To appear in: J. Wixted & H. Pashler (Eds.), *Stevens' Handbook of Experimental Psychology, Third Edition: Volume 4: Methodology in Experimental Psychology*. New York: Wiley.

Single Versus Multiple Systems of Learning and Memory

F. Gregory Ashby[1] & Shawn W. Ell
University of California, Santa Barbara

Many areas in cognitive psychology are currently debating whether learning and memory are mediated by one or more functionally distinct processing systems. Included in this list are the fields of memory, category learning, function learning, discrimination, and reasoning. Within each field, many of the multiple systems accounts have hypothesized at least two similar systems: an explicit system that is rule-based, and an implicit system that operates largely without conscious awareness. This chapter explores the debate between single and multiple systems. The focus is on the methodologies that have been proposed for testing between these two positions. In particular, we ask the following questions: 1) What constitutes a separate system? 2) What is the appropriate way to resolve this debate empirically?, and 3) What are the best empirical methodologies for testing between single and multiple systems? Finally, as a model of this debate, we focus on the question of whether human category learning is mediated by single or multiple systems.

One of the most hotly debated current issues in psychology and neuroscience is whether human learning and memory is mediated by a single processing system or by multiple qualitatively distinct systems. Although it is now generally accepted that there are multiple memory systems (Klein, Cosmides, Tooby, & Chance, in press; Squire, 1992; Schacter, 1987; Mishkin, Malamut, & Bachevalier, 1984; Zola-Morgan, Squire, & Mishkin, 1982; Cohen & Squire, 1980; O'Keefe & Nadel, 1978; Gaffan, 1974; Hirsh, 1974; Corkin, 1965), this issue is far from resolved in the case of learning and other cognitive processes. Even so, arguments for multiple systems have been made in such diverse fields as reasoning (Sloman, 1996), motor learning (Willingham, Nissen, & Bullimer, 1989), discrimination learning (Kendler & Kendler, 1962), function learning (Hayes & Broadbent, 1988), and category learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Brooks, 1978; Erickson & Kruschke, 1998). Interestingly, many of these papers have hypothesized at least two similar systems: 1) an explicit, rule-based system that is tied to language function and conscious awareness, and 2) an implicit system that may not have access to conscious awareness. In many cases, there has been resistance to these proposals, and a number of researchers have responded with papers arguing that single system models can account for many of the phenomena that have been used to support the notion of multiple systems (Nosofsky & Johansen, in press; Nosofsky & Zaki, 1998; Poldrack, Selco, Field, & Cohen, 1999).

This chapter explores the debate between single and multiple systems. The focus is on the methodologies that have been proposed for testing between these two positions. Thus, rather than attempting to resolve the debate by arguing for one position or another, our goals are to answer the following questions: 1) What constitutes a separate system? 2) What is the appropriate way to resolve this debate empirically?, and 3) What are the best empirical methodologies for testing between single and multiple systems? Many of the different areas currently engaged in the single versus multiple systems debate use similar methodologies to test between these two opposing arguments, and as mentioned above, they have all postulated similar explicit and implicit systems. For this reason, a detailed study of the debate in one area will most likely benefit the other areas as well. Thus, in the last major section, as a model of this debate, we focus on the question of whether human category learning is mediated by single or multiple systems.

---

[1] Preparation of this article was supported by grant BCS99-75037 f rom the National Science Foundation. Correspondence concerning this article should be addressed to F. Gregory Ashby, Department of Psychology, University of California, Santa Barbara, 93106, USA. Email: ashby@psych.ucsb.edu.

I. What Is A System?

Before one can examine methods for testing between single and multiple systems, one must first decide what is meant by a separate system. This question turns out to be as difficult as any that we will examine in this chapter. This is because all tasks in which we are interested are performed somewhere in the brain, and at one level, the brain is part of a single system (e.g., the central nervous system). At the other extreme, a strong argument can be made that each single cell, or even each single ion channel, forms its own system. So there is a continuum of levels, from macroscopic to microscopic, at which a system could be defined. It seems clear however, that the level chosen should match the task in question. Thus, a more macroscopic system is required to learn a new category of automobiles than to detect a sine-wave grating of a certain orientation. In the latter case, one could reasonably ask whether a column of cells in visual cortex defines the system, whereas in the former case this is clearly too reductionistic.

Given that an appropriate level and task are selected, what criteria should we use to decide whether some model postulates one or more systems? Suppose we have a model with two modules S_1 and S_2. The question is: do S_1 and S_2 define separate systems, or should they be viewed as two components of a single system? We believe there is no single criterion that can be used to answer this question. Instead, we propose a hierarchy of criteria – from the mathematical to the psychological to the neurobiological. Two modules that meet all these criteria are clearly separate systems. Modules that meet none of the criteria clearly do not constitute separate systems, and modules that meet some, but not all the criteria are in some ambiguous gray region along the single system - multiple system continuum.

Suppose the model for S_1 is characterized by a set of parameters denoted by the vector $\boldsymbol{\theta}_1$ and the model for S_2 is characterized by the parameters $\boldsymbol{\theta}_2$. For any specific set of numerical values of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, the models of S_1 and S_2, respectively, each predict a certain probability distribution of the relevant dependent variable, whatever that might be. Denote these probability distributions by $f_1(x|\boldsymbol{\theta}_1)$ and $f_2(x|\boldsymbol{\theta}_2)$, respectively. As the numerical values of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ change, these predicted probability distributions will also change. Therefore, let $\{f_1(x|\boldsymbol{\theta}_1)\}$ and $\{f_2(x|\boldsymbol{\theta}_2)\}$ denote the set of all possible probability density functions that can be generated from the S_1 and S_2 models, respectively (i.e., any numerical change in $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$ creates a new member of these sets). Then a mathematical criterion for S_1 and S_2 to be separate systems is that $\{f_1(x|\boldsymbol{\theta}_1)\}$ and $\{f_2(x|\boldsymbol{\theta}_2)\}$ are not identical, and neither is a subset of the other. In other words, the models of S_1 and S_2 are not mathematically equivalent and one is not a special case of the other – i.e., they each make at least some unique predictions. If the models were completely mathematically equivalent, so no experiment could ever be run that could produce data that might differentiate the two, then it is difficult to see how they could qualify as separate systems.

Note that an implicit assumption of this definition is that S_1 and S_2 each make predictions about observable behavior (since they each predict some probability distribution on the relevant dependent variable). This itself, is a stringent requirement that eliminates many possible models. For example, signal detection theory postulates separate sensory and decision processes, each described by its own parameter (d' and X_C, respectively). But either process, by itself, is incapable of making predictions about behavior. Instead, the two subsystems are assumed to always work together to produce a behavioral response. As such, standard signal detection theory is a single system theory, even though it postulates functionally separate sensory and decisional subsystems.

At the psychological level, to qualify as separate systems S_1 and S_2 should postulate that different psychological processes are required to complete the task in question successfully. For example, a multiple systems account of category learning might postulate separate prototype abstraction and rule-based systems, but a model that proposed two different prototype abstraction processes might be better described as a single system model. This criterion would also apply the single system label to a theory that postulated two separate signal detection systems, one say, with a more efficient sensory process and the other with a more efficient

decision process. This is because both systems would postulate similar (but not identical) sensory and decision processes that are active on all trials.

At the neurobiological level, separate systems should be mediated by separate neural structures or pathways. In most cases, there will be widespread agreement within the field of neurobiology about whether a pair of structures are part of the same or different systems, so this criterion should usually be straightforward to test. Within cognitive psychology, this should be the gold standard for establishing the existence of separate systems. It is highly likely that if the neurobiological condition is met, then the psychological and mathematical conditions will also be met. However, it is very easy to find examples in which the reverse implication fails. For example, one could easily construct two different exemplar-based category learning models that are mathematically identifiable (i.e., so the mathematical condition is met), but that postulate the same process of accessing category exemplars and computing their similarity to the presented stimulus, and therefore are also mediated by the same neural structures and pathways.

Just as the theoretical criteria for the existence of separate systems can be formulated at several different levels of analysis, so too is it vitally important to appeal to converging operations when empirically testing between single and multiple systems of learning and memory. It is extremely unlikely that any single experiment will yield data that definitively decides the question of whether there are single or multiple systems in any specific area of learning or memory. For any single set of data that purportedly supports the existence of multiple systems, for example, it is highly likely that a clever researcher will be able to construct a single system model that can account for those data. Thus, it is vital that when evaluating any new model, whether it postulate single or multiple systems, data is considered from many different experimental paradigms. Ideally, such data would come from several different levels of analysis – including behavioral neuroscience, traditional cognitive psychology, as well as cognitive neuroscience and neuropsychology.

II. Specific Methodological Tests of Single Versus Multiple Systems

A formal investigation of the efficacy of various methods for testing between single and multiple systems of learning and/or memory requires more structure than our previous discussions. Consider an experiment with several different conditions in which the dependent variable on condition i is denoted by the random variable $\mathbf{X}_i$. Denote the probability density function (pdf) of $\mathbf{X}_i$ in condition *i* by $g_i(x)$. As concrete examples, $\mathbf{X}_1$ and $\mathbf{X}_2$ might be the response times (RTs) from an experiment with two different conditions that load on different putative memory systems, or they might be the number of trials required to reach some criterion accuracy level in this same experiment. In the former case, $g_i(x)$ might be the RT distribution produced by a single subject in condition i, but in the latter case $g_i(x)$ would be the trials-to-criterion distribution across a group of subjects who all participated in condition i (i.e., because each subject produces many RTs, but only one value for trials-to-criterion in each condition).

Next consider an organism with two separate memory systems, either of which might be sufficient to complete the experimental task by itself. Let $\mathbf{X}_{Ai}$ and $\mathbf{X}_{Bi}$ denote the value of the dependent variable on trials when condition i is completed by systems A and B, respectively, and let $f_A(x|\,i)$ and $f_B(x|\,i)$ denote their respective pdfs. The pdf $g_i(x)$ is the distribution of observable data values and so it can always be estimated directly. As we will see, however, whether the pdfs $f_A(x|\,i)$ and $f_B(x|\,i)$ can be estimated directly depends on the model we assume.

In this section, we will consider three different types of multiple systems models. In the *strong model*, the observer uses only system A in experimental condition 1, and only system B in experimental condition 2. Thus,

$$g_1(x) = f_A(x|\,1) \text{ and } g_2(x) = f_B(x|\,2). \quad (1)$$

The assumption that different systems are used in the two tasks has been called *selective influence* in the single versus multiple systems literature (Dunn & Kirsner, 1988), after a similar assumption in the response time literature that was identified by Sternberg (1969). Almost all of the formal analysis of methodologies that purport to test between single and multiple systems (e.g., double dissociations) are based on this strong model.

