1$  and a correct response occurs, then  $Z_r(n) = Z_r(n - 1) + \Delta_C$ , where  $\Delta_C$  is a positive constant. If rule  $R_r$  is used on trial  $n - 1$  and an error occurs, then  $Z_r(n) = Z_r(n - 1) - \Delta_E$ , where  $\Delta_E$  is also a positive constant. The parameter  $\gamma$  is a positive constant that reflects the tendency of the participant to perseverate. The larger the value of  $\gamma$ , the greater the tendency to perseverate on the current rule, even in the presence of negative feedback. Finally,  $X$  is a positive-valued random variable that represents the attempt of the anterior cingulate to select rule  $R_j$ . We have assumed that  $X$  is Poisson distributed.

Step 3: Choose the rule for trial  $n + 1$  with the greatest weight  $Y_r$ ; that is, choose rule  $R_r$  on trial  $n + 1$  if  $Y_r(n) = \max\{Y_1(n), Y_2(n), \dots, Y_m(n)\}$ .

This algorithm has a number of attractive properties. First, the more salient the rule, the higher the probability that it will be selected. For example, with rectangles, shape is more salient than height; that is,  $Z_{\text{shape}}(0) > Z_{\text{height}}(0)$  (e.g., Krantz & Tversky, 1975). Thus, on the first trial of a category learning task in which the exemplars are all rectangles, we expect the verbal system to select a rule on shape with greater probability than a rule on height. Second, after the first trial, feedback is used to adjust the selection probabilities up or down, depending on the success of the rule type. Third, the model has separate selection and switching parameters, reflecting the assumption of COVIS that these are separate operations. The random variable  $X$  models the selection operation. The greater the mean of  $X$ , the greater the probability that the selected rule ( $R_j$ ) will become active. Following Ashby et al. (1998), we assume that this parameter (i.e., the mean of  $X$ ) increases with the dopamine level in anterior cingulate. In contrast, the parameter  $\gamma$  models switching

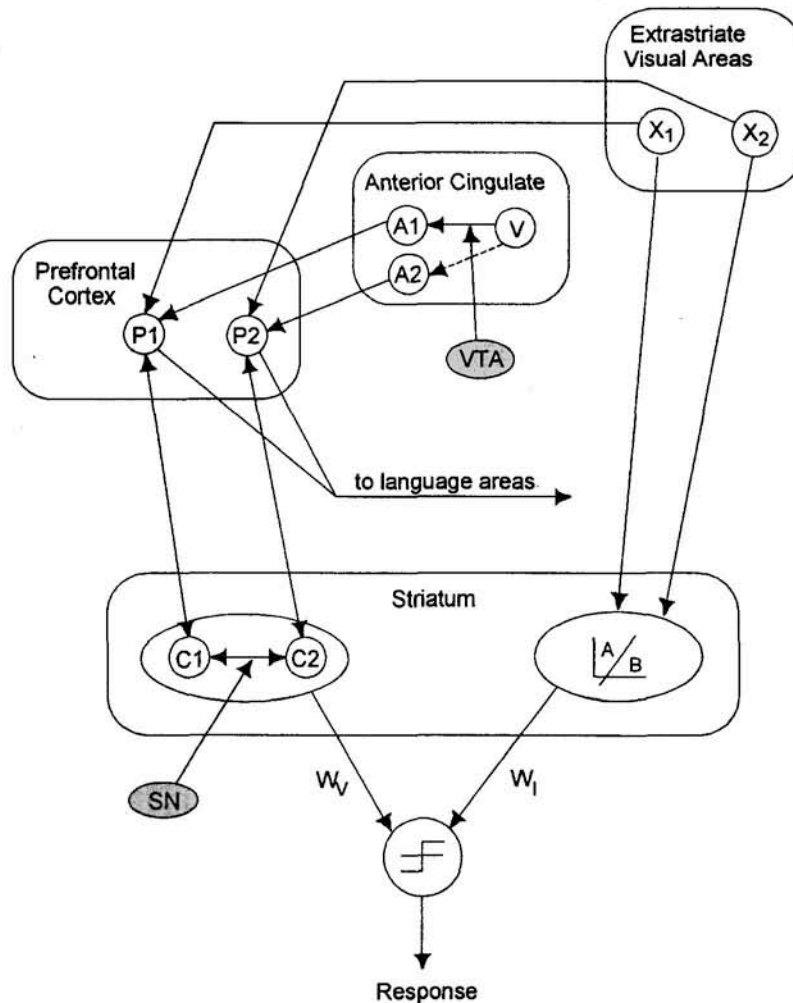


Figure 3. A schematic illustrating a connectionist implementation of COVIS (competition between verbal and implicit system). VTA = ventral tegmental area; SN = substantia nigra.

because, when  $\gamma$  is large, it is unlikely that the system will switch to the selected rule  $R_j$ .

Once a rule type is selected, the verbal system computes its own discriminant function. With the circular stimuli of Figure 1, only two verbal rules are salient. If the verbal system chooses to set a criterion on length, then the relevant discriminant function is  $h_{R_1}(x_n) = x_1(n) + b_{R_1}(n)$ ; if the choice is to set a criterion on orientation, the discriminant function is  $h_{R_2}(x_n) = x_2(n) + b_{R_2}(n)$ , where  $b_{R_1}(n)$  and  $b_{R_2}(n)$  are weights that the network updates during the course of learning. The verbal system in the COVIS connectionist network is similar to the rule competition model proposed by Busemeyer and Myung (1992). The two models make different assumptions about the rule selection process, but the global-level descriptions are identical.

Next, the network weights the verbal and implicit discriminant values by the system weights  $W_V(n)$  and  $W_I(n)$ . These values measure the degree to which the network believes in or trusts the component systems. Initially,  $W_V(n)$  is much greater than  $W_I(n)$ , both because the verbal system is the only one controlled by consciousness and because, at the outset of the experiment, the implicit system has not yet learned anything. As the experiment progresses, feedback is used to adjust the two system weights up or down depending on the success of the relevant component system.

To select a response, the network compares the two weighted discriminant values. The one with the larger absolute value determines the response. For example, suppose

$$|W_V(n)h_{R_i}(x_n)| > |W_I(n)h_I(x_n)|, \quad (2)$$

where  $i = 1$  or  $2$ , depending on whether the verbal system selected Rule Type 1 (i.e., set a criterion on length) or 2 (i.e., set a criterion on orientation). Then the decision rule of the network is "Respond A if  $h_{R_i}(x_n) < 0$ ; respond B if  $h_{R_i}(x_n) > 0$ ." If, instead,

$$|W_V(n)h_{R_i}(x_n)| < |W_I(n)h_I(x_n)|,$$

then the overall decision rule is "Respond A if  $h_I(x_n) < 0$ ; respond B if  $h_I(x_n) > 0$ ."

As mentioned earlier, the absolute magnitude of the discriminant value can be interpreted as a measure of response confidence. A discriminant value near zero occurs when the stimulus is near the decision bound. In this case, confidence is low because an error of any kind (e.g., perceptual or criterial noise) could change the response (i.e., move the stimulus to the other side of the bound). A discriminant value far from zero occurs when the stimulus is far from the bound. In this case, confidence is high because even a substantial error would not change the response.<sup>11</sup>

In this model, learning occurs in four places. First, the implicit system must learn the values of  $a_1(n)$ ,  $a_2(n)$ , and  $b_1(n)$  that are associated with the optimal bound. Second, the verbal system must learn which rule type is most effective. Third, the verbal system must learn the values of the response criteria  $b_{R_1}(n)$  and  $b_{R_2}(n)$  that maximize the accuracy of the verbal rules  $R_1$  and  $R_2$ , respectively. Finally, the overall network must learn whether the verbal system or the implicit system is more accurate; that is, it must learn the optimal values of  $W_V(n)$  and  $W_I(n)$ . A

detailed description of all parts of the network is provided in Appendix A.

Consider again the categorization task described in Figure 1. Suppose that, by the end of training, the verbal system has settled on a verbal rule on orientation (i.e., on each trial, the probability that the verbal system selects a rule on orientation is much greater than the probability that it selects a rule on length). Suppose, also, that the implicit system has learned the optimal bound for this task. Under these conditions, what is the observable performance of the overall network? In other words, what is the effect of the competition between the verbal and implicit systems on observable categorization behavior? It is straightforward to show that, in the absence of noise, the overall network's Category A and Category B responses will be partitioned perfectly by a linear bound that is intermediate in slope to the effective bounds of the implicit and verbal systems. The argument is sketched out in Figure 4.

The dotted lines in Figure 4 illustrate the asymptotic bounds learned by the verbal and implicit systems. These two bounds partition the psychological space into four regions. In the regions marked "1" and "2," the verbal and implicit systems agree on their category response (i.e., A in Region 1 and B in Region 2). In Regions 3 and 4, which are shaded, the two systems disagree. Suppose the verbal and implicit weights are equal, that is,  $W_V(n) = W_I(n)$ . In this case, the network emits the response associated with the system for which the current stimulus is farthest from its bound. For example, consider the stimulus at the point illustrated in Region 4. This point is closer to the verbal bound than the implicit bound. As such, the implicit system is more confident in its suggestion to respond B than the verbal system is in its suggestion to respond A. As a consequence, the network responds B to this stimulus. In fact, it is easy to see that any stimulus in Region 4 that falls below the solid line bound is farther from the implicit bound than from the verbal bound, so the network responds B to all of these stimuli. Similarly, any stimulus that falls above the solid line bound in Region 4 is farther from the verbal bound, so the network responds A. In Region 3, the argument is reversed. Points above the solid line are farther from the implicit bound, so the network responds A, whereas points below the solid line are farther from the verbal bound, so the network responds B. In summary, the network responds A to any stimulus falling above the solid line bound and B to any stimulus falling below,

<sup>11</sup> For simplicity of presentation, we have expressed confidence as the absolute value of the weighted discriminant value. It is straightforward to show that the discriminant value  $h(x)$  can be interpreted as the log of the likelihood ratio of the estimated category distributions (e.g., Ashby, 1992; Ashby & Alfonso-Reese, 1995). Balakrishnan and Ratcliff (1996) recently reported evidence that response confidence is more closely related to the distance between the percept and the decision bound than to the likelihood ratio. This result suggests that a more accurate model would replace Equation 2 with

$$|W_V(n)D_V(x_n)| > |W_I(n)D_I(x_n)|,$$

where  $D_J(x_n)$  is the distance from  $x_n$  to the decision bound used by system  $J$  (where  $J = V_i$  or  $I$ ). Even so, for the applications reported in this article, the two models of response confidence make the same predictions.

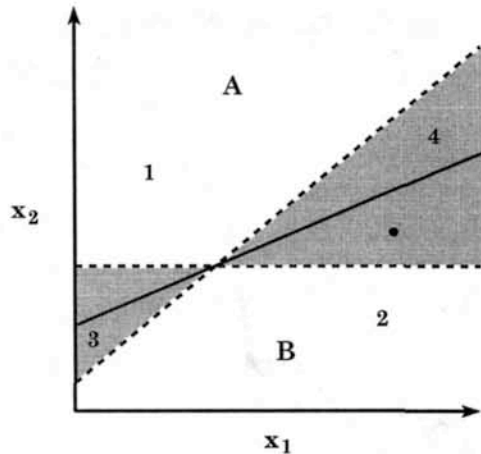


Figure 4. A demonstration that the connectionist version of COVIS (competition between verbal and implicit systems) predicts sharp category bounds (in the absence of perceptual and criterial noise). The category response regions are denoted by "A" and "B."

even though this bound is not used by either the verbal or implicit system. If the verbal and implicit system weights are unequal and both are greater than zero—that is,  $W_v(n) \neq W_i(n) \neq 0$ —then the observable bound will still be intermediate in slope to the verbal and implicit bounds, but it will be closer to the bound of the system having the larger weight.

As an initial test of this connectionist implementation of COVIS, we asked whether the network could account for learning curves from standard perceptual categorization experiments. There are many such curves published in the literature. However, in almost all cases, these curves are averaged across participants in a block-by-block fashion. Unfortunately, this practice can obscure important qualitative properties of the individual participant learning curves. For example, when the optimal rule is verbal, COVIS predicts that accuracy will often increase dramatically over the course of just a few trials, with the jump occurring when the participant suddenly discovers the correct rule. It is well known that sudden jumps of this type will not appear in the averaged data, even if they appear in the data of each participant (because they are likely to occur on different trials for different participants; see, e.g., Estes, 1956; Laming, 1973). A better way to present learning data is to plot the so-called backward learning curves. The idea is to find the first block of trials for which each participant first achieves a criterion accuracy level. This block is then designated as Block 0 for each participant, and, for example, the immediately preceding block becomes -1 and the immediately succeeding block becomes +1. After this renumbering is complete, the data can be safely averaged over participants.

The first data set that we modeled came from the 3 participants in the circle condition of the Maddox and Ashby (1993) experiment described in Figure 1. The averaged backward learning curve from this experiment is shown in Figure 5a. Each block included 100 trials, and criterion accuracy was defined as 75% correct for two consecutive blocks (this was the most stringent criterion met by all 3 participants). Note that accuracy appears to increase gradually over the course of the experiment,

until an asymptotic value is reached. None of the individual learning curves showed any dramatic jumps. Thus, learning appears to be incremental.

When simulating the performance of COVIS in this experiment, we assumed that the two verbal rule types are equally salient:  $Z_{size}(0) = Z_{orientation}(0)$ . We also assumed that the initial system weights strongly favored the verbal system:  $W_v(0) = .99$ ,  $W_i(0) = .01$ . The other parameters in the network (e.g., learning rates) were all set to certain prespecified values as well (see Appendix A for details), with the exception of the decay rate on the output node, which was coarsely estimated from the data. Thus, the model had only one free parameter in this application. The resulting learning curve, averaged over 50 replications, is shown in Figure 6b. The criterion of two consecutive blocks of at least 75% correct was met in 48 of the 50 replications. The data from the 2 replications that failed to meet this

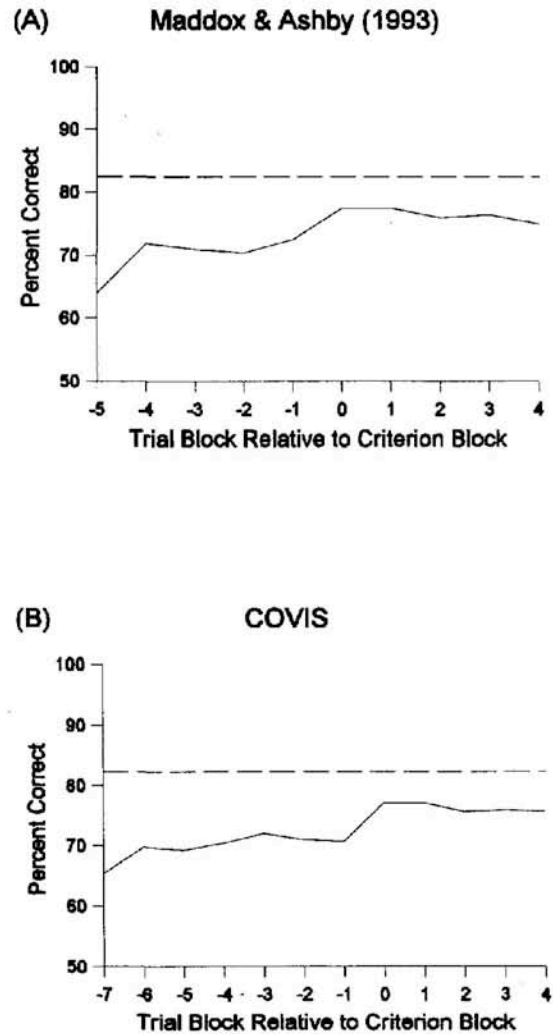


Figure 5. Backward (i.e., shifted) learning curves from the circle condition of an experiment reported by Maddox and Ashby (1993; i.e., illustrated in Figure 1). a: Learning curve from the 3 participants in the experiment. b: Simulated learning curve from the COVIS (competition between verbal and implicit systems) connectionist network.

criterion are not included in the Figure 5b curve. Note the incremental increase in accuracy from an initial percentage correct rate of about 65% to a final value of about 77%. These values agree nicely with the data shown in Figure 5a (which increase from 64% to 77.5%). As expected, the final system weights favored the implicit system, but it is interesting to note that the verbal weight asymptoted at a substantial nonzero value:  $W_v(1,200) = .42$ ,  $W_i(1,200) = .58$ . Thus, the model predicts that, even after 1,200 trials of practice, participants were still using verbal rules on a significant proportion of trials.