In practice however, it seems possible that both systems would contribute to performance in both conditions, with the relative contributions of systems A and B varying from condition 1 to condition 2. For example, explicit memory systems may contribute to performance on putative implicit memory tasks (and vice versa). There are two obvious models of how this division of labor might proceed. In the *mixture model*, the observable response is determined by a single system on each trial, but memory system A determines the response on some trials and memory system B determines the response on the remaining trials. Let p_i denote the probability that memory system A determines the response in condition *i*. Then the mixture model predicts that the observable pdf is a probability mixture of the two component pdfs – that is,

$$g_i(x) = p_i\, f_A(x|\,i) + (1 - p_i)\, f_B(x|\,i). \tag{2}$$

The third possibility that we will consider is that both systems contribute to the observable response on every trial. In fact, in the *averaging model* the observable dependent variable is a weighted average of the outputs of the two component systems. In particular,

$$\mathbf{X}_i = r_i\, \mathbf{X}_{Ai} + (1 - r_i)\, \mathbf{X}_{Bi}, \tag{3}$$

where $0 \le r_i \le 1$ is the weight given memory system A in condition i. The observable pdf is found from a generalization of the so-called convolution integral:

$$g_i(x) = \frac{1}{r_i(1-r_i)} \int_{-\infty}^{\infty} f\left(\tfrac{x-w}{r_i}, \tfrac{w}{1-r_i}\,\middle|\, i\right) dw, \tag{4}$$

where $f(x_A, x_B|\,i)$ is the joint pdf of $\mathbf{X}_{Ai}$ and $\mathbf{X}_{Bi}$.

Equations 2 and 3 are in a similar form, but mathematically their behavior is very different. For example, suppose systems A and B can both complete task i, but that system A is much better adapted to performing this task than system B. Then $f_A(x|\,i)$ and $f_B(x|\,i)$ will have very different means. In the mixture model, this will be obvious because on trials when the observer uses system A, RT will be short, but RT will be long on trials when the observer uses system B. In fact, if the A and B means are far enough apart, then the observable pdf, $g_i(x)$, will be bimodal. However, in the averaging model the observer does the same thing on every trial, and as a result, RT will always be of intermediate value and $g_i(x)$ will therefore be unimodal. For these reasons, mixture models will generally be easier to discriminate from single system models than will averaging models, which like single system models assume observers do the same thing on all trials.

The Fixed-Point Property of Binary Mixtures

An obvious signature of a mixture model would be a bimodal pdf (in the case of binary mixtures). Unfortunately, mixture models will produce unimodal pdfs unless the component distributions are far apart. Thus, it is important to find some other less obvious signature left by mixture models. A solution to this problem was discovered more than 30 years ago.

The issue of whether choice RT was mediated by a mixture model or a single system model achieved

intense scrutiny during the 1960's and 1970's (e.g., Falmagne, 1968; Falmagne & Theios, 1969; Lupker & Theios, 1977; Townsend & Ashby, 1983; Yellott, 1969; 1971). The interest was generated by Yellott's (1969) proposal that some proportion of responses in speeded choice tasks were simple guesses, and thus the observable RTs were a mixture of fast guesses and slower times from trials when complete processing occurred. In response, Falmagne (1968) proposed a clever test of mixture models that he called the *fixed-point property*. Consider a special case of Equation 2 in which the mixture probability p_i varies across the experimental conditions (i.e,. varies with i), but the component system pdfs do not – that is,

$$f_A(x|i) = f_A(x) \text{ and } f_B(x|i) = f_B(x), \text{ for all values of i.}$$

In each experimental condition, all we can estimate, of course, is the observable pdf, $g_i(x)$. The fixed-point property of binary-mixtures states that all such mixtures must intersect at the same time point, if they intersect at all (Falmagne, 1968).

Figure 1 shows examples of $g_i(x)$ when the component pdfs, $f_A(x)$ and $f_B(x)$, are each normal distributions with equal variance, and the mixture probability p_i varies across conditions from 0.2 to 0.8. Note that the resulting pdfs (which are not themselves normal) all intersect at the point $x = 0.5$. Although it is mathematically possible that a single-system model could coincidentally mimic this result, such a possibility seems highly unlikely, so a set of empirical pdfs that satisfy the fixed-point property should be taken as strong evidence of multiple systems. On the other hand, the converse result is much weaker. There are many reasons why the mixture model might fail to display the fixed-point property, so data in which the fixed-point property fails do not constitute strong evidence against the mixture model. For example, it might be the case that the component pdfs change across conditions, in addition to the mixture probability p_i.

The fixed-point property has not been used to test for single versus multiple systems of learning or memory, but there is no reason, in principle, why it could not. For example, consider the category structures shown in Figure 2. Suppose a researcher believes that learning of these structures will depend heavily on memorization when there are only a few exemplars per category, but as the number of exemplars is increased, observers begin learning and applying a more abstract rule. This dual-system hypothesis could be tested via the fixed-point property. For example, consider the stimulus labeled T in Figure 2. Note that this stimulus appears in every condition. Suppose the conditions are ordered so that the smallest categories are learned first and more exemplars are successively added (so the order is Figure 2a - 2b - 2c). In each condition, enough data is collected to estimate the RT distribution for stimulus T. If the theory is right, then in Figure 2a, the RT distribution for stimulus T will be determined primarily by a memorization strategy and in Figure 2c by applying an abstract rule. If during the transition, the observer intermixes trials in which the response to stimulus T is generated by these two systems, then the stimulus T RT distributions across conditions should satisfy the fixed-point property.

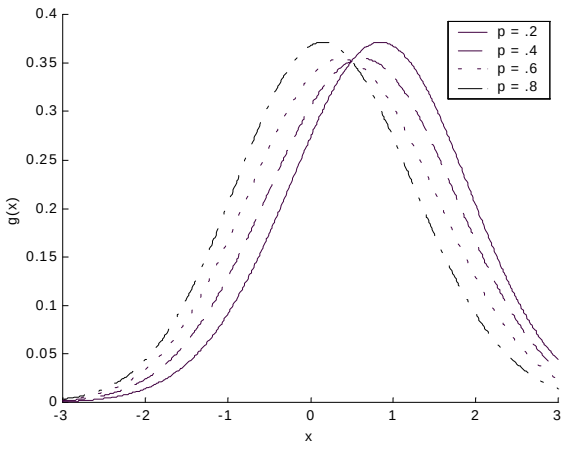

Figure 1. Examples of probability density functions that satisfy the fixed-point property (see text for details).

In this case, dual systems are supported if the observable RT pdfs all intersect at the same point. Unfortunately, however, if they do not satisfy the fixed-point property, it is difficult to draw any strong conclusions. Recall that a necessary condition for the fixed-point property to hold is that $f_A(x|\mathrm{i}) = f_A(x)$ and $f_B(x|\mathrm{i}) = f_B(x)$ – in other words, the component system pdfs for the time to categorize stimulus T are the same in all three conditions shown in Figure 2. This is a strong assumption that could fail for a variety of reasons. For example, the rule-based system might use a slightly different rule in the three conditions. There is much evidence that categorization RT is strongly affected by the distance from the stimulus to the category boundary (Ashby, Boynton, & Lee, 1994; Maddox, Ashby, & Gottlob, 1998), so if the boundary (i.e., rule) changes, then the distance between T and the boundary will change, and so will the time it takes the rule-based system to categorize stimulus T. Similarly, it may be that the memorization system slows down when the number of exemplars that must be memorized increases. This would cause the pdf from the memorization system to change (i.e., move to the right) as more stimuli are added from one condition to the next.

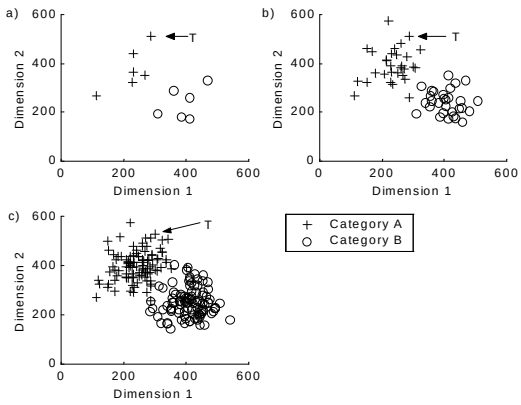

Figure 2. Example of category structures to which the fixed point property might be applied. a) A memorization strategy may be utilized to learn this structure with few exemplars. However, as the number of exemplars increases (in b and c), it seems more likely that an abstract rule may be applied.

Double Dissociations

The most widely used current method for establishing that there are multiple systems of learning or memory is to find a double dissociation between two tasks that load differently on the two systems. Many such examples exist. To name one, several studies have found that rats with lesions of the tail of the caudate nucleus are impaired in visual discrimination learning but not in spatial learning, whereas rats with lesions to the fornix (the output structure of the hippocampus) show the opposite pattern – namely, they are impaired in spatial learning but are normal in visual discrimination learning (Packard, Hirsch, & White, 1989; Packard & McGaugh, 1992; McDonald & White, 1994). An example of the pattern of results one would expect in such a situation is given in Figure 3. Note that the dependent variable is trials-to-criterion.

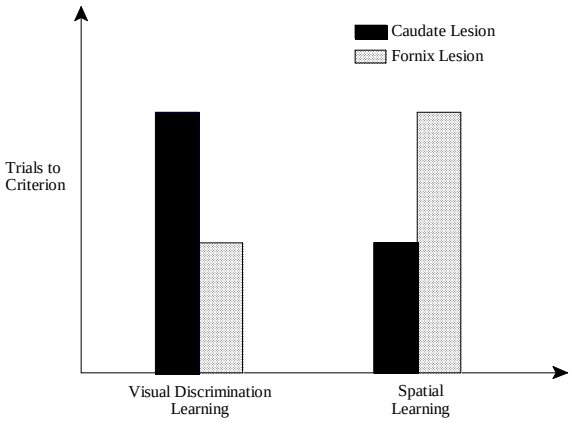

Figure 3. Hypothetical results showing a double dissociation between visual discrimination learning and spatial learning for two different types of lesions (tail of the caudate nucleus or fornix).

There are several properties of the Figure 3 results that are necessary for them to qualify as a double dissociation (a term first coined by Teuber, 1955). First, the interaction must be of the cross-over type. A non-crossover interaction does not qualify as a double dissociation, no matter what its level of statistical significance. This is because it is relatively easy for single system models to account for non-crossover

interactions (this is demonstrated below). Second, the cross-over interaction must come from measuring the same dependent variable in two different tasks. Thus, a cross-over interaction, by itself is not sufficient to qualify as a double dissociation. Again, this is because it is straightforward for single system models to account for cross-over interactions in 2 × 2 designs when only one task is used and different dependent variables are measured (more detail on this is provided later in this section).

A third condition, which is not strictly necessary but greatly strengthens the argument that a double dissociation supports multiple systems, is that the two groups in the experiment each are representative of some homogeneous population. In the Figure 3 example, the same results would be assumed to hold for any group of rats that received these same lesions. McCloskey (1993) in particular, has forcefully argued this point. Of the phrase "homogeneous population", both words are important. For example, McCloskey (1993) showed that spurious conclusions are possible (or perhaps likely) if each group contains a mixture of observers with different types of lesions. This homogeneity requirement makes the interpretation of a double dissociation especially problematic if each group comprises humans who have suffered some particular type of lesion. Since human lesions are generally the result of accident or stroke, no two are alike. For example, they are often unilateral and do not respect the neuroanatomical boundaries established by Broadman and others. From this perspective, neurodegenerative disease groups (e.g., Parkinson's disease) are probably better candidates for double dissociation studies, but even in Parkinson's disease there is widespread individual difference in the neuroanatomical locus and extent of damage (e.g., van Domburg & ten Donkelaar, 1991). For this reason, it is important that, whenever possible, any double dissociations reported in humans are replicated in nonhuman animals under more controlled conditions.

The term "population" in the phrase "homogeneous population" is equally important. For example, suppose one of our groups is normal, healthy, adult humans, and that a single neuropsychological patient is discovered who, when defined as the second group, produces data that satisfies a double dissociation. Several researchers have emphasized the dangers in attempting to make inferences from such data (e.g., Shallice, 1988; Van Orden, Pennington, & Stone, in press). For example, since we have no data from this particular patient before his or her neurological trauma, we do not know whether the patient would have produced these idiosyncratic data before the trauma, and thus, that the peculiar data are the result of the neurological damage. When one samples from any variable population, eventually an extreme outlier is encountered that might not be representative of any existing population.

Another popular argument against double dissociation logic is that it leads to the conclusion that there are too many functionally separate systems (e.g., Van Orden et al., in press). For example, consider two tasks – both are YES/NO detection tasks where the signal is a sine wave grating and the noise is a uniform field. In the first task, however, the frequency of the signal grating is f_1 degrees and in the second task the signal has frequency f_2 degrees. Our two groups are animals with lesions to specific spatial frequency columns in primary visual cortex. Group 1 has a lesion to columns sensitive to spatial frequencies centered at f_1 degrees, and Group 2 has a lesion to columns sensitive to frequencies centered at f_2 degrees. This experiment should produce a double dissociation, so the standard conclusion would be that there are separate systems for the detection of gratings of f_1 and f_2 degrees. Furthermore, if we repeat this experiment with other frequencies, we will have to conclude that a number of other such systems also exist. In a sense, our logic is correct since visual psychophysiologists often treat different cortical columns (or hypercolumns) as separate (mini) systems. On the other hand, from the perspective of cognitive psychology this conclusion seems too reductionistic. Cognitive psychologists might be satisfied to learn, for example, only that there are separate systems for spatial frequency and orientation perception. At this point, any more detail would just overwhelm theory development.

From a practical perspective, the problems arise in our hypothetical detection experiment because the

two tasks are so similar[2]. According to standard signal detection theory, they require the same sensory and decision processes. Therefore, a practical solution to the problem is to use current theory regarding the function of the postulated systems to aid in selecting the tasks to be used in the double dissociation experiment. In particular, two tasks should be used only if there is current theoretical debate as to whether they are mediated by one or more separate systems.

In the remainder of this section, we formally examine the validity of claims that a double dissociation is strong evidence for multiple systems. We assume throughout this discussion that the double dissociation was produced in an experiment that satisfies all of the guidelines described above (and avoids the pitfalls).

To begin, consider the strong multiple systems model described in Equation 1. Suppose system A is based in the hippocampus (e.g., the fornix) and specializes in spatial memory tasks and system B is based in the caudate nucleus and specializes in visual discrimination tasks. Denote the pdf of system A in the spatial memory task when the fornix is lesioned by $f_A'(x|\mathrm{S})$, and the pdf of system B in the visual discrimination task when the caudate is lesioned by $f_B'(x|\mathrm{V})$. Such lesions will impair the two systems. We can document this by assuming that lesions affect the entire pdfs. Specifically, we assume that the performance of system A in the normal and lesioned groups is related via

$$\mathrm{P}(\mathbf{X}_A \le x) \ge \mathrm{P}(\mathbf{X}_A' \le x), \text{ for all values of } x. \tag{5}$$

These two functions are called the cumulative probability distribution functions, denoted by $\mathrm{F}_A(x)$ and $\mathrm{F}_A'(x)$, respectively, so Equation 5 is equivalent to

$$\mathrm{F}_A(x) \ge \mathrm{F}_A'(x), \text{ for all } x. \tag{6}$$

Similarly, we assume

$$\mathrm{F}_B(x) \ge \mathrm{F}_B'(x), \text{ for all } x. \tag{7}$$

Note that the orderings specified by Equations 6 and 7 guarantee that the means will also be ordered (although in the reverse direction – i.e., lesions will increase mean trials-to-criterion). Figure 4a presents hypothetical cumulative distribution functions (left) and the relative ordering of the means (right) predicted by Equations 6 and 7.

Let $\mathrm{G}_{IJ}(x)$ denote the cumulative distribution function of trials-to-criterion for group J (J = F or C for fornix or caudate lesions) in task I (I = S or V for spatial memory or visual discrimination). We assume that this function provides a complete description of the dependent variable of interest (e.g., trials-to-criterion).

In this strong multiple systems model, the observable cumulative distribution functions in the four conditions are:

	Spatial Memory Task	Visual Discrimination Task
Fornix Lesion	$G_{SF}(x) = F_A'(x\|\mathrm{S})$	$G_{VF}(x) = F_B(x\|\mathrm{V})$
Caudate Lesion	$G_{SC}(x) = F_A(x\|\mathrm{S})$	$G_{VC}(x) = F_B'(x\|\mathrm{V})$

[2] In fact, one might easily argue that they are so similar that they should be considered the same task, a conclusion that would violate our earlier condition that two different tasks are needed to test for a double dissociation.

Equations 6 and 7 guarantee that this model produces the cross-over double dissociation. Figure 4b presents a graphical example of these orderings.

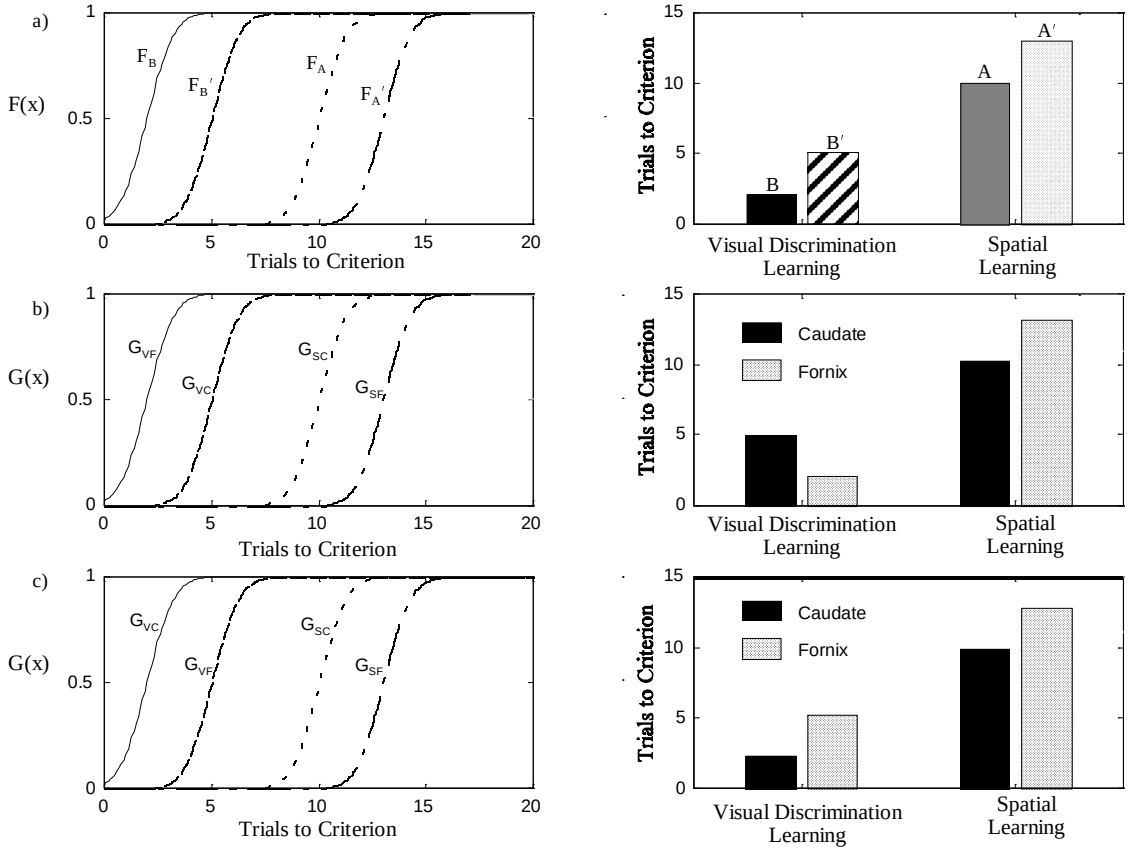

Figure 4. Cumulative distribution functions (left) and means (right) in four conditions of a hypothetical experiment (see text for details). a) Orderings induced by Equations 6 and 7. b) Predictions of the strong multiple systems model. c) Predictions for a single system model that satisfies Equation 8 (i.e., fornix lesions are more detrimental than caudate lesions).

Next, consider what a single system model predicts in this experiment. Even if the same system is used on every trial of all conditions, that system might not be equally suited to the two types of task, and the two types of lesions might not inflict the same amount of damage to the system. With these caveats in mind, single system models predict:

	Spatial Memory Task	Visual Discrimination Task
Fornix Lesion	$G_{SF}(x) = F_F'(x\|\mathrm{S})$	$G_{VF}(x) = F_F'(x\|\mathrm{V})$
Caudate Lesion	$G_{SC}(x) = F_C'(x\|\mathrm{S})$	$G_{VC}(x) = F_C'(x\|\mathrm{V})$

where the subscripts F and C refer to the fornix and caudate, respectively. Now, if the fornix lesion causes

more damage to the system than the caudate lesion, then we assume that the ability of the system to perform in any task is poorer with fornix lesions than with caudate lesions. Thus,

$$F_C'(x|S) \geq F_F'(x|S) \text{ and } F_C'(x|V) \geq F_F'(x|V), \text{ for all } x. \tag{8}$$

Similarly, if the caudate lesion causes more damage, then

$$F_F'(x|S) \geq F_C'(x|S) \text{ and } F_F'(x|V) \geq F_C'(x|V), \text{ for all } x. \tag{9}$$

In either case, there is no cross-over interaction and therefore, no double dissociation (see Figure 4c for an example of the Equation 8 predictions).