In the Maddox and Ashby (1993) rectangle condition, learning occurred so suddenly and so early in training that the learning curves were essentially flat for all participants. The qualitative difference between the learning curves in the circle and rectangle conditions supports the hypothesis of separate category learning systems, but it is not much of a challenge for COVIS to account for the flat learning curves from the rectangle condition (this is easily done by assuming that the most salient verbal rule is one on shape). Learning curves that are qualitatively different from those in Figure 5, and that provide a much more difficult challenge for the model, were reported by J. D. Smith et al. (1993). As mentioned previously, in these experiments a number of depressed and nondepressed people participated in two categorization tasks. In both tasks, two categories of four exemplars each were constructed from stimuli (either letter strings or schematic faces) that varied on four binary-valued dimensions. In one task, the optimal rule was verbal (i.e., unidimensional), and, in the other task, it was nonverbal. The learning curves for both the depressed and nondepressed groups are shown in Figure 6a for the verbal condition in the experiment with letter strings. The learning curves from the verbal condition in the experiment with schematic faces exhibited the same qualitative properties as the curves shown in Figure 6a (i.e., see Figure 8 of J. D. Smith et al., 1993). In both experiments, each of the eight exemplars that defined the two categories was shown in random order on each of six consecutive blocks of trials. Note that participants in both groups dramatically increased to perfect accuracy within the course of a single block (although this critical block number differed across participants and groups).

Two properties of these data make them especially difficult to model. First, of course, is the sudden jump in accuracy exhibited by both groups over the course of a single block of eight trials (from about 65% correct to 100% correct). The second challenge is to account for the difference between the depressed and nondepressed groups. On average, the depressed participants required about one extra block of training to reach criterion accuracy. To our knowledge, no existing computational models of category learning can account for such differences.

In the network simulations, we assumed that the four verbal rule types (i.e., the four unidimensional rules) had equal initial salience and, again, that the initial system weights favored the verbal system:  $W_v(0) = .99$ ,  $W_i(0) = .01$ . We assumed that the only difference between the depressed and nondepressed groups was that the depressed group had a higher tendency toward perseveration (so  $\gamma$  was larger for the depressed group than for the nondepressed group). All other parameters for the two groups were the same (see Appendix A for details). Because the perseveration parameter (i.e.,  $\gamma$ ) has no effect on the

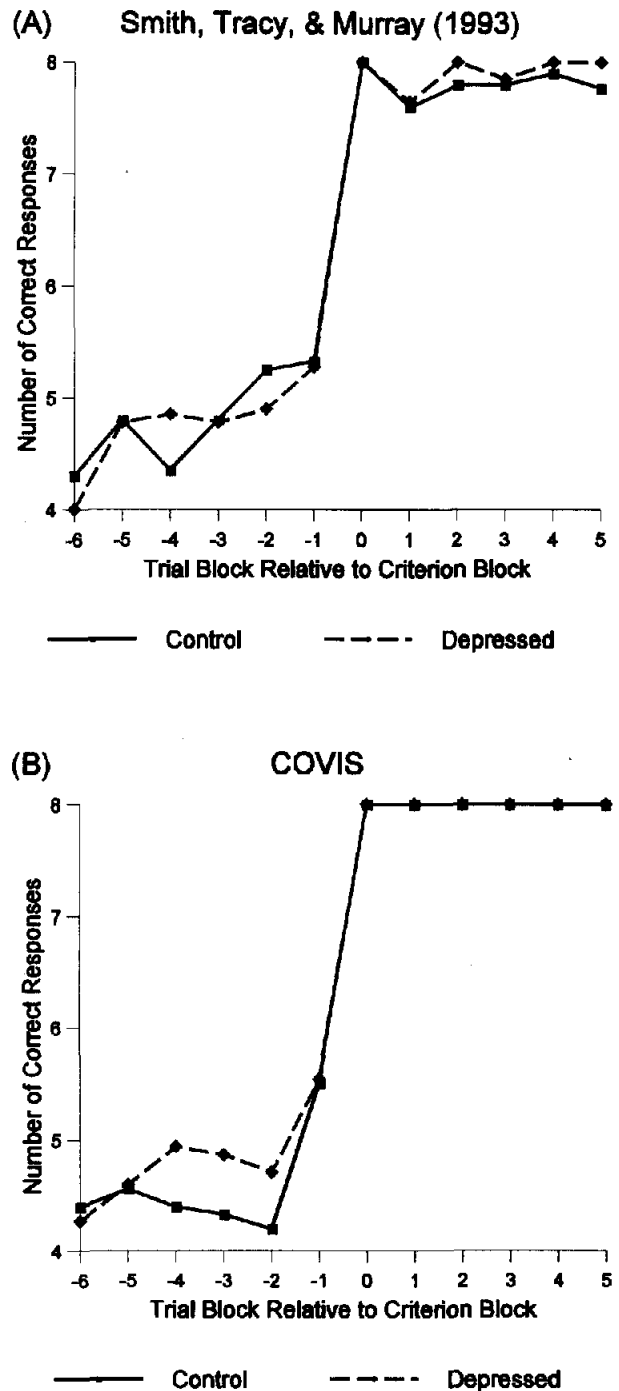


Figure 6. Backward (i.e., shifted) learning curves from an experiment reported by J. D. Smith, Tracy, and Murray (1993) in which the optimal rule was verbal. Panel (a): Learning curves from depressed and control participants. Panel (b): Simulated learning curves from the COVIS (competition between verbal and implicit systems) connectionist network.

implicit system, these assumptions mean that the COVIS network correctly predicts no difference between the depressed and nondepressed groups in the nonverbal conditions of the J. D. Smith et al. (1993) experiments.

Figure 6b shows the learning curves from the COVIS simulations, and Figure 7 shows the mean number of blocks until criterion for the J. D. Smith et al. (1993) participants and for the COVIS network. A comparison of Figure 6a and Figure 6b indicates that the COVIS network accurately reproduced the dramatic jump in accuracy during the last block of trials before criterion was reached (i.e., from about 65% correct to perfect accuracy). In COVIS, this jump is due primarily to the verbal system. Three of the four verbal rules achieve chance accuracy, and one yields perfect accuracy. Thus, the accuracy of the verbal system jumps dramatically on the trial when the verbal system first switches to the correct rule. A comparison of Figures 5 and 6 shows that COVIS successfully accounts for the qualitative difference between learning curves in "verbal" and "nonverbal" categorization experiments. COVIS also correctly accounted for the slower learning in the depressed group, although, overall, COVIS learned slightly slower than the Smith et al. (1993) participants.

We believe that the difference between Figures 5a and 6a represents a qualitative difference in how the verbal and implicit systems learn. In the COVIS implicit system, learning is incremental. As the bound gradually moves toward its optimal value, accuracy often will gradually increase. In contrast, in the verbal system, learning often occurs suddenly. Typically, different verbal rules will yield very different accuracy values, so accuracy will often increase dramatically on the trial in which the optimal verbal rule is first instantiated. On the other hand, exceptions should sometimes be expected. For example, if the optimal rule is verbal but complex, then learning may appear incremental because the participant may begin with a simple rule and successively generalize it in a series of discrete steps that eventually produce the optimal rule. In contrast, learning may be sudden in some tasks in which the optimal rule is implicit, if the contrasting categories are widely separated. This is because many bounds are optimal in such a case, so learning in the implicit system need not be as precise as in tasks in which only a single

bound maximizes accuracy (e.g., as in the Maddox & Ashby, 1993, experiments).

### A Global Description of COVIS Dynamics

The neuropsychological version of COVIS sketched in Figure 2 describes the "hardware" used by the separate verbal and implicit categorization systems. The connectionist network shown in Figure 3 provides a description of the algorithms used by the separate systems to select a response. Although the COVIS network can be used to generate predictions about trial-by-trial learning, the form of the model makes it difficult to derive global-level or qualitative predictions. In this section, we show how to overcome this traditional limitation of connectionist networks.

A connectionist network is an example of a dynamical system. A useful method of characterizing an unfamiliar dynamical system is to study its trajectories through the so-called "state space" (e.g., Strogatz, 1994; Wiggins, 1990). For example, consider the implicit component of the COVIS connectionist network. For the experiment illustrated in Figure 1, we assumed that the output of the striatal units could always be described by a linear discriminant function (i.e., by Equation 1). Thus, the state of the implicit system on trial  $n$  is given by the numerical values of the Equation 1 parameters, that is, by the triple  $[a_1(n), a_2(n), b_1(n)]$ . The state space is defined as the space in which there is one dimension for each parameter value (so, in the present case, the state space has dimensions  $a_1$ ,  $a_2$ , and  $b_1$ ). On each trial, the state of the system is represented by a single point in the state space, and the system trajectory is the path taken by the system (over trials) through the state space.

In the COVIS network, the implicit system adjusts its weights via a modified form of the delta rule (see Appendix A). As a result, the trial-by-trial learning of the implicit system is equivalent to a process of gradient descent over the probability of error surface (Rumelhart, Hinton, & Williams, 1986). Every point in the state space—that is, every triple  $[a_1(n), a_2(n), b_1(n)]$ —is associated with some probability of error in the categorization task. A gradient descent algorithm is one that always travels through the state space in the direction associated with the quickest decrease in probability of error. Gradient descent learning algorithms are extremely popular in the category learning literature. They are the basis, for example, of Kruschke's (1992) attention learning covering map (ALCOVE) model and of the hill-climbing model of category learning proposed by Busemeyer and his colleagues (Busemeyer & Myung, 1987; Busemeyer, Swenson, & Lazarte, 1986; Myung & Busemeyer, 1989).

Gradient descent is also a general property of any connectionist model that uses back propagation or the delta rule (Rumelhart et al., 1986). One well-known property of gradient descent learning is that it is highly susceptible to local minima. Specifically, once a gradient descent system reaches a local minimum, it must remain there for all time. Thus, all category learning models based on gradient descent predict that category learning problems in which the probability of error surface is characterized by local minima (over the relevant parameter space) will be difficult to solve.

The parameters  $a_1(n)$ ,  $a_2(n)$ , and  $b_1(n)$  describe the linear bound used by the implicit system. However, only two of the

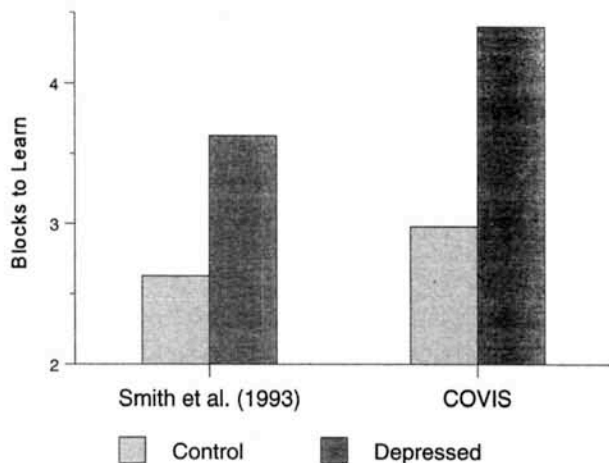


Figure 7. Number of blocks until learning occurred for the participants in an experiment reported by J. D. Smith, Tracy, and Murray (1993) and for the COVIS (competition between verbal and implicit systems) simulations of this experiment.

three parameters are free (only two parameters are needed to describe a line). It is much easier to visualize trajectories in two dimensions than in three dimensions, so, before proceeding, it is worthwhile to reparameterize the linear bound predicted by the implicit system. The usual parameterization of a line is in the slope–intercept form. Unfortunately, this form cannot be used to describe a vertical line (because the slope is infinite). As a consequence, we developed an alternative two-parameter description of linear decision bounds. Details are provided in Appendix B. Briefly, the parameters are  $\theta$  and  $\tau$ . The  $\theta$  parameter is the angle of rotation of the linear bound (i.e., the angle the bound subtends relative to the abscissa), so the slope of the bound is  $\tan(\theta)$ . The  $\tau$  parameter determines the intercept on the  $x_1$  or  $x_2$  dimension. To define it precisely, note that in Figure 1, all stimuli have positive values on both stimulus dimensions. Therefore, any reasonable decision bound must eventually intersect either the  $x_1$ -axis or the  $x_2$ -axis (or both) at some nonnegative value. Now, let  $\omega$  denote the smallest positive intercept (of the  $x_1$  and  $x_2$  intercepts). Then  $\tau$  is defined as follows:

$$\tau = \begin{cases} \omega, & \text{if the smallest positive intercept} \\ & \text{is the } x_1 \text{ intercept.} \\ -\omega, & \text{if the smallest positive intercept} \\ & \text{is the } x_2 \text{ intercept.} \end{cases}$$

The only purpose of this reparameterization is that it allows us to express any linear bound using only two parameters. As a result, we gain the convenience of expressing COVIS trajectories for the Figure 1 experiment in a two-dimensional, rather than a three-dimensional, state space.

The probability of error surface for the implicit system in the Figure 1 experiment is shown in Figure 8. The diagonal valley that cuts across Figure 8 shows a trade-off between  $\theta$  and  $\tau$  that keeps the overall error rate reasonably small. The pole, which indicates the minimum of the surface, defines the optimal bound. The gradient at any point on the surface gives a measure of how quickly error rate changes with changes in the bound. The trajectories of the implicit system in the COVIS network tend to follow the steepest downhill course over this surface.

With the circular stimuli, there are two types of salient verbal rules. The first is to set a criterion on length. This produces a vertical line (i.e.,  $\theta = 90^\circ$ ) decision bound (when the abscissa is length and the ordinate is orientation). The second type sets a criterion on orientation and produces a horizontal decision bound (i.e.,  $\theta = 0^\circ$ ). For each rule type, accuracy changes with the value of the criterion. The COVIS network adjusts these values via a modified form of the delta rule, so, for each rule type, the verbal system performs gradient descent over a one-dimensional state space (i.e., over alternative values of the response criterion). Figure 9 shows the one-dimensional probability of error surfaces associated with both rule types.