There are several points worth noting here. First, even if Equation 8 or 9 holds, an interaction is possible in the single system model – only a cross-over interaction is precluded. Additive effects (i.e., no interaction) would occur only if the deleterious effect of the more damaging lesion was exactly the same in both tasks. This might occur, but there is no reason it should be expected.

Second, this analysis makes it clear that a single system model can predict a double dissociation if Equations 8 and 9 both fail -- that is, if the deficit is more severe with the first lesion in one task and with the second lesion in the other task. For example, single system models predict a double dissociation if

$$F_C'(x|S) \geq F_F'(x|S) \text{ and } F_F'(x|V) \geq F_C'(x|V), \text{ for all } x. \tag{10}$$

This point was noted by Dunn and Kirsner (1988), who called Equation 10 a negative relation between the tasks. With lesion data, it is difficult to imagine how this might occur in a true single-system model. One possibility though, is that the single system is composed of several subsystems – one of which is knocked out by fornix lesions and another by caudate lesions. A double dissociation could result if the subsystem damaged by the fornix lesion was more important in the spatial memory task and the subsystem damaged by the caudate lesion was more important in the visual discrimination task. There are several problems with this scenario, however. First, if the subsystems are arranged in series, with the output of one serving as the input for the other, then it is not clear that a double dissociation would result. Damage to the upstream subsystem would cause poor performance on both tasks because the input to the downstream, undamaged subsystem would be corrupted. On the other hand, damage to the downstream subsystem would affect performance only on one task, because the input and processing in the upstream subsystem would be unaffected by such a lesion. Thus, the only way the double dissociation is guaranteed is if the two subsystems operate in parallel. Such a parallel system, however, shares many properties with multiple systems, so it is unclear that its existence should be taken as support for a single system.

If different dependent variables are used for the two groups, then it becomes easy for single system models to predict cross-over double dissociations. For example, consider the hypothetical categorization RT data shown in Figure 5a. In this experiment, subjects must decide whether each presented stimulus is or is not a member of category A. Figure 5a shows mean RT for “A” and “not A” responses as a function of the similarity between the stimulus and the category A prototype. These data are easily predicted by a single system model that assumes subjects compute the similarity of the stimulus to the category A prototype, and then compare this similarity to a criterion. Similarities above the criterion elicit an “A” response and similarities below the criterion elicit a “not A” response. Such a model predicts the Figure 5a data if the time to determine whether the similarity is above or below criterion decreases with the magnitude of the difference between the similarity and the criterion. Clearly, in such a case, it would be a mistake to infer from Figure 5a that there are separate systems on “A” and “not A” trials.

From the perspective of double dissociation logic, there are several problems with the Figure 5a example. First, there are neither two groups nor two tasks. Instead, the Figure 5a data are from one group of subjects in one task. Second, data from two different types of response are plotted in Figure 5 – RT for "A" responses and RT for "not A" responses. Note that this contrasts with the double dissociation shown in Figure 3, in which the response is the same in all conditions. In Figure 5a, data from one experimental condition are divided into two categories (according to the response given). Then a variable is constructed (similarity-to-prototype) that subdivides these two categories in such a way that a cross-over interaction occurs. It is important to note, however, that other variables could be defined that subdivide the categories differently, and for which the interaction might disappear. For example, the same data are replotted in Figure 5b against the variable "psychological distance to category bound."

If performance in some task is mediated by a single system, then it is natural that there may exist negative relations between different kinds of responses, or different dependent variables (e.g., speed versus accuracy). Clearly, it would be a mistake to apply double dissociation logic to a cross-over interaction in such a case.

These analyses provide a rigorous justification for the practice of inferring multiple systems when double dissociations are found, but only under a fairly limited set of circumstances (e.g., different tasks, same response, separate homogeneous populations). On the other hand, the only multiple systems model we have so far considered is the strong model that assumes selective influence -- that is, that the observer uses separate systems in the two tasks under study. Perhaps a more plausible multiple systems alternative is that the observer uses both systems in both conditions, but the two tasks load differently on the two systems and the observable response is determined either by only one of the systems on any given trial or by a weighted average of the two system outputs. In other words, it is of interest to consider the conditions under which the mixture and averaging models predict a double dissociation. To our knowledge, this question has not previously been investigated.

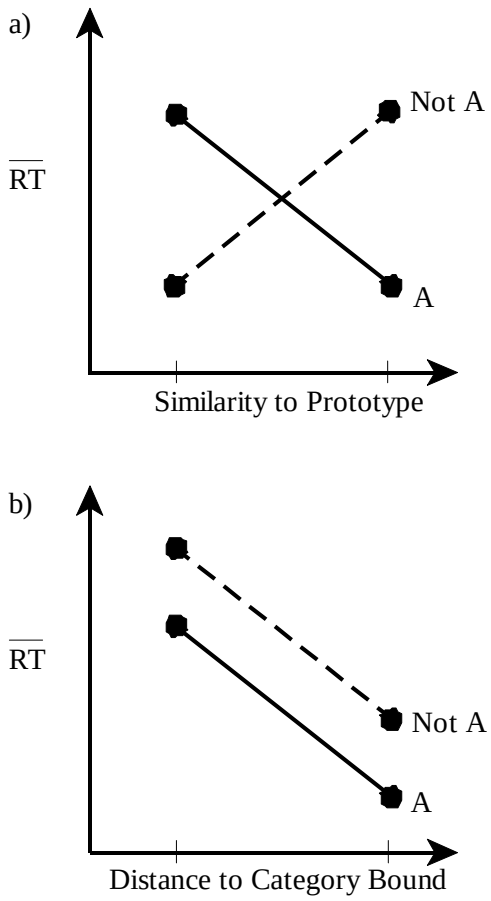

Figure 5. Hypothetical categorization RT data. a) Mean RT plotted as a function of similarity to prototype in an A-not A task. b) Data from the same experiment plotted as a function of distance to category bound (see text for details).

We begin with the mixture model. Let p_S and p_V denote the probability that the hippocampal-based system is used on any given trial of the spatial memory task and the visual discrimination task, respectively. We assume that observers are more likely to use the hippocampal system in the spatial memory task and the caudate system in the visual discrimination task. This means that $p_S > ½ > p_V$. As before, we assume that the effect of the lesions is as described in Equations 6 and 7. Under these assumptions, the cumulative distribution functions in each condition are given by:

	Spatial Memory Task	Visual Discrimination Task
Fornix Lesion	$G_{SF}(x) = p_S F_A'(x\|S) + (1 - p_S)F_B(x\|S)$	$G_{VF}(x) = p_V F_A'(x\|V) + (1 - p_V)F_B(x\|V)$
Caudate Lesion	$G_{SC}(x) = p_S F_A(x\|S) + (1 - p_S)F_B'(x\|S)$	$G_{VC}(x) = p_V F_A(x\|V) + (1 - p_V)F_B'(x\|V)$

It is not difficult to show[3] that this mixture model predicts a (cross-over) double dissociation if and only if for all values of x,

$$\frac{p_S}{1-p_S} > \frac{F_B(x|S) - F_B'(x|S)}{F_A(x|S) - F_A'(x|S)}, \tag{10}$$

and

$$\frac{1-p_V}{p_V} > \frac{F_A(x|V) - F_A'(x|V)}{F_B(x|V) - F_B'(x|V)}. \tag{11}$$

Since $p_S > ½ > p_V$, the left side is greater than 1 in both equations. By Equations 6 and 7, the numerator and denominator of the right hand side are positive in both equations. Thus, the mixture model predicts a double dissociation anytime the effects of the lesions are the same on the two systems. If they are not – for example if the caudate lesion more effectively impairs the caudate-based system than the fornix lesion impairs the hippocampal-based system – then whether or not the mixture model predicts a double dissociation depends on the mixture probabilities p_S and p_V. If the experimenter is effective at finding two tasks that each load heavily on different systems, then p_S will be near 1 and p_V will be near 0, and the left side of Equations 10 and 11 will both be large. In this case, a double dissociation will occur even if there are large differences in the efficacy of the various lesions. Thus, with the mixture model of multiple systems, a double dissociation is not guaranteed, but it should generally be possible to find tasks and conditions (e.g., lesions) that produce one.

The predictions of the averaging model are qualitatively similar to those of the mixture model if we shift our focus from the cumulative distribution functions, $F_A(x)$ and $F_B(x)$, to the means $E(\mathbf{X}_{Ai})$ and $E(\mathbf{X}_{Bi})$ (e.g., this allows us to avoid dealing with the convolution integral of Equation 4). Let r_S and r_V denote the weights given the hippocampal-based system on any given trial of the spatial memory task and the visual discrimination task, respectively. We assume that observers weight the hippocampal system more heavily in the spatial memory task and the caudate system more heavily in the visual discrimination task. Thus $r_S > ½ > r_V$. As before, we assume the lesions impair performance [i.e., since the dependent variable is trials-to-criterion, this

[3] If the caudate group performs better than the fornix group in the spatial memory task, then

$$p_S F_A(x|S) + (1 - p_S)F_B'(x|S) > p_S F_A'(x|S) + (1 - p_S)F_B(x|S), \text{ for all } x,$$

which implies that

$$p_S [F_A(x|S) - F_A'(x|S)] > (1 - p_S) [F_B(x|S) - F_B'(x|S)], \text{ for all } x.$$

Equation 10 follows readily from this result. Equation 11 follows in a similar fashion from the result that a double dissociation requires the fornix group to perform better than the caudate group in the visual discrimination task.

means that $E_A'(\mathbf{X}) > E_A(\mathbf{X})$ and $E_B'(\mathbf{X}) > E_B(\mathbf{X})$]. Under these assumptions, the observable means in each condition are given by:

	Spatial Memory Task	Visual Discrimination Task
Fornix Lesion	$E_{SF}(\mathbf{X}) = r_S E_A'(\mathbf{X}\mid S) + (1 - r_S)E_B(\mathbf{X}\mid S)$	$E_{VF}(\mathbf{X}) = r_V E_A'(\mathbf{X}\mid V) + (1 - r_V)E_B(\mathbf{X}\mid V)$
Caudate Lesion	$E_{SC}(\mathbf{X}) = r_S E_A(\mathbf{X}\mid S) + (1 - r_S)E_B'(\mathbf{X}\mid S)$	$E_{VC}(\mathbf{X}) = r_V E_A(\mathbf{X}\mid V) + (1 - r_V)E_B'(\mathbf{X}\mid V)$

Note the similarity to the structure of the cumulative distribution functions in the mixture model. As a result, the averaging model predicts a double dissociation if

$$\frac{r_S}{1-r_S} > \frac{E_B'(\mathbf{X}|S) - E_B(\mathbf{X}|S)}{E_A'(\mathbf{X}|S) - E_A(\mathbf{X}|S)}, \quad (12)$$

and

$$\frac{1-r_V}{r_V} > \frac{E_A'(\mathbf{X}|V) - E_A(\mathbf{X}|V)}{E_B'(\mathbf{X}|V) - E_B(\mathbf{X}|V)}. \quad (13)$$

The conclusions are therefore similar to the case of the mixture model. The averaging model predicts a double dissociation if the effects of the two lesions are approximately equal. If one lesion is more severe than the other, then a double dissociation can still be predicted if the two tasks load heavily on different systems.

We believe this analysis provides strong theoretical justification for the current practice of interpreting a double dissociation as evidence of multiple systems. However, we have also noted some important and severe limitations on this methodology. For example, it is essential that the observed interaction be of the cross-over type, and not just any interaction that achieves statistical significance. Also, the same dependent variable should be measured in two different tasks that sample from separate populations of homogeneous subjects. It is also important to note that there is an asymmetry in interpreting double dissociation results. Whereas the existence of a double dissociation (under the appropriate experimental conditions) is strong evidence for multiple systems, the failure to find a double dissociation must be interpreted more cautiously, because there are several reasonably plausible ways in which multiple systems models could produce this null result (e.g., see our discussion of the mixture model).

Single Dissociations

Although other definitions are possible, we operationally define a single dissociation as an interaction of the type described in the last section for which there is no crossover. As already mentioned, in the absence of extenuating circumstances, it is difficult or impossible to draw strong conclusions about whether such data were produced by single or multiple systems. As we have seen, in many cases it is straightforward for single system models to predict single dissociations. Even so, there are certain special circumstances in which single dissociation data have been used to argue for multiple systems.

Perhaps the most common argument that a single dissociation signals multiple systems has been in cases where two groups perform equally on one task, but one of these groups is impaired, relative to the other, on a second task. For example, amnesic patients perform poorly on explicit memory tests but they often are relatively normal on a variety of tests of implicit memory (e.g, Warrington & Weiskrantz, 1970). It is dangerous, however, to infer simply from this result that there are separate explicit and implicit memory systems. For example, there have been several formal demonstrations that certain single system models can account for such data (e.g., Nosofsky, 1988; Nosofsky & Zaki, 1998). In addition, recently it has been argued that even garden-variety single system models can account for single dissociations of this type if the explicit memory tests are more reliable than the implicit tests (Buchner & Wippich, 2000; Meier & Perrig, 2000).

These arguments generally assume no *a priori* knowledge about the nature of the tasks that are used. When such knowledge is considered, then stronger tests are sometimes possible. One such attempt employs what has been called the *logic of opposition* to test for unconscious learning (Jacoby, 1991; Higham, Vokey, & Pritchard, 2000). Consider a categorization task with two categories, denoted A and B. To begin, subjects are trained to identify members of these two categories. There are two different test conditions. In the control condition, subjects are shown a series of stimuli and are asked to respond "Yes" to each stimulus that belongs to Category A *or* B and to respond "No" to stimuli that are in neither category. In the opposition condition, subjects respond "Yes" only if the stimulus belongs to Category A. If it belongs to Category B or to neither category, then the correct response is "No". The key test is to compare the accuracy rates in the opposition condition for these two kinds of stimuli (i.e., those in Category B and those in neither category). The idea is that, if responding is based solely on conscious learning, then the accuracy rates to these two kinds of stimuli should be equal, but unconscious learning could cause Category B exemplars to become associated with the notion that these stimuli are valid category members, thereby causing more "Yes" responses to Category B exemplars than to stimuli in neither category. This logic, which is not without controversy (Redington, 2000), takes advantage of our knowledge that subjects were trained on Category B exemplars but not on the stimuli in neither category.

Another possible use of *a priori* knowledge is to focus on the relative difficulty of the two tasks. For example, consider the two tasks described by the performance operating curves shown in Figure 6. When full resources are available, Task 1 is easier to learn than Task 2 (i.e., criterion performance is achieved in fewer trials for Task 1 than for Task 2). As resources are withdrawn, performance naturally declines in both tasks, although at different rates. A small to moderate decline in the available resources is more deleterious to the more difficult Task 2 (e.g., when R_b resources are available for both tasks). However, as performance on Task 2 nears floor (i.e., worst possible performance), Task 1 performance begins to narrow the gap, until eventually performance on both tasks is equally bad. The point marked R_C in Figure 6 denotes the critical level of resources in which the rate of decline on Task 1 first exceeds the rate of decline on Task 2.

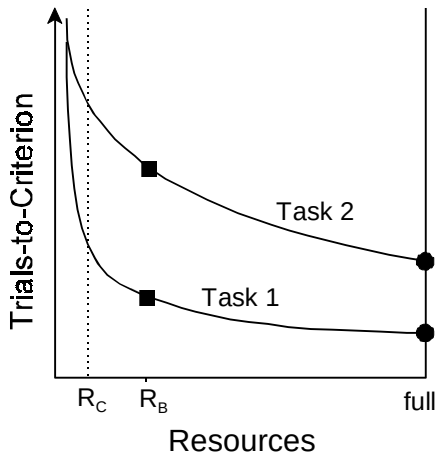

Figure 6. Performance operating characteristics of two tasks.

Now, suppose Tasks 1 and 2 are both learned by the same system, and consider an experiment with two conditions. In one, observers learn the two tasks with full resources available. This condition produces data points denoted by the closed circles in Figure 6. In the second

condition, observers learn the tasks with reduced resources. This could be accomplished either by requiring observers to perform a simultaneous dual task, or perhaps through instruction (e.g., by forcing a quick response). As long as the observer has available R_C or more resources in this latter condition, single-system models predict that the reduced resources condition will cause more problems in the more difficult Task 2. For example, with resources equal to R_B, the reduced resources condition produces data points denoted by the closed squares in Figure 6. The only potential problem with this prediction is if the observer had available less than R_C resources for the learning task in the reduced resources condition. This possibility should be easy to avoid however, by ensuring that performance on Task 2 is well below ceiling.

Next, consider predictions in this experiment if the observer uses different systems to learn Tasks 1 and 2, and for some reason the experimental intervention to reduce resources works more effectively on the system that learns Task 1. In this case, the greater interference will be with Task 1 – a result that is problematic for single system models.

This was the strategy of a recent experiment reported by Waldron and Ashby (in press). Participants in this study learned simple and complex category structures under typical single-task conditions and when performing a simultaneous numerical Stroop task. In the simple categorization tasks, each set of contrasting categories was separated by a unidimensional, explicit rule that was easy to describe verbally. An example is shown in Figure 7 for the rule "respond A if the background color is blue, and respond B if the background color is yellow". On the other hand, the complex tasks required integrating information from three stimulus dimensions and resulted in implicit rules that were difficult to verbalize. An example is shown in Figure 8. Ashby et al. (1998) hypothesized that learning in such tasks will be dominated by different systems – in particular, that the simple categories would be learned by an explicit, rule-based system that depends heavily on frontal cortical structures, whereas the complex categories would be learned primarily by an implicit, procedural learning system that depends heavily on subcortical structures. Stroop tasks are known to activate frontal cortex (Bench et al., 1993), and so it was hypothesized that the concurrent Stroop task would interfere with the explicit system more strongly than with the implicit system. In support of this prediction, the concurrent Stroop task dramatically impaired learning of the simple explicit rules, but did not significantly delay learning of the complex implicit rules. These results support the hypothesis that category learning is mediated by multiple learning systems.

Mapping Hypothesized Systems Onto Known Neural Structures

Testing between single and multiple systems of learning and memory will always be more difficult when the putative systems are hypothetical constructs with no known neural basis. For example, the Waldron and Ashby (in press) dual task study was more effective because it had earlier been hypothesized that the putative explicit system relied on frontal cortical structures much more strongly than the implicit system. Given this, and the neuroimaging evidence that Stroop tasks activate frontal cortex (Bench et al., 1993), it becomes much easier to argue that if there are multiple systems, then the concurrent Stroop task should interfere more strongly with the learning of the simpler, rule-based category structures.

In general, the memory literature has enthusiastically adopted this constraint. Most of the memory systems that have been proposed have become associated with a distinct neural basis. For example, cognitive neuroscience models of working memory focus on prefrontal cortex (e.g., Fuster, 1989; Goldman-Rakic, 1987, 1995), declarative memory models focus on the hippocampus and other medial temporal lobe structures (e.g., Gloor, 1997, Gluck & Myers, 1997; McClelland, McNaughton, & O'Reilly, 1995; Polster, Nadel, & Schacter, 1991; Squire & Alvarez, 1995), procedural memory models focus on the basal ganglia (e.g., Jahanshahi, Brown, & Marsden, 1992; Mishkin et al., 1984; Saint-Cyr, Taylor, & Lang, 1988; Willingham et al., 1989),

and models of the perceptual representation system focus on visual cortex (Curran & Schacter, 1996; Schacter, 1994; Tulving & Schacter, 1990).

III. Category Learning as a Model of the Single Versus Multiple Systems Debate

Category learning is a good example of an area in which the single versus multiple systems debate is currently being waged. The issues that have arisen in the category learning literature are similar to issues that are being discussed in other areas that are wrestling with this same debate. This is partly because similar methodologies are used in the different areas to test between single and multiple systems, and partly because the different sub-disciplines engaged in this debate – motor learning, discrimination learning, function learning, category learning, and reasoning – have all postulated similar explicit and implicit systems. So, there is a very real possibility that if there are multiple systems of category learning, these same (or highly similar) systems might also mediate other types of learning. For this reason, this section examines the debate as to whether there are single or multiple systems of category learning.

Within the field of categorization, the debate as to whether there is one or more than one learning system is just beginning. There have been no attempts to test the fixed point property, and empirical demonstrations of double dissociations are rare. Nevertheless, there have been some encouraging attempts to map category learning systems onto distinct neural structures and pathways, and as mentioned above, there has been at least one attempt to test for multiple systems by exploiting a known *a priori* ordering of task difficulty. Even so, in the case of category learning, the single versus multiple systems debate is far from resolved. Not only is there insufficient empirical evidence to decide this issue, but there is still strong theoretical disagreement. Although there have been a number of recent articles arguing for multiple category learning systems (Ashby et al., 1998; Erickson & Kruschke, 1998; Pickering, 1997; Waldron & Ashby, in press), there have also been recent papers arguing for a single system (e.g., Nosofsky & Johansen, in press; Nosofsky & Zaki, 1998).