The verbal system in the COVIS network also learns which of the alternative verbal rule types are most accurate by iteratively adjusting the weights associated with the rule types in such a way that, as training progresses, the more accurate verbal rule type becomes more and more likely to be selected (see the previous section for details). This procedure tends to minimize the probability that the verbal system makes an error. A graphical method of deriving probability of error predictions for the verbal

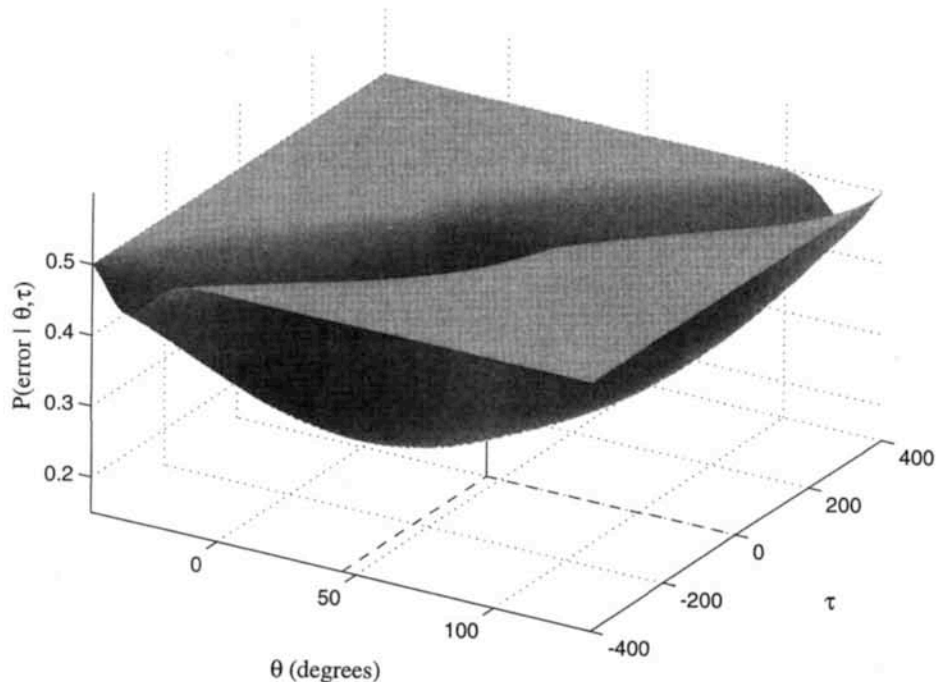


Figure 8. Probability of error for every possible linear bound in the circle condition of the Maddox and Ashby (1993) experiment.

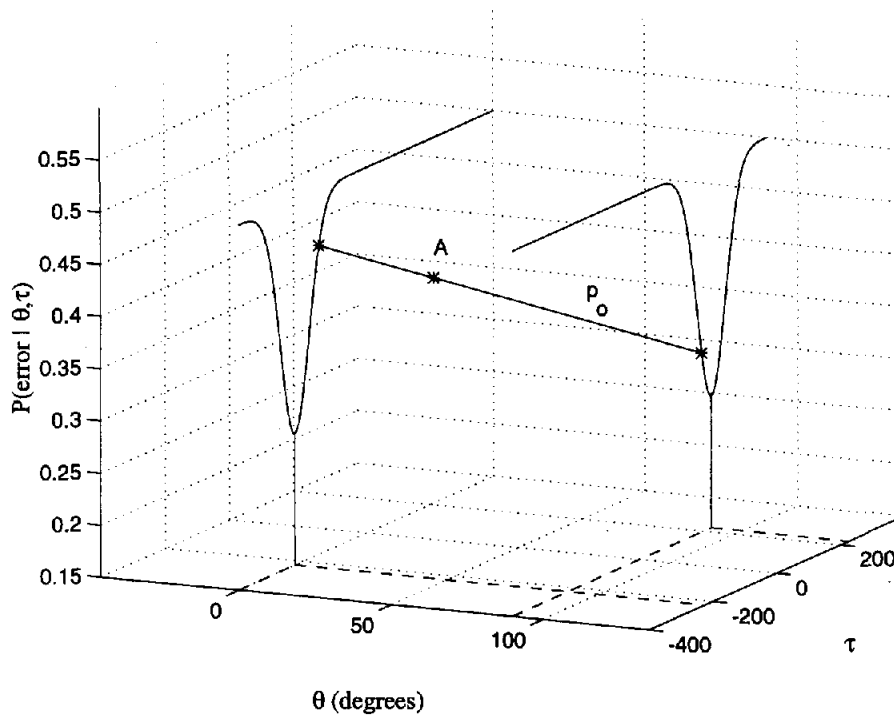


Figure 9. Probability of error for every possible version of the two salient verbal rule types in the circle condition of the Maddox and Ashby (1993) experiment. Also shown is a graphical method for deriving the overall probability of error predictions for the verbal system.

system as a whole, illustrated in Figure 9, is to draw a line that connects the probability of error associated with the current state of the two verbal rule types. Call the length of this line one unit and denote the  $\theta = 0^\circ$  endpoint as the orientation endpoint and the  $\theta = 90^\circ$  endpoint as the length endpoint. Suppose the probability that the verbal system selects a rule on orientation equals  $p_o$ . Then the probability of an error for the overall verbal system equals the height of the point A on the connecting line that is distance  $p_o$  from the length endpoint. The dynamical behavior of  $p_o$  can be anticipated by imagining that the line is a wire and the point A is a bead that slides freely downhill on that wire.

Using this method, it is straightforward to show that the COVIS network virtually always predicts that the verbal system will eventually settle on one rule type (and ignore the other), even when the two rule types are equally accurate (as in Figure 9). In other words, the state in which  $p_o = 0.5$  is not stable. Thus, in the circle condition of Figure 1, COVIS predicts that, by the end of the experiment, participants will consciously be focusing on one of the two stimulus dimensions (i.e., through the anterior attentional system), even though their implicit system may be attending to both dimensions equally (i.e., through the posterior attentional system).

The final step in understanding the global dynamics of COVIS is to predict trajectories of the overall combined COVIS network. The key to solving this problem is to recall that, in the absence of noise, the overall network's category A and category B responses will be partitioned perfectly by a linear bound that

is intermediate in slope to the effective bounds of the implicit and verbal systems (see Figure 4). Let  $\theta_I$  and  $\theta_V$  denote the slope parameters of the active decision bounds of the implicit and verbal systems, respectively (so  $\theta_V = 0^\circ$  or  $90^\circ$ ), and let  $\tau_I$  and  $\tau_V$  denote the corresponding intercept parameters. Then, in fact, the slope parameter of the observable overall bound equals

$$\theta_o = \frac{W_I \theta_I + W_V \theta_V}{W_I + W_V},$$

where  $W_I$  and  $W_V$  are the weights on the implicit and verbal systems, respectively. To find the intercept of the observable bound, note from Figure 4 that the observable bound passes through the point at which the verbal and implicit bounds intersect. Given the slope (i.e.,  $\theta_o$ ) and a point on the line, it is straightforward to solve for  $\tau_o$ .

Suppose the verbal system is restricted to using a rule on orientation (i.e., so  $\theta_V = 0^\circ$ ). Then the trajectories of the verbal system through the  $(\theta, \tau)$  state space have only a single degree of freedom ( $\tau$  is free to vary, but  $\theta$  is not). On the other hand, the trajectories of the implicit system have two degrees of freedom (both  $\theta$  and  $\tau$  are free to vary). Because the verbal and implicit systems move independently, the trajectories of the overall COVIS network in this example have three degrees of freedom. In other words, every triple  $(\theta_I, \tau_I, \tau_V)$  determines a single observable bound  $(\theta_o, \tau_o)$ . But because three degrees of freedom are used to fix the two parameters of the observable bound, any observable bound  $(\theta_o, \tau_o)$  can be realized by more



than one triple  $(\theta_I, \tau_I, \tau_V)$ . The implication of this mapping from a three-dimensional parameter space to a two-dimensional space is that the observable trajectories of the COVIS network through the  $(\theta_o, \tau_o)$  space can cross, even in the absence of noise. Trajectories of (deterministic) dynamical systems cannot cross (e.g., Wiggins, 1990), so there is no two-dimensional dynamical system equivalent to the COVIS network. One consequence of this result is that COVIS predicts that, at two different times in learning, a participant could have exactly the same observable bound, yet performance could deteriorate after the first time and converge to the optimal bound after the second time. In contrast, gradient descent models predict that participants can never spontaneously revisit the same state, and if some experimental intervention induces them back to the same state, then they must follow exactly the same learning trajectory the second time.

Another straightforward prediction of COVIS is that the observable trajectories will not satisfy gradient descent, even though all of its component networks learn by gradient descent. There are several reasons why gradient descent fails in the overall system. First, trajectories in gradient descent systems cannot cross (e.g., Wiggins, 1990). Second, and more important, gradient descent fails because the verbal system will typically jump back and forth between horizontal and vertical decision bounds. As a result, the overall learning trajectories should periodically make large discrete jumps. Because of such jumps, COVIS predicts that people will usually "jump out" of local minima. On the other hand, as learning progresses and participants discover which type of verbal rule is best, the probability of shifting from one type of verbal rule to another will gradually decrease. This process endows COVIS with a natural "simulated annealing" mechanism (Geman & Geman, 1984). As a consequence, COVIS predicts that the probability of escaping from a local minimum decreases with experience.

Alfonso-Reese (1996) obtained empirical estimates of learning trajectories in two categorization experiments using the randomization technique. In both experiments, the exemplars from the two categories were lines that varied in length and orientation, and the optimal decision bound was linear but not easily verbalized. The trajectories for all participants violated gradient descent. The trajectories initially showed large discrete jumps between the alternative verbal rules. Later in learning, the jumps became less frequent, and the trajectories gradually began converging on the optimal bound. Thus, the empirical trajectories displayed the same qualitative properties predicted by COVIS.

Because the various component networks learn by gradient descent, if there is no noise, the effective implicit system bound in COVIS should be nearly optimal at asymptote, so long as accurate feedback is provided on every trial. In addition, the verbal system should have learned which rule type is most accurate and the optimal value of the criterion associated with that rule. Under these conditions, the overall observable decision bound is determined by the value of the system weights,  $W_I$  and  $W_V$ . COVIS assumes that these weights are also learned via a modified form of the delta rule (see Appendix A), so in the absence of noise and in a task in which the optimal rule is nonverbal, the ratio  $W_I/(W_I + W_V)$  eventually should be driven to 1.0. Even so, there are at least two potential problems with this prediction.

First,  $W_I$  might not ever dominate  $W_V$  because the implicit system fails to learn the optimal bound. Second, the initial bias in favor of the verbal system means that a huge amount of change in the  $W_I$  and  $W_V$  parameters is required before the ratio  $W_I/(W_I + W_V)$  is close to 1.0. For example, in the simulations of the COVIS network described in the previous section, the initial values of the system weights were  $W_I(0) = .01$  and  $W_V(0) = .99$ . If the best verbal rule is nearly as accurate as the optimal bound, then the gradient driving  $W_I$  up and  $W_V$  down will be nearly flat. In this case, any noise in the system will probably cause learning of the system weights to stop, well before the ratio  $W_I/(W_I + W_V)$  is close to 1.0. As a consequence, the verbal system would contribute significantly to categorization behavior, even after learning had ceased and even though the optimal bound is nonverbalizable.

An experiment in which such an asymptotic suboptimality might occur is described in Figure 10. This experiment used the randomization technique (Ashby & Gott, 1988) described earlier with two categories, denoted *A* and *B*, each composed of lines of various lengths and orientations. Each plus sign represents an exemplar from Category *A*, and each circle represents an exemplar from Category *B*. The decision rule that maximizes overall accuracy is to respond *A* to any stimulus with length and orientation that fall above the solid line shown in Figure 10 and to respond *B* to any stimulus that falls below this line.

The optimal bound in this experiment had a slope of 0.6; therefore, the optimal decision rule is "Respond *A* if Orientation  $- 0.6 \times$  Length  $> 0$ ; respond *B* if Orientation  $- 0.6 \times$  Length  $< 0$ ." Because orientation and length are expressed in different units, this rule cannot be described verbally (or, at least, it has extremely low saliency). Two verbal rule types are salient: (a) "Respond *A* if the line is long; respond *B* if the line is short" and (b) "Respond *A* if the line is tilted up; respond *B* if the line is tilted down." Rules of the first type produce linear bounds that are vertical, whereas rules of the second type produce horizontal bounds. The most accurate versions of both of these rule types are shown by the dotted line bounds in Figure 10. Also shown are the predicted accuracies associated with each of these bounds.

Suppose that the implicit system learns the optimal bound for the experiment illustrated in Figure 10 and the verbal system learns both the type of verbal rule that is most accurate and the optimal intercept value associated with that rule type. In other words, suppose the implicit system learns to use the solid line bound in Figure 10 and the verbal system learns to use the dotted line horizontal bound. Figure 11 shows how accuracy changes in this experiment as a function of the ratio  $W_I/(W_I + W_V)$ . This entire function is extremely flat, so if there is any perceptual or criterial noise, one would expect the system weights to stabilize long before  $W_I$  approaches 1. As a result, COVIS predicts that asymptotic performance in this task is likely to be suboptimal.<sup>12</sup> The suboptimality will be difficult to see in a crude measure such as percentage correct, because the verbal system can achieve an accuracy that is nearly optimal.

<sup>12</sup> The analyses in this section all assumed that the implicit system uses a linear decision bound. If a more complex bound is used (e.g., quadratic), then the state space for the implicit system will be more

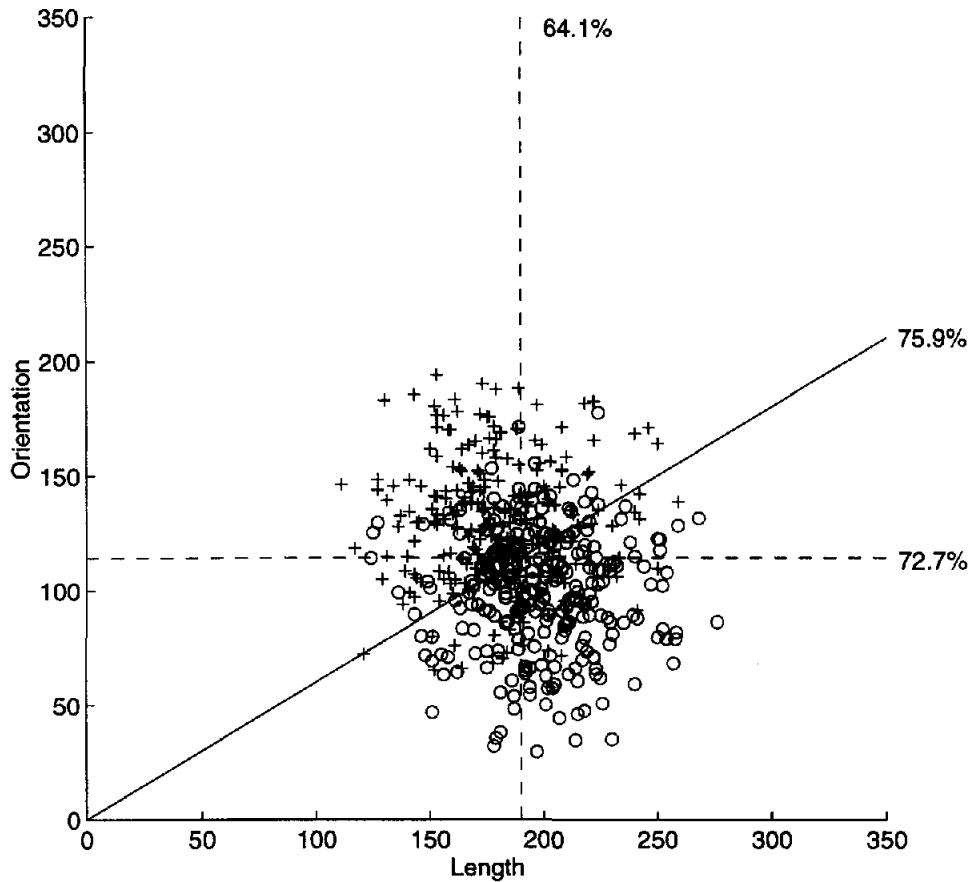


Figure 10. Category structure of a hypothetical experiment. A plus sign indicates an exemplar from Category A, and a circle indicates an exemplar from Category B. The optimal decision bound is the line  $x_2 = 0.6x_1$ . The vertical and horizontal dotted lines represent the most accurate verbal rules on length and orientation, respectively. The numerical values are the accuracies associated with each bound.

A better measure of this suboptimality is the slope of the decision bound that best accounts for the observed data. COVIS predicts that this slope frequently will be shallower than optimal. This prediction is tested in the next section.

Before ending this discussion, it is important to note that predicted suboptimality depends critically on the fact that some verbal rule is nearly as accurate as the optimal bound. In experiments in which no verbal rules achieve high accuracy, COVIS predicts that there will be no systematic bias toward verbal

rules, the reason being that accuracy will increase significantly with the ratio  $W_i/(W_i + W_v)$ . Thus, for example, COVIS correctly predicts the nearly optimal performance found in randomization experiments in which the optimal bound is quadratic (Ashby & Maddox, 1993).

#### A Bias Toward Verbal Rules

As described in the previous section, COVIS predicts that, so long as the verbal and implicit systems learn to use the most accurate possible bounds, participants in the Figure 10 experiment typically will not use the optimal bound. Instead, there will always be some residual pull of the most accurate verbal rule: "Respond A if the line is rotated more than 37.3° counterclockwise." As a consequence, COVIS predicts that even experienced participants will adopt a bound with a slope of less than 0.6. To our knowledge, COVIS is the only model of category learning that predicts (a priori) such an asymptotic bias in categorization performance. For example, consider Nosofsky's (1986) generalized context model, which, arguably, has been the most successful of the formal exemplar models. Nosofsky

than two dimensional. The main tools for deriving qualitative predictions from the COVIS connectionist network are the error surface for the implicit system illustrated in Figure 8, the error surfaces for the verbal system illustrated in Figure 9, and the error surface for the system weights illustrated in Figure 11. If the implicit system uses a complex decision bound, then the dimensionality of the error surface for the implicit system increases, but the dimensionality of the error surfaces for the verbal system and the system weights is unchanged. As a result, a qualitative analysis of this more complex version of COVIS would proceed in a similar fashion.

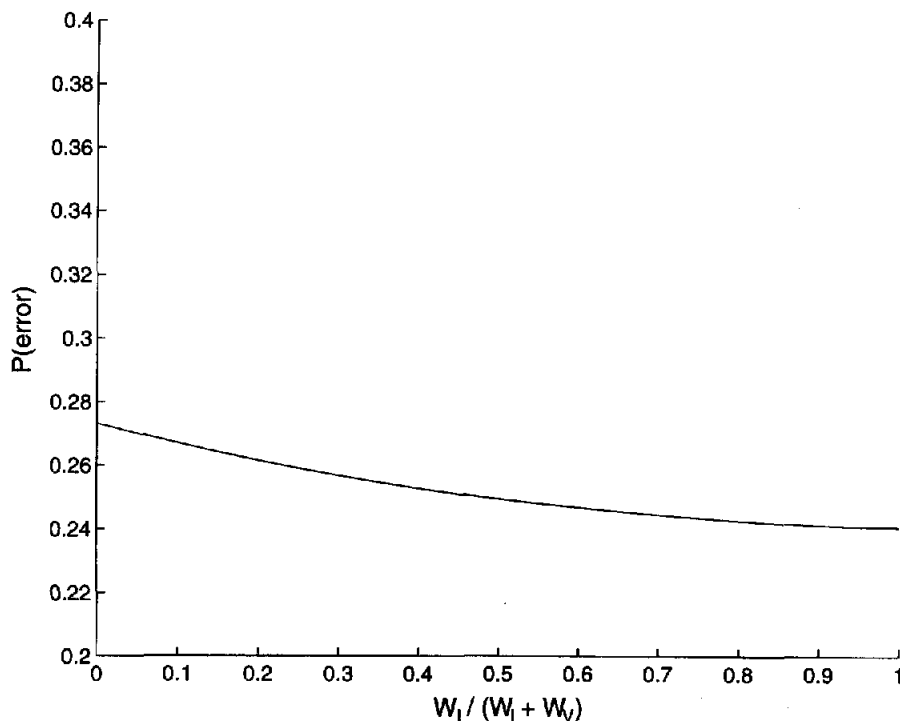


Figure 11. Probability of error for the experiment described in Figure 10 as a function of the relative strength of the implicit system, that is, of  $W_I / (W_I + W_V)$ .

(1986) assumed "that subjects will distribute attention among the component dimensions so as to optimize performance in a given categorization paradigm. That is, it is assumed that the attention parameters will tend toward those values that maximize the average percentage of correct categorizations" (p. 41). It is straightforward to show that the generalized context model predicts maximal accuracy in this experiment when equal attention is allocated to the two stimulus dimensions, regardless of the level of overall discriminability, and that under these conditions the "decision bound" of the model is identical to the optimal bound.<sup>13</sup> Thus, the generalized context model predicts a priori that there will be no systematic bias in this experiment.

One difficulty with this design is that decision bounds that are shallower than optimal might occur simply because participants are allocating too much attention to the orientation dimension, perhaps because orientation is more salient than length. In this case, it would be a mistake to infer a competition between separate verbal and implicit systems. To guard against this possibility, a second condition can be created by rotating the stimuli shown in Figure 10 about the line  $x_2 = x_1$ . The result is shown in Figure 12. Now the optimal bound has a slope of 1.67, and the most accurate nearby verbal rule is "Respond A if the length is greater than 114 pixels." According to COVIS, if the verbal rules on length and orientation are equally salient, then participants in this condition should settle on decision bounds with slopes that are greater than optimal (i.e., 1.67). However, if more attention is allocated to orientation than to length, then in this new experiment the best-fitting bounds should have slopes less than optimal.

In summary, there are four prominent possibilities in this experi-

ment. First, asymptotic performance could be optimal in both conditions. This result would be problematic for COVIS but easily accommodated by virtually all current category learning models (e.g., Kruschke's, 1992, ALCOVE model and Nosofsky, Kruschke, & McKinley's, 1992, exemplar learning model). Second, the best-fitting decision bounds could be shallower than optimal in both conditions, and, third, they could be steeper than optimal in both conditions. Finally, participants might use bounds that are shallower than optimal in the Figure 10 condition (i.e., the shallow condition) and bounds that are steeper than optimal in the Figure 12 condition (i.e., the steep condition). This last possibility is easily accommodated by COVIS and is the one that

<sup>13</sup> The generalized context model (GCM) predicts no decision bound per se. However, the analogue in the GCM is the equivocality contour (i.e., the contour on which the probability of responding A equals the probability of responding B). It follows directly from Propositions 3 and 4 of Ashby and Maddox (1993) that, in the Figure 10 experiment, the equivocality contour in the GCM is identical to the optimal decision bound if  $cw = c(1 - w)$ , where  $w$  is the attention allocated to the length dimension,  $1 - w$  is the attention allocated to orientation, and  $c$  is a measure of overall discriminability. Because predicted accuracy is maximized in the GCM when the equivocality contour equals the optimal decision bound, it follows that the GCM predicts a priori that equal attention will be allocated to the two dimensions and, thus, that there will be no systematic bias of the type predicted by COVIS. Note that this prediction holds regardless of the overall level of discriminability. In fact, if one adds the reasonable assumption that overall discriminability increases with experience, then this prediction holds even if allocation of attention is not optimal (Ashby & Alfonso-Reese, 1995).

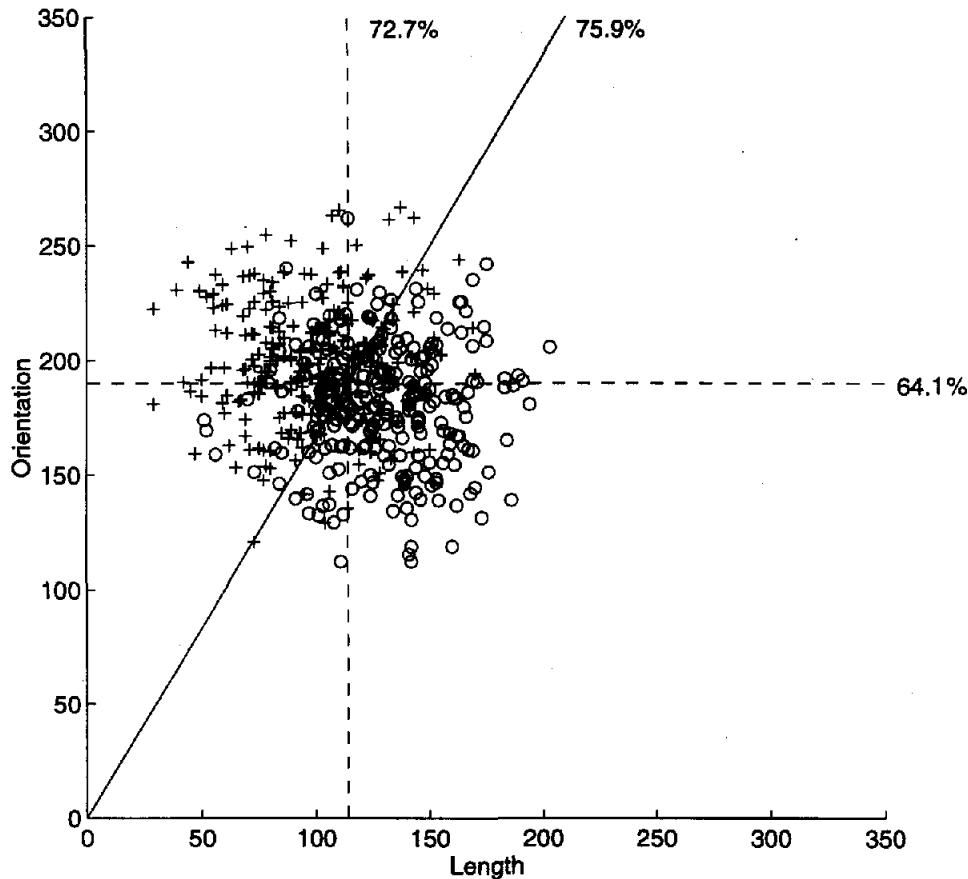


Figure 12. Category structure of an experiment in which the optimal decision bound is the line  $x_2 = 1.67x_1$ . The vertical and horizontal dotted lines represent the most accurate verbal rules on length and orientation, respectively. The numerical values are the accuracies associated with each bound. This is the category structure used in the steep condition. The structure used in the shallow condition is shown in Figure 10. A plus sign indicates an exemplar from Category A and a circle indicates an exemplar from Category B.

would be most problematic for other current models of category learning. We tested these predictions in the following experiment.

**Method**

*Participants.* Thirteen graduate students at the University of California, Santa Barbara, participated in this experiment. Five were assigned to the shallow condition, and 8 were assigned to the steep condition. Eleven students were each paid \$32 for four 40-min sessions. The other 2 students were members of F. Gregory Ashby's laboratory and participated in four sessions without pay. None of the participants had any prior knowledge of the design or goals of the experiment.

*Stimuli and apparatus.* The stimuli were lines varying in length and orientation. Each stimulus in both conditions was generated by randomly sampling from one of two bivariate normal category distributions. The category A distribution was specified by a mean vector  $\mu_A$  and a variance-covariance matrix  $\Sigma$ . The category B distribution was defined by a mean vector  $\mu_B$  and by the same covariance matrix  $\Sigma$ . The distribution parameter values in the shallow condition were as follows:

$$\mu_A = \begin{bmatrix} 180 \\ 130 \end{bmatrix}, \mu_B = \begin{bmatrix} 200 \\ 98 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} 729 & 0 \\ 0 & 729 \end{bmatrix}.$$

The stimulus set for the shallow condition is depicted in Figure 10. The category A stimuli are plotted as plus symbols, and the category B stimuli are plotted as circles. The optimal bound is  $x_2 = 0.6x_1$ . An ideal observer using the optimal bound could obtain 75.9% correct. The most accurate verbal rule on orientation or length could obtain 72.7% or 64.1% correct. The stimuli in the steep condition were generated by rotating the category distributions from the shallow condition about the line  $x_2 = x_1$ . As a result, the optimal bound is  $x_2 = 1.67x_1$  (see Figure 12). Optimal percentage correct is again 75.9%, and the maximum possible accuracy using unidimensional rules on orientation or length is now 64.1% or 72.7% correct.

Each  $(x_1, x_2)$  pair was converted to a line stimulus by letting  $x_1$  determine line length and  $x_2$  determine orientation. For example, the category A prototype in the shallow condition is converted to a line 180 pixels long rotated  $130(\pi/550)$  radians counterclockwise from horizontal. The  $(\pi/550)$  scaling factor was chosen in an attempt to equalize the salience

Table 2  
*Summary Statistics for the Last Experimental Session in the Steep and Shallow Conditions*

Condition	Participant	Percentage correct	General linear classifier statistic	
			AIC	Best-fitting bound
Shallow	1	61.6	455	$x_2 = -.06x_1 + 138$
	2	64.8	471	$x_2 = .22x_1 + 77$
	3	65.6	394	$x_2 = -.06x_1 + 129$
	4	67.8	454	$x_2 = .07x_1 + 87$
	5	75.2	283	$x_2 = .50x_1 + 15$
Steep	1	68.4	352	$x_2 = 6.9x_1 - 498$
	2	67.6	400	$x_2 = 21.4x_1 - 2,374$
	3	70.6	329	$x_2 = .75x_1 + 107$
	4	70.4	365	$x_2 = 1.09x_1 + 72$
	5	72.2	330	$x_2 = 1.87x_1 - 24$
	6	74.8	229	$x_2 = 1.52x_1 + 33$
	7	74.4	396	$x_2 = 1.27x_1 + 47$
	8	65.8	458	$x_2 = -56.7x_1 + 6,344$

Note. AIC = "A" information criterion.

of orientation and length. The orientation of the lines shown in the shallow condition varied from 10° to 63°, and their visual angle varied from 1.7° to 4.3°. The orientation of the lines shown in the steep condition varied from 37° to 87°, and their visual angle varied from 0.5° to 3.2°. The stimuli were computer generated and displayed on a Mitsubishi Electric Color Display Monitor (Model C-9918NB) in a dimly lit room.