Category Learning Theories

As one might expect, the early theories of category learning all assumed a single system. There were a number of such theories, but four of these have been especially important. *Rule-based theories* assume that people categorize by applying a series of explicit logical rules (e.g., Bruner, Goodnow, & Austin, 1956; Murphy & Medin, 1985; Smith & Medin, 1981). Various researchers have described this as a systematic process of hypothesis testing (e.g., Bruner et al., 1956) or theory construction and testing (e.g., Murphy & Medin, 1985). Rule-based theories are derived from the so-called classical theory of categorization, which dates back to Aristotle, although in psychology it was popularized by Hull (1920). The classical theory assumes that categorization is a process of testing whether or not each stimulus possesses the necessary and sufficient features for category membership (Bruner et al., 1956). Much of the work on rule-based theories has been conducted in psycholinguistics (Fodor, Bever, & Garrett, 1974; Miller & Johnson-Laird, 1976) and in psychological studies of concept formation (e.g., Bruner et al., 1956; Bourne, 1966).

Prototype theory assumes that the category representation is dominated by the prototype, or most typical member, and that categorization is a process of comparing the similarity of the stimulus to the prototype of each relevant category (Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968, 1970; Reed, 1972; Rosch, 1973, 1977; Smith & Minda, 2000). In its most extreme form, the prototype *is* the category representation, but in its weaker forms, the category representation includes information about other exemplars (Busemeyer, Dewey, & Medin, 1984; Homa, Dunbar, & Nohre, 1991; Shin & Nosofsky, 1992).

Exemplar theory assumes people compute the similarity of the stimulus to the memory representation of every exemplar of all relevant categories and select a response on the basis of these similarity computations (Brooks, 1978; Estes, 1986a; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). The assumption that the similarity computations include *every* exemplar of the relevant categories is often regarded as intuitively unreasonable. For example, Myung (1994) argued that "it is hard to imagine that a 70 year-old fisherman would remember every instance of fish that he has seen when attempting to categorize an object as a fish" (p. 348). Even if the exemplar representations are not consciously retrieved, a massive amount of activation is assumed by exemplar theory. Nevertheless, exemplar models have been used to account for asymptotic categorization performance from tasks in which the categories: (1) were linearly or non-linearly separable (Medin & Schwanenflugel, 1981; Nosofsky, 1986, 1987, 1989), (2) differed in baserate (Medin & Edelson, 1988), (3) contained correlated or uncorrelated features (Medin, Alton, Edelson, & Freko, 1982), (4) could be distinguished using a simple verbal rule (or a conjunction of simple rules; Nosofsky, Clark, & Shin, 1989), and (5) contained differing exemplar frequencies (Nosofsky, 1988).

Finally, *decision bound theory* (also called general recognition theory) assumes there is trial-by-trial variability in the perceptual information associated with each stimulus, so the perceptual effects of a stimulus are most appropriately represented by a multivariate probability distribution (usually a multivariate normal distribution). During categorization, the observer is assumed to learn to assign responses to different regions of the perceptual space. When presented with a stimulus, the observer determines which region the perceptual effect is in and emits the associated response. The decision bound is the partition between competing response regions (Ashby, 1992; Ashby & Gott, 1988; Ashby & Lee, 1991, 1992; Ashby & Maddox, 1990, 1992, 1993; Ashby & Townsend, 1986; Maddox & Ashby, 1993). Thus, decision bound theory assumes that although exemplar information may be available, it is not used to make a categorization response. Instead, only a response label is retrieved.

Three Different Category Learning Tasks

Each of these theories has intuitive appeal, especially in some types of categorization tasks. For example, rule-based theories seem especially compelling when the rule that best separates the contrasting categories (i.e., the optimal rule) is easy to describe verbally (Ashby et al., 1998), and an exemplar-based memorization strategy seems ideal when the contrasting categories have only a few highly distinct exemplars. Not surprisingly, proponents of the various theories have frequently collected data in exactly those tasks for which their pet theories seem best suited. If there is only one category learning system, then this strategy is fine. However, if there are multiple systems, then the different tasks that have been used might load differently on the different systems. In this case, two researchers arguing that their data best supports their own theory might both be correct. As we will see below, there is neuropsychological and neuroimaging evidence supporting this prediction. So, before we examine the single versus multiple systems debate within the categorization literature, we take some time to describe three different types of categorization tasks that each seems ideally suited to the specific psychological processes hypothesized by the different theories.

As mentioned above, rule-based theories seem most compelling in tasks in which the rule that best separates the contrasting categories (i.e., the optimal rule) is easy to describe verbally (Ashby et al., 1998). As a result, observers can learn the category structures via an explicit process of hypothesis testing (Bruner et al., 1956) or theory construction and testing (Murphy & Medin, 1985). Figure 7 shows the stimuli and category structure of a recent rule-based task that used 8 exemplars per category (Waldron & Ashby, in press). The categorization stimuli were colored geometric figures presented on a colored background. The stimuli varied on four binary-valued dimensions: background color (blue or yellow; depicted as light or dark gray,

respectively), embedded symbol color (red or green; depicted as black or white, respectively), symbol numerosity (1 or 2), and symbol shape (square or circle). This yielded a total of 16 possible stimuli. To create rule-based category structures, one dimension is selected arbitrarily to be relevant. The two values on that dimension are then assigned to the two contrasting categories. At the end of training, observers are able to describe the rule they used in rule-based tasks quite accurately. Most categorization tasks used in studies that have argued for rule-based learning have been designed in a similar fashion (e.g., Bruner et al., 1956; Salatas & Bourne, 1974), as are virtually all categorization tasks used in neuropsychological assessment, including the well known Wisconsin Card Sorting Test (e.g., Grant & Berg, 1948; Kolb & Whishaw, 1990).

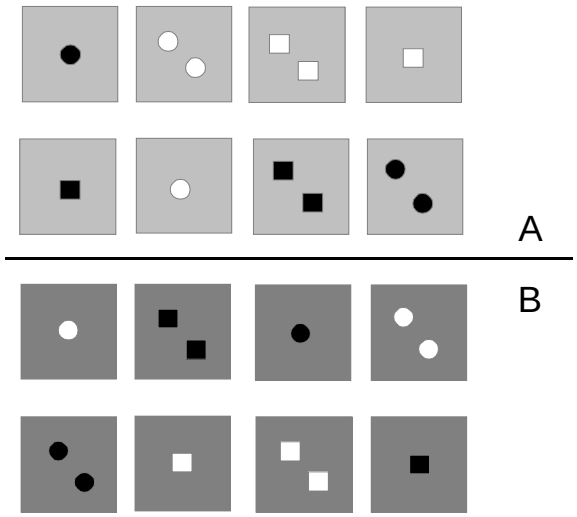

Figure 7. Category structure of a rule-based category learning task. The optimal explicit rule is: Respond A if the background color is blue (depicted as light gray), and respond B if the background color is yellow (depicted as dark gray).

Information-integration tasks are those in which accuracy is maximized only if information from two or more stimulus components (or dimensions) must be integrated at some pre-decisional stage (Ashby & Gott, 1988; Shaw, 1982). A conjunction rule (e.g., respond A if the stimulus is small on dimension *x* and small on dimension *y*) is a rule-based task rather than an information-integration task because separate decisions are first made about each dimension (e.g., small or large) and then the outcome of these decisions is combined (integration is not pre-decisional). In many cases, the optimal rule in information-integration tasks is difficult or impossible to describe verbally (Ashby et al., 1998). That people readily learn such category structures seems problematic for rule-based theories, but not for prototype, exemplar, or decision bound theories. The neuropsychological data reviewed below suggests that performance in such tasks is qualitatively different depending on the size of the categories – in particular, when a category contains only a few highly distinct exemplars, memorization is feasible. However, when the relevant categories contain many exemplars (e.g., hundreds), memorization is less efficient. An exemplar strategy seems especially plausible when the categories contain only a few highly distinct exemplars. Not surprisingly, most articles arguing for exemplar-based category learning have used such designs (e.g., Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1986; Smith & Minda, 2000).

Figure 8 shows the stimuli and category structure of a recent information-integration task that used only 8 exemplars per category (Waldron & Ashby, in press). The categorization stimuli were the same as in Figure 7. To create these category structures, one dimension was arbitrarily selected to be irrelevant. For example, in Figure 8, the irrelevant dimension is symbol shape. Next, one level on each relevant dimension was arbitrarily assigned a value of +1 and the other level was assigned a value of 0. In Figure 8, a background color of blue (depicted as light gray), a symbol color of green (depicted as white), and a symbol number of 2 were all assigned a value of +1. Finally, the category assignments were determined by the following rule:

The stimulus belongs to Category A if the sum of values on the relevant dimensions > 1.5;
Otherwise it belongs to Category B.

This rule is readily learned by healthy young adults, but even after achieving perfect performance, they can virtually never accurately describe the rule they used.

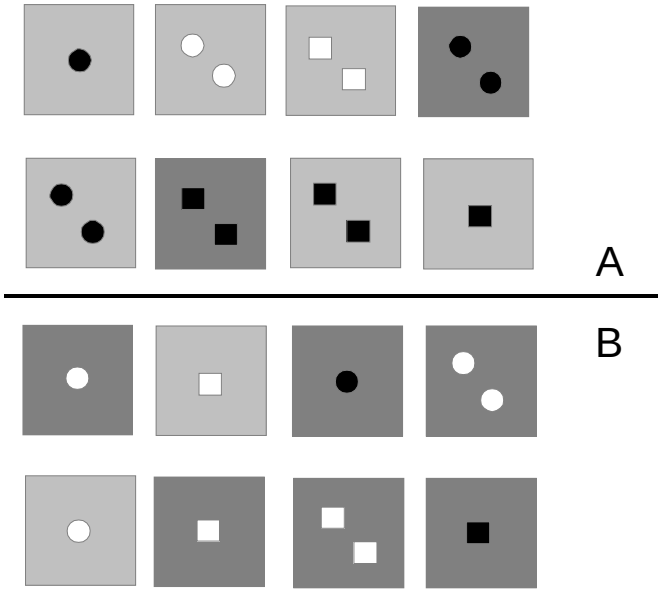

Figure 8. **Category structure of an information integration category learning task with only a few exemplars in each category.**

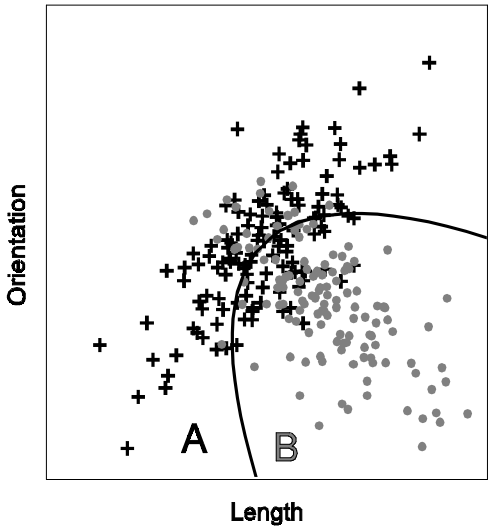

Figure 9. **Category structure of an information integration category learning task with many exemplars per category.** Each stimulus is a line that varies across trials in length and orientation. Every black plus depicts the length and orientation of a line in Category A and every gray dot depicts the length and orientation of a line in Category B. The quadratic curve is the boundary that maximizes accuracy.