*Procedure.* The participants' task was to classify each line stimulus into category A or B by pressing an appropriate response key. A brief high-pitched tone was sounded if the response was correct, and a longer low-pitched tone was sounded if the response was incorrect. Each stimulus was response terminated and displayed for up to 5 s. The task consisted of four 40-min consecutive daily sessions, each containing 500 trials. At the end of each session, the percentage of correct responses appeared on the monitor. In the instructions, participants were told that about half of the stimuli came from category A and half from category B. They were also told that, as is common in real life, the categories were not clearly separated (e.g., some plants look like bushes, even though they are actually trees), so the best performance an expert could achieve was about 75%.

## Results

Some of the results from this experiment are summarized in Table 2. All participants in both experimental conditions achieved between 66% and 75% correct during their best experimental session. In most cases, this was the last (i.e., fourth) session, but 4 participants achieved highest accuracy in an earlier session. For every participant in both conditions, we fit a number of different decision bound models to the data collected from the final experimental session. All models were variants of the general linear classifier described by Maddox and Ashby (1993), which assumes the use of a linear decision bound. In the present application, this model had three free parameters: the slope and intercept of the decision bound (or, alternatively,  $\theta$  and  $\tau$ ) and the variance of perceptual and criterial noise. Special cases of the model assume the use of a verbal rule or the optimal decision bound. The optimal model has only one parameter (noise variance), whereas the verbal rule models have

two parameters (noise variance and the value of the response criterion on the relevant stimulus dimension).

The parameters were estimated using an iterative maximum likelihood procedure, and the goodness-of-fit statistic ("A" information criterion [AIC]) was  $AIC = 2r - 2 \log L$ , where  $r$  is the number of free parameters and  $L$  is the likelihood of the model given the data (Akaike, 1974; Takane & Shibayama, 1992). The AIC statistic penalizes a model for extra free parameters in such a way that the smaller the AIC, the closer a model is to the "true model," regardless of the number of free parameters. Thus, to find the best model among a given set of competitors, one simply computes an AIC value for each model and chooses the model associated with the smallest AIC value.

In the shallow condition, the model that assumes a verbal rule on orientation fit best in three cases (Participants 1, 3, and 4), and the general linear classifier fit best in two cases. In the steep condition, the general linear classifier fit best in five cases, the model that assumes a verbal rule on length fit best in two cases (Participants 2 and 8), and the optimal classifier fit best in one case (Participant 5). In both conditions, however, the mean AIC value across participants was smallest for the general linear classifier by a substantial amount. In addition, the null hypothesis that the general linear classifier is correct could not be rejected for any participant in either condition (using the data from the final session;  $p > .20$ ,  $n = 497$ ). Thus, overall, the fits of the general linear classifier were excellent. Furthermore, the general linear classifier was the only model that provided consistently good fits; the null hypotheses that participants used the optimal bound and that they used a verbal rule were both rejected for some participants ( $p < .20$ ). The poor performance of the optimal classifier relative to the general linear classifier allowed us to convincingly reject one of the four prominent possibilities that we considered earlier, namely, that participants will use the optimal rule in these conditions.

Table 2 lists the decision bound of the best-fitting general linear classifier, together with its associated AIC value. In the

shallow condition, the slope of the optimal bound was 0.6. In contrast, the slopes of all best-fitting bounds were less than optimal (two were actually negative), with a mean of 0.16. In the steep condition, two bounds were essentially vertical lines, two had slopes steeper than optimal but shallower than the nearby verbal rule on length, and four had slopes shallower than optimal. The mean slope was 2.09, and the slope of the optimal bound was 1.67.<sup>14</sup> Figure 13 shows the responses of typical participants in each condition during the last experimental session. Also shown are the optimal and best-fitting bounds (the dotted and solid lines, respectively). Thus, in both conditions, the mean slopes of the best-fitting general linear classifier supported the COVIS prediction of an asymptotic bias toward the nearby verbal rule.

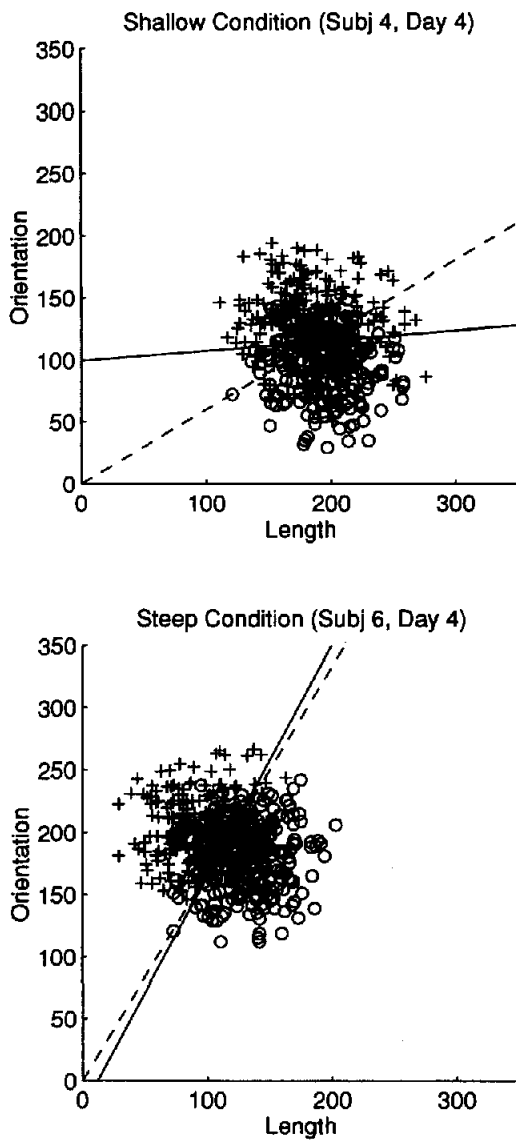


Figure 13. Responses of typical participants (Subj) in the shallow and steep conditions. The dotted lines denote the optimal bounds, and the solid lines denote the best-fitting bounds. A plus sign indicates an "A" response and a circle indicates a "B" response.

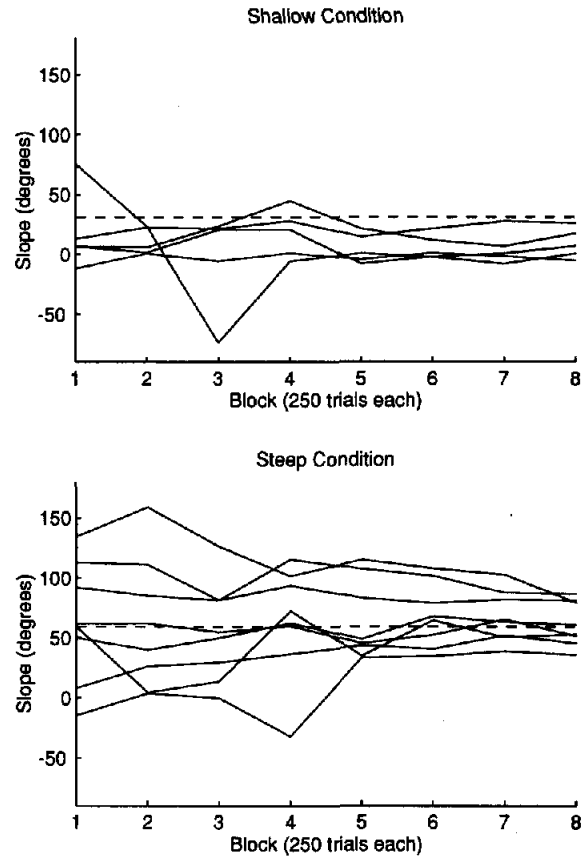


Figure 14. Slopes of the best-fitting bounds from the general linear classifier when it is fit separately to each 250-trial block of every participant.

As a more rigorous test of the four hypotheses discussed in the introduction to this experiment, we fit the general linear classifier (using maximum likelihood estimation) to each consecutive set of 250 trials (half an experimental session). The slopes of these best-fitting bounds are shown in Figure 14. In the shallow condition, all slopes were shallower than optimal in Blocks 5 through 8. In the steep condition, the results were more equivocal. In Blocks 6 and 7, five of eight best-fitting bounds were steeper than optimal, and, in Block 8, four of eight were steeper than optimal.

Discussion

Overall, the results of this experiment support the COVIS prediction of an asymptotic bias toward the nearby verbal rule, because the mean slopes of the best-fitting general linear classifier displayed this bias in both conditions. In the shallow condition, every participant showed this bias. In the steep condition,

<sup>14</sup> Because of the discontinuity associated with vertical lines in slope-intercept form, we computed mean slope in the steep condition using the  $(\theta, \tau)$  parameterization described earlier. The mean value of  $\theta$  was then converted to a slope via the equation  $\text{slope} = \tan(\theta)$ .

however, there were individual differences, because 2 participants used bounds that were considerably shallower than optimal.<sup>15</sup>

Our dynamical-level description of the COVIS network led us to predict an asymptotic bias toward the nearby verbal rule in this experiment. This argument depended on the assumption that stimulus, perceptual, and criterial noise would prevent the system weight on the verbal system (i.e.,  $W_V$ ) from being driven to 0. Even so, if the weight on the implicit system dominates the weight on the verbal system (i.e., at asymptote), then the bias may not be noticeable. To investigate this question, we simulated the performance of the COVIS network in 100 replications of the two conditions of this experiment (i.e., the shallow and steep conditions). The network was presented with exactly the same stimuli as the participants in this experiment, and, initially, we assumed that the two verbal rule types were equally salient. As in our earlier simulations, the initial bias toward verbal rules was incorporated into the network by setting  $W_V(0) = .99$  and  $W_I(0) = .01$ . The perceptual and criterial noise standard deviation was set equal to the mean of the general linear classifier's estimates of this parameter across all participants in the experiment. The only free parameters to be assigned were a set of learning rates (see Appendix A for details). However, these rates had little effect on asymptotic performance (as long as they were in the appropriate range over which learning occurred). Therefore, an intermediate set of values of the learning rates was chosen and used in all simulations.

Some of the statistics summarizing the performance of the network during the final experimental session (i.e., the last 500 trials of the 2,000-trial sequence) are shown in Table 3. Although optimal accuracy was 76%, note that the accuracy of the network was 69.6% across the two conditions. As a comparison, the participants in the shallow condition averaged 67% correct, and the participants in the steep condition averaged 70.9% correct. The average across both conditions was 69.2% correct, so the accuracy of the network was virtually identical to that of a typical human participant.

The general linear classifier was fit to the data from the final experimental session separately for each of the 100 replications. The mean AIC score across both conditions was 381 (see Table 3). In contrast, the mean AIC of the general linear classifier across all 13 participants in both conditions was 374. Thus, the general linear classifier accounts for the data from the COVIS connectionist network about as well as it accounts for the human data. The mean best-fitting bounds for the two conditions, which are described in Table 3, indicate a significant bias toward the nearby verbal rule. Specifically, the best-fitting bounds were shallower than optimal in the shallow condition and steeper than optimal in the steep condition. An examination of the asymptotic system weights reveals the source of this bias. Over all simulations, the mean asymptotic values of the system weights were  $W_I(2,000) = .59$  and  $W_V(2,000) = .41$  (see Table 3). Thus, the implicit system overcame the initial bias toward verbal rules, even though it was never able to completely eliminate the participation of the verbal system. Even after 2,000 trials of training, the network was still emitting the response suggested by the verbal system on a significant proportion of trials. Recall that optimal accuracy in this task was 76% and that the best verbal rule could achieve 72.5% correct. This small difference is

Table 3  
Summary Statistics for Simulations of the COVIS  
Connectionist Network in the Steep and Shallow Conditions

Statistic	Shallow condition	Steep condition
Mean percentage correct	69.4	69.7
Mean $W_V$	0.44	0.38
Mean $W_I$	0.56	0.62
General linear classifier statistic		
Mean AIC	385	376
Mean bound	$x_2 = .32x_1 + 51$	$x_2 = 2.6x_1 - 109$

Note. AIC = "A" information criterion.

enough for the implicit system weight to overtake and exceed the verbal system weight, but only by a small margin. Apparently, the verbal system is rewarded at a high enough rate to maintain an active presence throughout the course of the experiment.

These results confirm that, when the alternative verbal rule types are equally salient, the COVIS connectionist network predicts a significant bias in the observable category bound in the direction of a nearby verbal rule.<sup>16</sup> Thus, as it should, the COVIS

<sup>15</sup> Why is there individual difference in the steep condition but not in the shallow condition? And why is there more variability in the bounds used by participants in the steep condition than in the shallow condition? An experiment by Alfonso-Reese (1996) suggests an answer to these questions. Alfonso-Reese (1996) presented a reference line of fixed length and orientation on one side of a stimulus display and a second target line of random length and orientation on the other side. The participant's task was to adjust the length and orientation of the target line until it exactly matched the length and orientation of the reference line. Every participant performed the same adjustment procedure for each reference line many times. The means and variances (and covariance) of the adjusted target lines give an estimate of the perceptual noise associated with each of the reference lines. In the shallow condition, the amount of perceived variability in length was consistently greater than the amount of perceived variability in orientation. This difference decreases the slope of the optimal bound, but Alfonso-Reese (1996) showed that the observed suboptimality was much greater than predicted by this purely perceptual argument. Another possibility is that because perceptual noise lowers the accuracy of verbal rules, participants in the shallow condition were drawn to a verbal rule on orientation because of the smaller perceived variability on orientation. In the steep condition, variability depended on the length of the reference line. For the shorter reference line, variability was greater on the orientation dimension, but for the longer reference line, variability was greater on the length dimension. Thus, in the steep condition, participants who focused more on the shorter lines might be drawn to a verbal rule on length, and participants who focused more on the longer lines might be drawn to a verbal rule on orientation. This hypothesis accounts for the smaller mean bias we observed in our steep condition and also for the larger individual difference.

<sup>16</sup> We also ran simulations to test the unequal salience hypothesis suggested by the Alfonso-Reese (1996) data. Incorporating such salience differences into the COVIS network allowed it to account for the smaller mean bias in the steep condition and the greater individual difference.

connectionist network makes the same qualitative prediction as the dynamical version of COVIS.

### Alternative Models of Category Learning

An important distinguishing feature that sets COVIS apart from other models of category learning is the assumption that there are multiple categorization systems that compete against each other throughout learning (but see Erickson & Kruschke, *in press*). The major argument in favor of multiple systems is the overwhelming evidence that people treat categorization rules that can be verbalized in a fundamentally different way from rules that cannot be verbalized. In general, models postulating a unitary process fail to account for such qualitative differences. An important exception, however, would be a model postulating a single system similar to the implicit system proposed in this article, but in which a bias toward verbal rules was incorporated via a prior probability distribution over the set of initial decision bounds. Such a model has not been fully tested, but the notion of prior biases in categorization has been explored (e.g., Choi, McDaniel, & Busemeyer, 1993).