When there are many exemplars in each category, memorization strategies, which are necessarily exemplar-based, become more difficult to implement. In these situations, it seems especially plausible that observers learn to associate category labels with regions of perceptual space (as predicted by decision bound theory). Figure 9 shows the category structure of an information-integration categorization task in which there are hundreds of exemplars in each category (first developed by Ashby & Gott, 1988). In this experiment, each stimulus is a line that varies across trials in length and orientation. Each cross in Figure 9 denotes the length and orientation of an exemplar in Category A and each dot denotes the length and orientation of an exemplar in Category B**.** The categories overlap, so perfect accuracy is impossible in this example. Even so, the quadratic curve is the boundary that maximizes response accuracy. This curve is difficult to describe verbally, so this is an information-integration task. Many of the studies supporting decision bound theory have used this randomization design (Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992; Maddox & Ashby, 1993).

A prototype abstraction process does not work well in the Figure 9 experiment because prototype theory always predicts linear decision bounds (Ashby & Gott, 1988), and there is much data showing that quadratic bounds give a much better account of the resulting data than linear bounds (Ashby & Maddox, 1992). A prototype abstraction process seems most plausible in prototype distortion tasks in which each category is

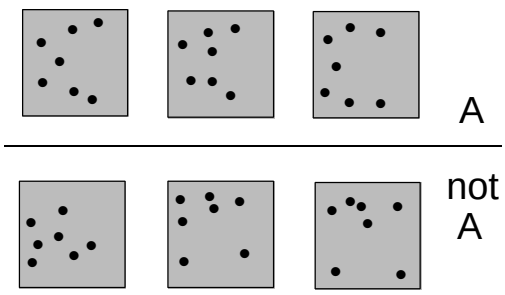

Figure 10. Some exemplars from a prototype distortion category learning task with random dot patterns.

created by first defining a category prototype and then creating the category members by randomly distorting these prototypes. In the most popular version of prototype distortion tasks the category exemplars are random dot patterns (Posner & Keele, 1968, 1970). An example of the random dot pattern task is shown in Figure 10. To begin, many stimuli are created by randomly placing a number of dots on the display. One of these stimuli is then chosen as the prototype for Category A. The others become stimuli not belonging to Category A. The other Category A exemplars are then created by randomly perturbing the position of each dot in the Category A prototype. Categories created from these random dot patterns have been especially popular with prototype theorists (e.g., Homa & Cultice, 1984; Homa, Cross, Cornell, Goldman, & Schwartz, 1973; Homa et al., 1981; Posner & Keele, 1968, 1970).

Explicit Versus Implicit Category Learning

As mentioned previously, many of the current theories that postulate multiple category learning systems propose separate explicit and implicit subsystems. The multiple memory systems literature also frequently uses the terms explicit and implicit, but usually in a slightly different fashion. So before proceeding further, we briefly discuss the existing criteria that are used to determine whether category learning is explicit or implicit.

There is widespread agreement, within both the category learning and memory literatures, that explicit processing requires conscious awareness (e.g., Ashby et al., 1998; Cohen & Squire, 1980). The disagreements relate more to how implicit processing is defined. Many memory theorists adopt the strong criteria that a memory is implicit only if there is no conscious awareness of its details and there is no knowledge that a memory has even been stored (e.g., Schacter, 1987). In a typical categorization task, for example any of those described in the last section, these criteria are impossible to meet when trial-by-trial feedback is provided (as it usually is). When an observer receives feedback that a response is correct, then this alone makes it obvious that learning has occurred, even if there is no internal access to the system that is mediating this learning. Thus, in category learning, a weaker criterion for implicit learning is typically used in which the observer is required only to have no conscious access to the nature of the learning, even though he or she would be expected to know that some learning had occurred.

The stronger criteria for implicit processing that have been adopted in much of the memory literature could be applied in unsupervised category learning tasks, in which no trial-by-trial feedback of any kind is provided. In the typical unsupervised task, observers are told the number of contrasting categories and are asked to assign stimuli to these categories, but are never told whether a particular response is correct or incorrect. Free sorting is a similar, but more unstructured task in which participants are not told the number of contrasting categories (e.g., Ashby & Maddox, 1998). Although unsupervised and free sorting tasks are ideal for using the stricter criteria to test for implicit learning, so far, the only learning that has been demonstrated in such tasks is explicit (Ashby, Queller, & Berretty, 1999; Medin, Wattenmaker, & Hampson, 1997).

One danger with equating explicit processing with conscious awareness is that this shifts the debate from how to define ‘explicit’ to how to define ‘conscious awareness’. Ashby et al. (1998) suggested that one pragmatic solution to this problem is to operationally define a categorization rule as explicit if it is easy to describe verbally. By this criterion, the rule that separates the categories in Figure 7 is explicit, whereas the rules best separating the categories in Figures 8 and 9 are implicit. This definition works well in most cases, but it seems unlikely that verbalizability should be a requirement for explicit reasoning. For example, the insight displayed by Köhler’s (1925) famous apes seems an obvious example of explicit reasoning in the absence of language. So ultimately, a theoretically motivated criterion for conscious awareness is needed.

One way to develop a theory of conscious awareness is by exploiting the relationship between awareness

and working memory. For example, the contents of working memory are clearly accessible to conscious awareness. In fact, because of its close association to executive attention, a strong argument can be made that the contents of working memory *define* our conscious awareness. When we say that we are consciously aware of some object or event, we mean that our executive attention has been directed to that stimulus. Its representation in our working memory gives it a moment-to-moment permanence. Working memory makes it possible to link events in the immediate past with those in the present, and it allows us to anticipate events in the near future. All of these are defining properties of conscious awareness.

The association between working memory and the prefrontal cortex makes it possible to formulate cognitive neuroscience models of consciousness. The most influential such model was developed by Francis Crick and Christof Koch (Crick & Koch, 1990; 1995a; 1998). The Crick-Koch hypothesis states that one can have conscious awareness only of activity in brain areas that project directly to the prefrontal cortex[4]. Primary visual cortex (Area V1) does not project directly to the prefrontal cortex, so the Crick-Koch hypothesis asserts that we cannot be consciously aware of activity in V1. Crick and Koch (1995a; 1998) described evidence in support of this prediction. Of course, many other brain regions also do not project directly to the prefrontal cortex. For example, the basal ganglia do not project directly to the prefrontal cortex (i.e., they first project through the thalamus), so the Crick-Koch hypothesis predicts that we are not aware of activity within the basal ganglia. Memory theorists believe that the basal ganglia mediate procedural memories (Jahanshahi et al., 1992; Mishkin et al., 1984; Saint-Cyr et al., 1988; Willingham et al., 1989), so the Crick-Koch hypothesis provides an explanation of why we don't seem to be aware of procedural (e.g., motor) learning.

Category Learning and Memory

The notion that there may be multiple category learning systems goes back at least to 1978, when Brooks hypothesized that category learning is mediated by separate "deliberate, verbal, analytic control processes and implicit, intuitive, nonanalytic processes" (p.207). Nevertheless, most quantitative accounts of category learning have assumed the existence of a single system (e.g., Estes, 1986a; Hintzman, 1986; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). Recently, however, quantitative models that assume multiple category learning systems have been developed (e.g., Ashby et al., 1998; Erickson & Kruschke, 1998). For example, Ashby et al. (1998) proposed a formal neuropsychological theory of multiple category learning systems called COVIS (COmpetition between Verbal and Implicit Systems), which assumes separate explicit (rule-based) and implicit (procedural learning-based) systems. In response to these multiple systems proposals, Nosofsky and Zaki (1998) and Nosofsky and Johansen (in press) argued that single system (exemplar) models can account for many of the phenomena that have been used to support the notion of multiple systems.

Another way to study category learning systems is to emphasize the relationship between category learning and memory. Of course, every category learning system requires memory. In fact, one could characterize category learning as the process of establishing some durable record – that is, a memory – of the structure of the relevant categories, or possibly of a rule for correctly assigning new stimuli to one of the categories. Since much is now known about the neurobiology of memory, this might be a way to learn quickly about the neurobiology of category learning.

The multiple memory systems that have been proposed each are thought to have a distinct neural basis.

---

[4] Crick and Koch (1998) did not take the strong position that working memory is necessary for conscious awareness. Even so, they did argue that some short-term memory store is required. However, they left open the possibility that an extremely transient iconic memory might be sufficient.

Cognitive neuroscience models of working memory focus on prefrontal cortex (e.g., Fuster, 1989; Goldman-Rakic, 1987, 1995), declarative memory models focus on the hippocampus and other medial temporal lobe structures (e.g., Gloor, 1997; Gluck & Myers, 1997; McClelland et al., 1995; Polster et al., 1991; Squire & Alvarez, 1995), procedural memory models focus on the basal ganglia (Jahanshahi et al., 1992; Mishkin et al., 1984; Saint-Cyr et al.,1988; Willingham et al., 1989), and models of the perceptual representation system focus on visual cortex (e.g., Curran & Schacter, 1997; Schacter, 1994).

In addition, each of the category learning theories described above maps in a natural way onto a different one of these memory systems. To learn and apply explicit rules, one must construct and maintain them in working memory. Executive attention is also required to select and switch among alternative rules. Thus, rule-based theories depend on working memory. Exemplar theory assumes people store and access detailed representations of specific exemplars they have seen. The declarative memory system seems tailor made for this type of memory encoding and storage. Indeed, it has specifically been proposed that medial temporal lobe structures (i.e., the hippocampus) mediate the encoding and consolidation of exemplar memories (Pickering, 1997). On the other hand, declarative memory retrieval is typically thought to occur with conscious awareness (e.g., Cohen & Squire, 1980), whereas exemplar theorists are careful to assume that activation of the exemplar memories does *not* require awareness (e.g., Nosofsky & Alfonso-Reese, 1999; Nosofsky & Zaki, 1998; Nosofsky, 1986).