To take a concrete example, consider a model in which a bound is chosen for use on the first trial of a new category learning experiment by sampling from a prior probability distribution. Suppose that subsequent learning follows a process of gradient descent and the prior distribution specifies that, with a probability of 1, an initial bound is chosen that corresponds to a verbal rule. Such a model might account for an observed bias toward verbal rules by hypothesizing that learning was not complete at the end of the experiment, so the participants' final bounds still reflected the prior distribution. The model also provides an account of other important category learning effects. For example, it accounts for the widely reported result that people prefer unidimensional rules in unsupervised tasks by hypothesizing that the absence of feedback prevents learning, so the rules chosen by participants simply mirror their prior distributions.

Despite its attraction, there are problems with the hypothesis that people treat verbal decision rules differently than nonverbal rules only because of a prior bias toward verbal rules. First, the results shown in Figure 14 indicate that the performance of virtually all participants in our experiment stabilized long before the end of their final experimental session. If the results were due completely to prior biases, then the bias toward verbal rules should have decreased with training until it eventually disappeared. There is little evidence for this prediction in our data. Second, it is not at all clear how the prior bias hypothesis could account for the neuropsychological results discussed earlier in this article.

A third bit of evidence that is problematic for the prior bias hypothesis comes from the Ashby et al. (1997) unsupervised study. Recall that in each of the four conditions of this study, two widely separated categories were created from lines that varied in length and orientation. In two of the conditions, a verbal (i.e., unidimensional) rule was optimal, and, in two conditions, the optimal rule was nonverbal. Participants used verbal rules in all four conditions. With respect to the prior bias hypothesis, the most important result was that there was no evidence of learning in the implicit conditions but strong evidence of learning in the verbal conditions. In both verbal conditions, a number of participants initially tried a verbal rule of the incor-

rect type, and then, in every case, they spontaneously switched to the correct verbal rule. A model that hypothesizes only an initial prior bias toward verbal rules predicts no such learning or predicts that an equal amount of learning must occur in the implicit conditions.

There is abundant evidence that prior biases do affect category learning. For example, Salatas and Bourne (1974) investigated ease of learning for eight different complex verbal rules and found substantial differences in difficulty. COVIS does not deny the existence of prior biases. In fact, prior biases are formalized in COVIS through its salience parameters, that is, the  $Z_i(0)$ . For example, with categories composed of rectangles, COVIS displays a prior bias toward using a verbal rule on shape because the salience of a shape rule is assumed to be greater than the salience of rules on height, width, or size (e.g., Krantz & Tversky, 1975).

The prior bias hypothesis is a very different model of category learning from COVIS. Much more similar to COVIS is a multiple systems model that postulates separate rule-based and instance- or exemplar-based categorization systems (Brooks, 1978; Erickson & Kruschke, *in press*). As mentioned previously, exemplar-based and procedural-learning-based (decision bound) models have both been impressively successful at accounting for asymptotic categorization performance, and formal modeling attempts have been unable to distinguish empirically between the two (e.g., Maddox & Ashby, 1993; McKinley & Nosofsky, 1995). We favor the procedural-learning-based account of implicit category learning, primarily because of the neuropsychological data. As reviewed earlier in this article, the neuropsychological evidence is nicely consistent with a striatal-based procedural learning categorization system. Currently, there is no accepted neuropsychological version of exemplar theory, and there are several significant challenges that must be met before such a theory is possible. Given current understanding of the neurobiology of memory (e.g., Schacter & Tulving, 1994a), it seems clear that the hippocampus must be a key structure in any neuropsychological account of exemplar-based category learning. For example, current theories of amnesia assume that damage to hippocampal structures impairs the consolidation of instance-based memories (i.e., rather than their retrieval; e.g., Gluck & Myers, 1997; McClelland, McNaughton, & O'Reilly, 1995; Polster, Nadel, & Schacter, 1991; Squire & Alvarez, 1995). So, according to this view, patients with medial temporal lobe amnesia should be unable to lay down the memory traces that exemplar models assume are necessary for correct categorization. Therefore, one challenge for a neuropsychological version of exemplar theory is to account for the fact that amnesic patients are unimpaired at category learning, at least during the first 50 trials of training<sup>17</sup> (Knowl-

<sup>17</sup> Higham and Vokey (1994) argued that instance-based memory is intact in amnesic patients and that their only problem is that they have lost conscious access to those memories. As already noted, this retrieval hypothesis contradicts current theories of amnesia (e.g., McClelland et al., 1995; Polster et al., 1991; Squire & Alvarez, 1995). For example, if the retrieval hypothesis is correct, then anterograde and retrograde amnesia should be equally likely and equally severe in medial temporal lobe amnesia. The data, however, show that anterograde amnesia is more likely and generally much more severe than retrograde amnesia (e.g., Zola, 1997).



ton et al., 1992, 1994; Knowlton & Squire, 1993; Kolodny, 1994). A second challenge is to account for the category learning deficits that have been observed in patients with striatal dysfunction (e.g., Parkinson's and Huntington's disease). These data are problematic to exemplar theories, because the striatum is not usually implicated in instance-based memory.

Perhaps the most promising current foundation on which to build a neuropsychological version of exemplar theory is Gluck and Myers's (1993) computational model of hippocampal function. Although applications of this model have focused on classical Pavlovian conditioning, Gluck, Oliver, and Myers (1996) did show that the model could account for the Knowlton et al. (1994) category learning data with amnesic patients.

### Conclusions

In this article, we had a number of goals. First, following Squire (1992), we have argued that a cognitive form of procedural learning plays a significant role in perceptual categorization, especially in tasks in which the optimal rule is difficult to verbalize. There is now encouraging evidence in support of this hypothesis. First, the decision bound model has had impressive success in accounting for categorization data (e.g., Ashby & Gott, 1988; Ashby & Lee, 1991; Ashby & Maddox, 1990, 1992; Maddox & Ashby, 1993). Second, patients with Parkinson's or Huntington's disease, who are impaired in terms of procedural learning, have recently been shown to also have category learning deficits (Knowlton, Mangels, & Squire, 1996; Knowlton, Squire, et al., 1996). Finally, amnesic patients, who have normal procedural learning and memory but an impaired instance-based memory, show relatively normal category learning (Knowlton et al., 1992; Knowlton & Squire, 1993; Kolodny, 1994).

Second, we have argued for the existence of multiple category learning systems. Because it is now commonly believed that there are multiple memory systems (e.g., Schacter & Tulving, 1994b), an important advantage of COVIS is that it ties research in categorization firmly to current work in memory. In fact, the existence of multiple memory systems is itself suggestive of multiple categorization systems. Any category learning system requires memory. If there is a categorization system that uses one type of memory, then why should there not also be a categorization system that uses some other type? COVIS assumes that the two most important categorization systems use semantic and procedural memory systems, but we do not rule out the possibility that other systems may also be involved (i.e., an exemplar-based system and a perceptual priming system).

A third goal of this article was to identify some of the neuropsychological structures and pathways that mediate categorization. In particular, we have argued for important roles of the caudate nucleus and the cingulate–prefrontal cortex. Our interest in the caudate was motivated by Squire (1992). In COVIS, the caudate computes the decision of the implicit system, and it may also learn the response criteria of the various verbal rule types. In contrast, the cingulate–prefrontal cortex selects the type of verbal rule to be used on each trial. As such, we attribute to it the same basic functions as Posner and Petersen (1990). One advantage of this neuropsychological approach is that it allows predictions to be derived about many special populations. For example, in this article, we have derived specific predictions

for children and elderly people; for patients suffering from frontal lesions, Parkinson's disease, Huntington's disease, amnesia, and major depression; and for nonhuman animals. Although most of these predictions have not been fully tested, preliminary evidence supports the COVIS predictions.

Finally, a fourth goal of this article was to improve on current formal models of category learning. Toward this end, we have proposed a model that describes the behavior of COVIS at the level of trial-by-trial learning data (i.e., the algorithmic level) and at a global (i.e., computational) level. The global-level model is useful for investigating the effects of a number of important variables on categorization performance. In particular, the model makes predictions about the effects of verbal rules on performance in tasks in which the optimal rule is nonverbal. It also describes the effects of varying the quality of feedback and the salience of alternative verbal rule types. The connectionist version of COVIS makes predictions about trial-by-trial learning. We believe that these two models, together with the neuropsychological version of COVIS, provide a powerful theory of categorization behavior in a wide variety of tasks and in a diversity of populations (i.e., both human and nonhuman).

### References

- Ågren, H., Reibring, L., Hartvig, P., Tedroff, J., Bjurling, P., Lundqvist, H., & Långström, B. (1993). Monoamine metabolism in human prefrontal cortex and basal ganglia. PET studies using [ $\beta$ - $^{11}$ C] 5-hydroxytryptophan and [ $\beta$ - $^{11}$ C] L-Dopa in healthy volunteers and patients with unipolar major depression. *Depression, 1*, 71–81.
- Ahn, W. K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16*, 81–121.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia cortex. *Annual Review of Neuroscience, 9*, 357–381.
- Alexander, M. E., & Wickens, J. R. (1993). Analysis of striatal dynamics: The existence of two models of behaviour. *Journal of Theoretical Biology, 163*, 413–438.
- Alfonso-Reese, L. (1996). *Dynamics of category learning*. Unpublished doctoral dissertation, University of California, Santa Barbara.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Erlbaum.
- Arnauld, E., Jeantet, Y., Arsaut, J., & Demotes-Mainard, J. (1996). Involvement of the caudal striatum in auditory processing: c-fos response to cortical application of picrotoxin and to auditory stimulation. *Brain Research: Molecular Brain Research, 41*, 27–35.
- Aschkenasy, J. R., & Odom, R. D. (1982). Classification and perceptual development: Exploring issues about integrality and differential sensitivity. *Journal of Experimental Child Psychology, 34*, 435–448.
- Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Erlbaum.
- Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39*, 216–233.
- 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*, 33–53.
- Ashby, F. G., Isen, A. M., & Turken, A. U. (1998). *A neuropsychological*

- theory of positive affect and its influences on cognition*. Manuscript submitted for publication.
- Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*, 150–172.
- Ashby, F. G., & Lee, W. W. (1992). On the relationship among identification, similarity and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General*, *121*, 385–393.
- Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of perceptual theory* (pp. 369–399). Amsterdam: Elsevier.
- Ashby, F. G., & Maddox, W. T. (1989, November). *Toward a theory of natural categorization*. Paper presented at the Thirtieth Annual Meeting of the Psychonomic Society, Atlanta, GA.
- Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 598–612.
- 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.
- Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372–400.
- Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and integrality in classification. *Journal of Mathematical Psychology*, *38*, 423–466.
- Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Handbook of perception and cognition: Measurement, judgment, and decision making* (pp. 251–301). New York: Academic Press.
- Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*, 124–150.
- Ashby, F. G., Queller, S., & Berretty, P. M. (1997). *On the dominance of unidimensional rules in unsupervised categorization*. Manuscript submitted for publication.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Balakrishnan, J. D., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 615–633.
- Beatty, J. (1995). *Principles of behavioral neuroscience*. Dubuque, IA: Brown & Benchmark.
- Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Hoeber.
- Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowiak, R. S. J., & Dolan, R. J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia*, *33*, 907–922.
- Beninger, R. J. (1983). The role of dopamine in locomotor activity and learning. *Brain Research*, *287*, 173–196.
- Berg, E. A. (1948). A simple objective test for measuring flexibility and thinking. *Journal of General Psychology*, *39*, 15–22.
- Berger, H. J. C., van Hoof, J. J. M., van Spaendonck, K. P. M., Horstink, M. W. I., van den Bercken, J. H. L., Jaspers, R., & Cools, A. R. (1989). Haloperidol and cognitive shifting. *Neuropsychologia*, *27*, 629–639.
- Berman, K. F., Doran, A. R., Pickard, D., & Weinberger, D. R. (1993). Is the mechanism of prefrontal hypofunction in depression the same as in schizophrenia? Regional cerebral blood flow during cognitive activation. *British Journal of Psychiatry*, *162*, 183–192.
- Berns, G. S., & Sejnowski, T. J. (1996). How the basal ganglia make decisions. In A. R. Damasio, H. Damasio, & Y. Christen (Eds.), *Neurobiology of decision-making* (pp. 101–113). Berlin: Springer-Verlag.
- Bliss, T. V. P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology (London)*, *232*, 331–356.
- Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.
- Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain*, *111*, 323–345.
- Brown, V. J., Desimone, R., & Mishkin, M. (1995). Responses of cells in the tail of the caudate nucleus during visual discrimination learning. *Journal of Neurophysiology*, *74*, 1083–1094.
- Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.
- Busemeyer, J. R., & Myung, I. J. (1987). Resource allocation decision making in an uncertain environment. *Acta Psychologica*, *66*, 1–19.
- Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, *121*, 177–194.
- Busemeyer, J. R., Swenson, K. N., & Lazarte, A. (1986). An adaptive approach to resource allocation. *Organizational Behavior and Human Decision Processes*, *38*, 318–341.
- Calne, D. B., & Langston, J. W. (1983). Aetiology of Parkinson's disease. *Lancet*, *8365*, 1457–1459.
- Choi, S., McDaniel, M. A., & Busemeyer, J. R. (1993). Incorporating prior biases in network models of conceptual rule learning. *Memory & Cognition*, *21*, 413–423.
- Chudler, E. H., Sugiyama, K., & Dong, W. K. (1995). Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Research*, *674*, 33–45.
- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 17–30.
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing what. *Science*, *210*, 207–209.
- Collins, B., & Tellier, A. (1994). Differences in conceptual flexibility with age as measured by a modified version of the Visual Verbal Test. *Canadian Journal on Aging*, *13*, 368–377.
- Cook, G. L., & Odom, R. D. (1988). Perceptual sensitivity to dimensional and similarity relations in free and rule-governed classification. *Journal of Experimental Child Psychology*, *45*, 319–338.
- Cook, G. L., & Odom, R. D. (1992). Perception of multidimensional stimuli: A differential-sensitivity account of cognitive processing and development. *Journal of Experimental Child Psychology*, *54*, 213–249.
- Cools, A. R., van den Bercken, J. H. L., Horstink, M. W. I., van Spaendonck, K. P. M., & Berger, H. J. C. (1984). Cognitive and motor shifting aptitude disorder in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *47*, 443–453.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience*, *11*, 2383–2402.
- Crow, T. J. (1980). Molecular pathology of schizophrenia: More than one disease process? *British Medical Journal*, *280*, 66–68.
- Crow, T. J. (1982). Two syndromes in schizophrenia. *Trends in Neuroscience*, *5*, 351–354.
- Curran, T. (1995). On the neural mechanisms of sequence learning. *Psyche*, *2*(12). URL: <http://psyche.cs.monash.edu.au/volume2-1/psyche-95-2-12-sequence-1-curran.html>.

- Damasio, A. R. (1985). Prosopagnosia. *Trends in Neuroscience*, 8, 132–135.
- Darken, C., & Moody, J. (1992). Towards faster stochastic gradient search. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in neural information processing systems* (pp. 1009–1016). San Mateo, CA: Morgan Kaufman.
- Delis, D. C., & Massman, P. J. (1992). The effects of dopamine fluctuation on cognition and affect. In S. J. Huber & J. L. Cummings (Eds.), *Parkinson's disease: Neurobehavioral aspects* (pp. 405–427). New York: Oxford University Press.
- Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, 74, 24–40.
- DiFiglia, M., Pasik, T., & Pasik, P. (1978). A Golgi study of afferent fibers in the neostriatum of monkeys. *Brain Research*, 152, 341–347.
- Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1985). On consciousness in syntactic learning and judgment: A reply to Reber, Allen, and Regan. *Journal of Experimental Psychology: General*, 114, 25–32.
- Eacott, M. J., & Gaffan, D. (1992). Inferotemporal-frontal disconnection: The uncinat fascicle and visual associative learning in monkeys. *European Journal of Neuroscience*, 4, 1320–1332.
- Erickson, M. A., & Kruschke, J. K. (in press). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*.
- Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53, 134–140.
- Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, 18, 500–549.
- Estes, W. K. (1995). Response processes in cognitive models. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in text comprehension* (pp. 51–71). Hillsdale, NJ: Erlbaum.
- Estrada, C. A., Isen, A. M., & Young, M. J. (1997). Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organizational Behavior and Human Decision Processes*, 18, 285–299.
- Foard, C. F., & Kemler-Nelson, D. G. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General*, 113, 94–111.
- Franke, P., Maier, W., Hardt, J., Frieboes, R., Lichtermann, D., & Hain, C. (1993). Assessment of frontal lobe functioning in schizophrenia and unipolar major depression. *Psychopathology*, 26, 76–84.
- Freund, T. F., Powell, J. F., & Smith, A. D. (1984). Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience*, 13, 1189–1215.
- Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. J. (1991). Investigating a network model of word generation with positron emission tomography. *Proceedings of the Royal Society of London (B)*, 244, 101–106.
- Frith, C. D., Friston, K. J., Liddle, P. F., & Frackowiak, R. S. J. (1991a). A PET study of word finding. *Neuropsychologia*, 29, 1137–1148.
- Frith, C. D., Friston, K. J., Liddle, P. F., & Frackowiak, R. S. J. (1991b). Willed action and the prefrontal cortex in man: A study with PET. *Proceedings of the Royal Society of London (B)*, 244, 241–246.
- Gabrieli, J. (1995). Contribution of the basal ganglia to skill learning and working memory in humans. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 277–294). Cambridge, MA: Bradford.
- Gaffan, D., & Eacott, M. J. (1995). Visual learning for an auditory secondary reinforcer by macaques is intact after uncinat fascicle section: Indirect evidence for the involvement of the corpus striatum. *European Journal of Neuroscience*, 7, 1866–1871.
- Gaffan, D., & Harrison, S. (1987). Amygdalotomy and disconnection in visual learning for auditory secondary reinforcement by monkeys. *Journal of Neuroscience*, 7, 2285–2292.
- Gamble, E., & Koch, C. (1987). The dynamics of free calcium in dendritic spines in response to repetitive synaptic input. *Science*, 236, 1311–1315.
- Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Gholson, B., Levine, M., & Phillips, S. (1972). Hypotheses, strategies, and stereotypes in discrimination learning. *Journal of Experimental Child Psychology*, 13, 423–446.
- Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*, 3, 491–516.
- Gluck, M. A., & Myers, C. E. (1997). Psychobiological models of hippocampal function in learning and memory. *Annual Review of Neuroscience*, 48, 481–514.
- Gluck, M. A., Oliver, L. M., & Myers, C. E. (1996). Late-training amnesic deficits in probabilistic category learning: A neurocomputational analysis. *Learning and Memory*, 3, 326–340.
- Goldman-Rakic, P. S. (1995a). Cellular basis of working memory. *Neuron*, 14, 477–485.
- Goldman-Rakic, P. S. (1995b). Toward a circuit model of working memory and the guidance of voluntary motor action. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 131–148). Cambridge, MA: Bradford.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178–200.
- Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, 7, 497–510.
- Hayashi, A., Kagamihara, Y., Nakajima, Y., Narabayashi, H., Okuma, Y., & Tanaka, R. (1988). Disorder in reciprocal innervation upon initiation of voluntary movement in patients with Parkinson's disease. *Experimental Brain Research*, 70, 437–440.
- Heimer, L. (1995). *The human brain and spinal cord* (2nd ed.). New York: Springer-Verlag.
- Heindel, W. C., Butters, N., & Salmon, D. P. (1983). Impaired learning of a motor skill in patients with Huntington's disease. *Behavioral Neuroscience*, 102, 141–147.
- Heindel, W. C., Salmon, D. P., Shults, C. W., Walicke, P. A., & Butters, N. (1989). Neuropsychological evidence for multiple implicit memory systems: A comparison of Alzheimer's, Huntington's, and Parkinson's disease patients. *Journal of Neuroscience*, 9, 582–587.
- Hemmings, H. C., Jr., Walaas, S. I., Ouimet, C. C., & Greengard, P. (1987). Dopaminergic regulation of protein phosphorylation in the striatum: DARPP-32. *Trends in Neuroscience*, 10, 377–383.
- Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.
- Higham, P. A., & Vokey, J. R. (1994). Recourse to stored exemplars is not necessarily explicit: A comment on Knowlton, Ramus, & Squire (1992). *Psychological Science*, 5, 59.
- Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: Onset, progression, and mortality. *Neurology*, 17, 427–442.
- Huang, Y. Y., & Kandel, E. R. (1995). D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CA1 region of the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 2446–2450.
- Hyman, S. E., & Nestler, E. J. (1993). *The molecular foundations of psychiatry*. Washington, DC: American Psychiatric Press.
- Imai, S., & Garner, W. R. (1965). Discriminability and preference for

- attributes in free and constrained classification. *Journal of Experimental Psychology*, 69, 596–608.
- Jackson, G. M., Jackson, S. R., Harrison, J., Henderson, L., & Kennard, C. (1995). Serial reaction time learning and Parkinson's disease: Evidence for a procedural learning deficit. *Neuropsychologia*, 33, 577–593.
- Jahanshahi, M., Brown, R. G., & Marsden, C. (1992). The effect of withdrawal of dopaminergic medication on simple and choice reaction time and the use of advance information in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 55, 1168–1176.
- Jaspers, R. M. A., de Vries, T. J., & Cools, A. R. (1990a). Effect of intrastriatal apomorphine on changes in switching behaviour induced by the glutamate agonist AMPA injected into the cat caudate nucleus. *Behavioral Brain Research*, 37, 247–254.
- Jaspers, R. M. A., de Vries, T. J., & Cools, A. R. (1990b). Enhancement in switching motor patterns following local application of the glutamate agonist AMPA into the cat caudate nucleus. *Behavioral Brain Research*, 37, 237–246.
- Javoy-Agid, F., & Agid, Y. (1980). Is the mesocortical dopaminergic system involved in Parkinson's disease? *Neurology*, 30, 1326–1330.
- Jitsumori, M. (1993). Category discrimination of artificial polymorphous stimuli based on feature learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 244–254.
- Josiassen, R. C., Curry, L. M., & Mancall, E. L. (1983). Development of neuropsychological deficits in Huntington's disease. *Archives of Neurology*, 40, 791–796.
- Keele, S. W., & Curran, T. (in press). On the modularity of sequence learning systems in humans. In E. Covey (Ed.), *Neural representation of temporal patterns*. New York: Plenum.
- Kemler, D. G. (1978). Patterns of hypothesis testing in children's discriminative learning: A study of the development of problem-solving strategies. *Developmental Psychology*, 14, 653–673.
- Knopman, D., & Nissen, M. J. (1991). Procedural learning is impaired in Huntington's disease: Evidence from the serial reaction time task. *Neuropsychologia*, 29, 245–254.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, 273, 1399–1402.
- Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science*, 3, 172–179.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category level knowledge. *Science*, 262, 1747–1749.
- Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 79–91.
- Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 169–181.
- Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory*, 1, 106–120.
- Knowlton, B. J., Squire, L. R., Paulsen, J. S., Swerdlow, N. R., Swenson, M., & Butters, N. (1996). Dissociations within nondeclarative memory in Huntington's disease. *Neuropsychology*, 10, 538–548.
- Köhler, W. (1925). *The mentality of apes*. New York: Harcourt, Brace.
- Kolb, B., & Whishaw, I. Q. (1990). *Fundamentals of human neuropsychology* (3rd ed.). New York: Freeman.
- Kolodny, J. A. (1994). Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. *Psychological Science*, 5, 164–169.
- Kramer, A. F., Humphrey, D. G., Larish, J. F., & Logan, G. D. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging*, 9, 491–512.
- Krantz, D. J., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology*, 12, 4–34.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Laming, D. (1973). *Mathematical psychology*. London: Academic Press.
- Langston, J. W., & Palfreman, J. (1995). *The case of the frozen addicts*. New York: Pantheon Books.
- Lidow, M. S., Goldman-Rakic, P. S., Gallager, D. W., & Rakic, P. (1991). Distribution of dopaminergic receptors in the primate cerebral cortex: Quantitative autoradiographic analysis using [3H]Raclopride, [3H]spiperone, and [3H]SCH23390. *Neuroscience*, 40, 657–671.
- Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, 72, 94–104.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Lynch, G., Larson, J., Kelso, S., Barrionuero, G., & Schottler, F. (1983). Intracellular injections of EGTA block induction of hippocampal long-term potentiation. *Nature*, 266, 737–739.
- MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J., & Barker, J. L. (1986). NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spine cord neurones. *Nature*, 321, 519–522.
- Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Erlbaum.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, 53, 49–70.
- Malmö, H. P. (1974). On frontal lobe function: Psychiatric patient controls. *Cortex*, 10, 231–237.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1083–1100.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.
- McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, 107, 3–22.
- McDonald, R. J., & White, N. M. (1994). Parallel information processing in the water maze: Evidence for independent memory systems involving dorsal striatum and hippocampus. *Behavioral and Neural Biology*, 61, 260–270.
- 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.
- 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*, 1, 335–368.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, 19, 242–279.
- Miller, R., & Wickens, J. R. (1991). Corticostriatal cell assemblies in

- selective attention and in representation of predictable and controllable events. *Concepts in Neuroscience*, 2, 65–95.
- Mishkin, M., Malamut, B., & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, J. L. McGaugh, & N. M. Weinberger (Eds.), *Neurobiology of human learning and memory* (pp. 65–77). New York: Guilford Press.
- Mishkin, M., Ungerleider, L. G., & Mako, K. A. (1983). Object vision and spatial vision: Two critical pathways. *Trends in Neuroscience*, 6, 414–417.
- Monaghan, D. T., & Cotman, C. W. (1985). Distribution of N-methyl-D-aspartate sensitive L-[3H] glutamate binding sites in rat brain. *Journal of Neuroscience*, 5, 2909–2919.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Myung, I. J., & Busemeyer, J. R. (1989). Criterion learning in a deferred decision making task. *American Journal of Psychology*, 102, 1–16.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1–32.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 211–233.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53–79.
- Oakley, D. A. (1981). Brain mechanisms of mammalian memory. *British Medical Bulletin*, 37, 175–180.
- Owen, A. M., Roberts, A. C., Hodges, J. R., Summers, B. A., Polkey, C. E., & Robbins, T. W. (1993). Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain*, 116, 1159–1175.
- Packard, M. G., Hirsch, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience*, 9, 1465–1472.
- Packard, M. G., & McGaugh, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience*, 106, 439–446.
- Packard, M. G., & White, N. M. (1991). Dissociation of hippocampus and caudate nucleus memory systems by post-training intracerebral injection of dopamine agonists. *Behavioral Neuroscience*, 105, 295–306.
- Parkin, A. J., & Lawrence, A. (1994). A dissociation in the relation between memory tasks and frontal lobe tests in the normal elderly. *Neuropsychologia*, 32, 1523–1532.
- Pascual-Leon, A., Grafman, J., & Hallett, M. (1994). Modulation of cortical motor output maps during development of implicit and explicit knowledge. *Science*, 263, 1287–1289.
- Patterson, T. A., & Rose, S. P. R. (1992). Memory in the chick: Multiple cues, distinct brain locations. *Behavioral Neuroscience*, 106, 465–470.
- Perruchet, P., & Amorim, M. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 785–800.
- Pessin, M. S., Snyder, G. L., Halpain, S., Girault, J.-A., Aperia, A., & Greengard, P. (1994). DARPP-32/Protein Phosphatase-1/Na<sup>+</sup>/K<sup>+</sup> ATPase System: A mechanism for bidirectional control of cell function. In K. Fuxe, L. F. Agnati, B. Bjelke, & D. Ottoson (Eds.), *Trophic regulation of basal ganglia* (pp. 43–57). New York: Elsevier Science.
- Peterson, S. E., Fox, P. T., Posner, M. I., Mintan, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature*, 331, 585–588.
- Plaut, D., Nowlan, S., & Hinton, G. (1986). *Experiments on learning by back-propagation* (Techn. Rep. CMU-CS-86-126). Pittsburgh, PA: Department of Computer Science, Carnegie Mellon University.
- Poldrack, R. A., Prabhakaran, V., Seger, C. A., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997, November). *Learning to classify probabilistically: Changes in cortical activity revealed by functional MRI*. Paper presented at the 1997 Society for Neuroscience annual meeting, New Orleans, LA.
- Polster, M. R., Nadel, L., & Schacter, D. L. (1991). Cognitive neuroscience analyses of memory: A historical perspective. *Journal of Cognitive Neuroscience*, 3, 95–116.
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.
- Reading, P. J., Dunnett, S. D., & Robbins, T. W. (1991). Dissociable roles of the ventral, medial and lateral striatum on the acquisition and performance of a complex visual stimulus-response habit. *Behavioural Brain Research*, 45, 147–161.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863.
- Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 88–94.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 219–235.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory*, 1, 217–229.
- Ridley, R. M., Baker, H. F., Frith, C. D., Dowdy, J., & Crow, T. (1988). Stereotyped responding on a two-choice guessing task by marmosets and humans treated with amphetamine. *Psychopharmacology*, 95, 560–564.
- Roberts, A. C., De Salvia, M. A., Wilkinson, L. S., Collins, P., Muir, J. L., Everitt, B. J., & Robbins, T. W. (1994). 6-hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the Wisconsin Card Sort Test: Possible interactions with subcortical dopamine. *Journal of Neuroscience*, 14, 2531–2544.
- Robinson, A. L., Heaton, R. K., Lehman, R. A. W., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology*, 48, 605–614.
- 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* (Vol. 1, pp. 318–362). Cambridge, MA: Bradford.
- Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1988). Procedural learning and neostriatal dysfunction in man. *Brain*, 111, 941–959.
- Salatas, H., & Bourne, L. E., Jr. (1974). Learning conceptual rules: III. Processes contributing to rule difficulty. *Memory & Cognition*, 2, 549–553.
- Scatton, B., Rouquier, L., Javoy-Agid, F., & Agid, Y. (1982). Dopamine deficiency in the cerebral cortex in Parkinson's disease. *Neurology*, 32, 1039–1040.
- Schacter, D. L., & Tulving, E. (Eds.). (1994a). *Memory systems of 1994*. Cambridge, MA: MIT Press.
- Schacter, D. L., & Tulving, E. (1994b). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems of 1994* (pp. 1–38). Cambridge, MA: MIT Press.
- Scheel-Kruger, J., & Wilner, P. (1991). The mesolimbic system: Princi-

- ples of operation. In P. Wilner & J. Scheel-Krüger (Eds.), *The mesolimbic dopamine system: From motivation to action* (pp. 559–598). New York: Wiley.
- Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 681–696.
- Seger, C. A., Poldrack, R. A., Prabhakaran, V., & Gabrieli, J. D. E. (1997, November). *An fMRI study of visual concept learning*. Paper presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, PA.
- Selemon, L. D., & Goldman-Rakic, P. S. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *Journal of Neuroscience*, *5*, 776–794.
- Shanks, D. R., Green, R. E. A., & Kolodny, J. A. (1994). A critical examination of the evidence for unconscious (implicit) learning. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance 15: Conscious and nonconscious information processing* (pp. 837–860). Cambridge, MA: MIT Press.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
- Smith, E. E., Patalano, A. L., Jonides, J., & Koeppel, R. A., (1996, November). *PET evidence for different categorization mechanisms*. Paper presented at the 37th Annual Meeting of the Psychonomic Society, Chicago.
- Smith, J. D. (1989). Analytic and holistic processes in categorization. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 297–323). Hillsdale, NJ: Erlbaum.
- Smith, J. D., & Kemler-Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, *113*, 137–159.
- Smith, J. D., Tracy, J. L., & Murray, M. J. (1993). Depression and category learning. *Journal of Experimental Psychology: General*, *122*, 331–346.
- Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology*, *24*, 279–298.
- Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology*, *10*, 502–523.
- Squire, L. R. (1984). *The biology of learning*. Berlin: Springer-Verlag.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*, 143–145.
- Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinion in Neurobiology*, *5*, 169–177.
- Strange, P. G. (1992). *Brain biochemistry and brain disorders*. Oxford, England: Oxford University Press.
- Strick, P. L., Dum, R. P., & Picard, N. (1995). Macro-organization of the circuits connecting the basal ganglia with the cortical motor areas. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 117–130). Cambridge, MA: Bradford.
- Strogatz, S. H. (1994). *Nonlinear dynamics and chaos*. Reading, MA: Addison-Wesley.
- Takane, Y., & Shibayama, T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335–362). Hillsdale, NJ: Erlbaum.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*, 685–688.
- Thompson, L. A. (1994). Dimensional strategies dominate perceptual classification. *Child Development*, *65*, 1627–1645.
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, *40*, 385–398.
- Uylings, H. B. M., & van Eden, C. G. (1990). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Progress in Brain Research*, *85*, 31–63.
- van Domburg, P. H. M. F., & ten Donkelaar, H. J. (1991). *The human substantia nigra and ventral tegmental area*. Berlin: Springer-Verlag.
- Vokey, J. R., & Brooks, L. R. (1992). The salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 328–344.
- Ward, T. B. (1983). Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 103–112.
- Wattenmaker, W. D. (1992). Relational properties and memory-based category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 282–304.
- Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General*, *122*, 227–248.
- Wickens, J. R. (1990). Striatal dopamine in motor activation and reward-mediated learning: Steps towards a unifying model. *Journal of Neural Transmission*, *80*, 9–31.
- Wickens, J. (1993). *A theory of the striatum*. New York: Pergamon Press.
- Wickens, J. R., Alexander, M. E., & Miller, R. (1991). Two dynamic models of striatal function under dopaminergic-cholinergic control: Simulation and analysis of a model. *Synapse*, *8*, 1–12.
- Wiggins, S. (1990). *Introduction to applied nonlinear dynamical systems and chaos*. New York: Springer-Verlag.
- Wilkening, F., & Lange, K. (1987). When is children's perception holistic? Goals and styles in processing multidimensional stimuli. In T. Globerson & T. Zelniker (Eds.), *Cognitive style and cognitive development* (pp. 231–257). Norwood, NJ: Ablex.
- Willingham, D. B., & Koroshetz, W. J. (1993). Evidence for dissociable motor skills in Huntington's disease patients. *Psychobiology*, *21*, 173–182.
- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 1047–1060.
- Wise, R. A. (1982). Neuroleptics and operant behavior: The anhedonia hypothesis. *Behavioral and Brain Sciences*, *5*, 39–88.
- Wise, R. A., & Rompré, P. P. (1989). Brain dopamine and reward. *Annual Review of Psychology*, *40*, 191–225.
- Yeterian, E. H., & Pandya, D. N. (1995). Corticostriatal connections of extrastriate visual areas in rhesus monkeys. *Journal of Comparative Neurology*, *352*, 436–457.
- Zola, S. (1997). Amnesia: Neuroanatomic and clinical aspects. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral neurology and neuropsychology* (pp. 447–461). New York: McGraw-Hill.

(Appendixes follow)

## Appendix A

## The COVIS Connectionist Network

First we discuss applications to the Maddox and Ashby (1993) circle condition (i.e., which generated the simulated learning curve shown in Figure 5b), and to the new experiment reported in this article.

## Architecture

This section describes network architecture not discussed in the text. The values  $x_1$  and  $x_2$  illustrated in Figure 3 are the size and the orientation of the radial line, respectively, in the case of the circle stimuli used by Maddox and Ashby (1993) and length and orientation, respectively, in the case of the line stimuli used in the new experiment described in the text. Perceptual noise was modeled by adding (on every trial) random values  $\epsilon_1$  and  $\epsilon_2$  to  $x_1$  and  $x_2$ , respectively. Thus, the percept on trial  $n$  was  $x_p(n) = [x_1(n) + \epsilon_1(n), x_2(n) + \epsilon_2(n)]'$ . The independent random variables  $\epsilon_1$  and  $\epsilon_2$  were assumed to be normal with mean zero and standard deviation  $\sigma_p$ . In both of these applications, we set  $\sigma_p = 20.69$ , which was the mean of the general linear classifier's estimates of this parameter across all participants in the experiment reported in this article (i.e., described in Figures 10 and 12).

## Training and Testing

For both applications, the following procedures were followed. The network was presented the same stimuli as the participants in the experiment. The presentation order was randomized across each of the simulations (50 in the Maddox and Ashby experiment and 100 in the new experiment). The initial values for the weights of the implicit system,  $a_1(0)$  and  $a_2(0)$ , were chosen by randomly sampling from a uniform (0, 1) distribution. The initial value of the intercept of the implicit bound,  $b_I(0)$ , was set to zero. The initial salience values of the two alternative rule types were each set to 0.5—that is  $Z_1(0) = Z_2(0) = 0.5$ —the mean of the Poisson random variable  $X$  was set to 10,  $\Delta_C$  was set to .04,  $\Delta_E$  was set to .01, and  $\gamma$  was set to 12 in the Maddox and Ashby application and 10 in the application to the new experiment. The initial values of the verbal criteria,  $b_V(0)$  and  $b_{V_2}(0)$ , were chosen by randomly sampling from a uniform (0, 1) distribution. Finally, the initial system weights,  $W_V(0)$  and  $W_I(0)$ , were set to .99 and .01, respectively. None of these parameter settings were manipulated during the course of the model fits.

On each trial, a response was generated according to the algorithm described in the text. After the response was made, three error feedback signals were generated. The first indicated whether the overall system responded correctly, and the other two indicated whether the verbal and implicit systems responded correctly. These signals were used to update the appropriate network weights according to the delta rule. As a means of improving learning, however, a momentum term (Plaut, Nowlan, & Hinton, 1986) and learning rate annealing (Darken & Moody, 1992) were incorporated into the learning rules. In the present applications, we assumed that the verbal and implicit systems both receive their own feedback signal. In the connectionist literature, the problem of how to assign feedback to individual components on the basis of feedback that the overall system receives is called the credit assignment problem (e.g., Hertz, Krogh, & Palmer, 1991). In future applications, it may be necessary to consider more sophisticated solutions to the credit assignment problem, but in the present applications, our simple solution worked well.

The system weights,  $W_V(n)$  and  $W_I(n)$ , and the parameters describing the decision bounds of the implicit and verbal systems were adjusted via the modified delta rule (i.e., the delta rule with a momentum term

and learning rate annealing). The learning rates on trial  $n$  on the overall system weights and on the decision bound parameters of the implicit system were

$$\eta_K(n) = \frac{\eta_K(n-1)}{1 + \frac{r_K}{\delta_K}},$$

where  $K = O$  (for overall) or  $I$  (for implicit),  $\delta_K$  is the decay rate, and  $r_K$  is the number of incorrect responses emitted by system  $K$ . The initial learning rate is given by  $\eta_K(0)$ . The learning rate on the criteria used by the verbal system was

$$\eta_V(n) = \frac{\eta_V(0)}{1 + \frac{r_V}{\delta_V}}.$$

For the implicit and verbal systems, we set  $\eta_K(0) = 0.388$  and  $\delta_K = 42$  (for  $K = I$  or  $V$ ). The momentum values were set at 1.0 for the implicit system and 0.9332 for the verbal system. For the overall system,  $\eta_O(0) = 0.22$  and  $\delta_O = 430$ . No momentum term was used for updating the overall system weights,  $W_V(n)$  and  $W_I(n)$ . The following special considerations were observed. For the verbal system, only the criterion associated with the verbal rule used on trial  $n$  was updated on that trial. The weights of the output node were constrained to be nonnegative and to sum to 1. The vector of decision bound parameters for the implicit system,  $\mathbf{a}(n) = [a_1(n), a_2(n), b_I(n)]'$ , was normalized by dividing by its norm after each update. In the new experiment, the only parameters that were manipulated were three initial learning rates, that is,  $\eta_K(0)$  for  $K = I, V$ , or  $O$ . However, these rates had little effect on asymptotic performance (as long as they were in the appropriate range over which learning occurred). Therefore, an intermediate set of values of the learning rates was chosen and used in all simulations. In the Maddox and Ashby (1993) experiment, the initial learning rates were set to the same values as in the new experiment, and the only parameter manipulated was  $\delta_O$ .

In addition to trial-by-trial feedback, the participants in the experiment received feedback as to their overall percentage correct rate at the end of each 400-trial session (500 in the new experiment). This session-by-session feedback was modeled in the following way. After each session, if the percentage of correct responses was lower than a criterion level (62% in the present simulations), the learning rate parameters were set back to their original values. If accuracy exceeded this criterion, the learning rate parameters were not changed from their current values.

## Smith et al. (1993) Simulations

The basic architecture used to simulate the J. D. Smith et al. (1993) experiments was the same as that just described. However, because the Smith et al. stimuli were four dimensional, there were four input nodes rather than just two; that is, the input vector  $[x_1, x_2]'$  was replaced by  $[x_1, x_2, x_3, x_4]'$ . Also, the learning rules for the verbal and implicit systems were modified because of the qualitative differences between the Smith et al. study and the studies described earlier that used the randomization technique. Specifically, Smith et al. used binary-valued stimuli and categories that did not overlap, and each participant completed only 48 trials. In contrast, the Maddox and Ashby study and the study reported in this article used continuous-valued stimuli and categories that overlapped, and each participant completed more than 1,000 trials. For this reason, the standard perceptron learning rule, without a momentum term or learning rate decay (e.g., Hertz et al., 1991), was

used for both the verbal and implicit systems. Each condition in the Smith et al. (1993) experiments was simulated with 1,000 replications that used the same stimuli and methods as Smith et al. The perseveration parameter  $\gamma$  was set to 0.1 for the nondepressed group and 3.8 for the depressed group. The mean and variance of the random variable  $X$  were set to 2.0. The initial weights on the four stimulus dimensions were selected at random from the interval  $(-1, 1)$  according to a uniform

distribution for both the implicit and verbal systems. The learning rate for the perceptron learning rule was set to 0.3. The initial saliences for the four verbal rule types were all set to 0.25, that is,  $Z_i(0) = 0.25$  for  $i = 1, 2, 3, 4$ . The parameters  $\Delta_c$  and  $\Delta_E$  were both set to 0.02, and the initial system weights were set to  $W_v = .9$  and  $W_f = .1$ . In all other respects, the simulations of the Smith et al. (1993) experiments were the same as those described earlier.

## Appendix B

### An Alternative Parameterization of a Linear Decision Bound

In slope-intercept form, any linear bound can be written as  $bx_2 = ax_1 + c$ . Without loss of generality, we assume  $b = 1$ , except when the bound is a vertical line, in which case  $b = 0$ . In our new parameterization,  $\theta = \tan(a/b)$  when  $b \neq 0$ , and  $\theta = 90^\circ$  when  $b = 0$ . We restrict consideration to bounds that eventually intersect the upper right quadrant of the plane. There are two cases to consider.

$0^\circ < \theta < 90^\circ$ . In this case, the bound always intersects the origin if  $c = 0$ , it intersects the (positive)  $x_1$ -axis if  $c < 0$ , and it intersects the (positive)  $x_2$ -axis if  $c > 0$ . Note, however, that it never intersects both the  $x_1$ -axis and the  $x_2$ -axis. Throughout this range of  $\theta$ , note that  $b = 1$ , so the  $x_2$  intercept =  $c$  and the  $x_1$  intercept =  $-c/a$ . As a consequence,  $\tau = -c$  for  $c \geq 0$ , and  $\tau = -c/a$  for  $c < 0$ . The linear bounds in this range can therefore be expressed as  $x_2 = (\tan \theta)x_1 - \tau$ , if  $c \geq 0$ , and  $x_2/(\tan \theta) = x_1 - \tau$ , if  $c < 0$ .

$90^\circ \leq \theta \leq 180^\circ$ . In this case, the bound always intersects both the (positive)  $x_1$ - and  $x_2$ -axes. The  $x_1$  intercept is less than the  $x_2$  intercept if  $90^\circ < \theta < 135^\circ$ , and the  $x_2$  intercept is less than the  $x_1$  intercept if  $135^\circ < \theta < 180^\circ$ . Therefore,  $\tau = -c$  for  $135^\circ < \theta < 180^\circ$ , and  $\tau = -c/a$  for  $90^\circ \leq \theta \leq 135^\circ$ . The linear bounds in this range can therefore be expressed as  $x_2 = (\tan \theta)x_1 - \tau$ , if  $135^\circ \leq \theta \leq 180^\circ$ , and  $x_2/(\tan \theta) = x_1 - \tau$ , if  $90^\circ \leq \theta < 135^\circ$ .

This definition can be extended to bounds that do not intersect the upper right quadrant. However, in such cases, discontinuities arise that make this particular parameterization less attractive.

Received May 20, 1996

Revision received December 18, 1997

Accepted December 19, 1997 ■

### Low Publication Prices for APA Members and Affiliates

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.