Decision bound theory assumes people learn to associate abstract response programs (e.g., response labels) with groups of similar stimuli (Ashby & Waldron, 1999). Thus, the stored memories are of stimulus-response associations, rather than of rules or previously seen exemplars. This is a form of procedural memory (Ashby et al., 1998).

The prototype abstraction process assumed by prototype theory is perhaps the most difficult to map onto existing accounts of memory. The memory of a prototype is durable, so working memory, by itself is insufficient. Prototype theorists also have been clear that the prototype might not correspond exactly to any previously seen exemplar, which rules out simple declarative memory. Finally, prototypes are not tied to responses in any direct way, so procedural memory can also be ruled out. While it is not clear that such a result is necessary, we will present evidence later that prototype abstraction depends, at least sometimes, on perceptual learning, and as a result, on the perceptual representation memory system.

It is important to point out that even if multiple memory systems participate in category learning, this does not necessarily imply that there are multiple category learning systems. For example, it is logically possible that a single category learning system accesses different memory systems in different category learning tasks. Such a model could predict double or triple dissociations across tasks. As mentioned in the last major section, however, such a model also shares many properties with a multiple systems perspective. As such, it would probably lie somewhere in the middle of the continuum between pure single system and pure multiple system models. In our view, it would be counterproductive to place a sharp boundary on this continuum in an attempt to produce a criterion that classifies every model as postulating either single or multiple systems. Instead the goal, in all areas of learning and memory, should be to understand how humans perform this vitally important skill. In the case of category learning, understanding what memory systems are involved is an important first step in this process.

A good example of this blurring between single and multiple systems can be seen with prototype abstraction. If this process is mediated by a perceptual representation system that depends on perceptual learning in visual cortex, then it is not clear that prototype abstraction would meet our criteria as a separate system. When the stimuli are visual in nature, then any category learning system must receive input from the visual system. If some category learning system X depends on input from the brain region mediating prototype abstraction, then system X and the prototype abstraction system would not be mediated by separate neural pathways – a criterion that we earlier decided was a necessary condition for separate systems. For example, under this scenario, a double dissociation between system X and the prototype system should be

impossible. Damage to the neural structures downstream from visual cortex that mediate system X should induce deficits in category learning tasks mediated by system X, but not in prototype abstraction tasks. On the other hand, damage to visual cortex should impair all types of visual category learning. Thus, if prototype abstraction is mediated within visual cortex, then any group impaired in prototype abstraction should also be impaired on all other category learning tasks. In addition, it should be extremely difficult, or impossible, to find neuropsychological patient groups that are impaired in prototype abstraction, but not in other types of category learning. As we will shortly see, this latter prediction is supported by current neuropsychological category learning data.

Under the assumption that the category learning tasks described above differentially load on different memory systems, then theoretically, it should be possible to find neuropsychological populations that establish at least a triple dissociation across the tasks. Ashby and Ell (under review) reviewed the current neuropsychological category learning data to test this prediction. Presently, there is extensive category learning data on only a few neuropsychological populations. The best data come from four different groups: 1) patients with frontal lobe lesions, 2) patients with medial temporal lobe amnesia, and two types of patients suffering from a disease of the basal ganglia – either 3) Parkinson's or 4) Huntington's disease. Table 1 summarizes the performance of these groups on the three different types of category learning tasks.

Table 1. Performance of various neuropsychological populations on three category learning tasks

Neuropsychological Group	Task			
	Rule-Based	Information-integration		Prototype Distortion
		Many Exemplars	Few Exemplars	
Frontal Lobe Lesions	Impaired	?	Normal	?
Parkinson's Disease	Impaired	Impaired	Impaired	Normal
Huntington's Disease	Impaired	Impaired	Impaired	Normal
Medial Temporal Lobe Amnesia	Normal	Normal	Late Training Deficit	Normal

Note first that Table 1 does not establish a triple dissociation. At best, one could argue from Table 1 only for a double dissociation – between frontal lobe patients and medial temporal lobe amnesiacs on rule-based tasks and information-integration tasks with few exemplars per category. Specifically, frontal patients are impaired on rule-based tasks (e.g., the Wisconsin Card Sorting Test; Kolb & Whishaw, 1990) but medial temporal lobe amnesiacs are normal (e.g., Janowsky, Kritchevsky, & Squire, 1989; Leng & Parkin, 1988). At the same time, the available data on information-integration tasks with few exemplars per category indicates that frontal patients are normal (Knowlton, Mangels, & Squire, 1996), but medial temporal lobe amnesiacs are impaired (i.e., they show a late-training deficit -- that is, they learn normally during the first 50 trials or so, but thereafter show impaired learning relative to age-matched controls; Knowlton, Squire, & Gluck, 1994). Therefore, the neuropsychological data support the hypothesis that at least two memory systems participate in category learning. Of course, until more data are collected on the information-integration tasks, this

conclusion must be considered tentative.

Note also that Table 1 supports the prediction that it should be difficult to find patient groups that are impaired in the prototype distortion task, but not in the other types of tasks. We know of no data on the performance of frontal lobe patients in prototype distortion tasks, but if learning in these tasks is mediated within visual cortex, then frontal patients should not be impaired in prototype distortion tasks.

If three or more memory systems participate in category learning, then why doesn't Table 1 document a triple dissociation? There are several reasons why a triple dissociation might not be observed in Table 1, even if multiple memory systems are involved. First, Table 1 is incomplete. There are several cells with no known data. For example, we know of no data on the performance of frontal patients in information-integration tasks with many exemplars per category. Conclusions in some other cells are based on very little data. As mentioned above, this is the case for the late-training deficit reported for medial temporal lobe amnesiacs in information-integration tasks with few exemplars per category. Second, even with unlimited data in each cell, there is no guarantee that these are four patient groups appropriate for establishing a triple dissociation. The groups included in Table 1 were selected because they are the groups for which there is the most current data, rather than for some theoretical purpose. For example, the ideal groups might each have focal damage to a different memory system. This condition is surely not met for the Table 1 groups. For example, Parkinson's and Huntington's diseases affect similar structures (i.e., the basal ganglia). Of course, to select groups that satisfy this condition requires specific hypotheses about the neural structures and pathways that mediate the putatively separate systems. There are two ways to generate such hypotheses. One is to use Table 1 and recent neuroimaging results to make such inferences, and another is to examine current neuropsychological theories of multiple systems in category learning. We follow these two approaches in the next section.

The Neurobiological Bases of Category Learning

Patients with frontal or basal ganglia dysfunction are impaired in rule-based tasks (e.g., Brown & Marsden, 1988; Cools et al.,1984; Kolb & Whishaw, 1990; Robinson, Heaton, Lehman, & Stilson, 1980), but patients with medial temporal lobe damage are normal in this type of category learning task (e.g., Janowsky et al., 1989; Leng & Parkin, 1988). Thus, an obvious first hypothesis is that the prefrontal cortex and the basal ganglia participate in this type of learning, but the medial temporal lobes do not. Converging evidence for the hypothesis that these are important structures in rule-based category learning comes from several sources. First, an fMRI study of a rule-based task similar to the Wisconsin Card Sorting Test showed activation in the right dorsal-lateral prefrontal cortex, the anterior cingulate, and the right caudate nucleus (i.e., head) (among other regions) (Rao et al., 1997). Second, many studies have implicated these structures as key components of executive attention (Posner & Petersen, 1990) and working memory (e.g., Fuster, 1989; Goldman-Rakic, 1987), both of which are likely to be critically important to the explicit processes of rule formation and testing that are assumed to mediate rule-based category learning. Third, a recent neuroimaging study identified the (dorsal) anterior cingulate as the site of hypothesis generation in a rule-based category-learning task (Elliott & Dolan, 1998). Fourth, lesion studies in rats implicate the dorsal caudate nucleus in rule switching (Winocur & Eskes, 1998).

Next, note that in information integration tasks with large categories, only patients with basal ganglia dysfunction are known to be impaired (Filoteo, Maddox, & Davis, submitted; Maddox & Filoteo, in press). In particular, medial temporal lobe patients are normal (Filoteo, Maddox, & Davis, in press). So a first hypothesis should be that the basal ganglia are critical in this task, but the medial temporal lobes are not. If the number of exemplars per category is reduced in this task to a small number (e.g., 4 to 8), then medial temporal lobe amnesiacs show late training deficits – that is, they learn normally during the first 50 trials or

so, but thereafter show impaired learning relative to age-matched controls (Knowlton et al., 1994). An obvious possibility in this case, is that normal observers begin memorizing responses to at least a few of the more distinctive stimuli – a strategy that is not available to the medial temporal lobe amnesiacs, and which is either not helpful or impossible when the categories contain many exemplars. Since patients with basal ganglia dysfunction are also impaired with small categories requiring information-integration (Knowlton, Mangels et al., 1996; Knowlton, Squire et al., 1996), a first hypothesis should be that learning in such tasks depends on the basal ganglia and on medial temporal lobe structures. The hypothesis that the basal ganglia are active in information-integration tasks was supported by Poldrack, Prabhakaran, et al. (1999), who used fMRI to measure neural activation at four different time points of learning in a probabilistic version of the information-integration task with few exemplars per category. They reported learning related changes within prefrontal cortex and in the tail of the right caudate nucleus. Interestingly, they also reported a simultaneous suppression of activity within the medial temporal lobes. Thus, the available neuroimaging data predict that the deficits of basal ganglia disease patients in information-integration tasks may arise from dysfunction in the tail of the caudate nucleus.

Table 2. Brain Regions that Current Neuropsychological Data Implicate in the Various Category Learning Tasks

Brain Region	Task			
	Rule-Based	Information-integration		Prototype Distortion
		Many Exemplars	Few Exemplars	
Prefrontal Cortex	X			
Visual Cortex				X
Basal Ganglia	X	X	X	
Medial Temporal Lobe			X	

Finally, none of these four patient groups are impaired on the prototype distortion tasks, which suggests that learning on these tasks does not depend on an intact medial temporal lobe or basal ganglia (Knowlton, Squire et al., 1996; Knowlton, Ramus, & Squire, 1992; Kolodny, 1994; Meulemans, Peigneux, & Van der Linden, 1998). As mentioned above, it has been suggested instead that learning might depend on the perceptual representation memory system – through a perceptual learning process (Knowlton, Squire et al., 1996). In the random dot pattern experiments, this makes sense because all category A exemplars are created by randomly perturbing the positions of the dots that form the category A prototype (see Figure 9). Thus, if there are cells in visual cortex that respond strongly to the category A prototype, they are also likely to respond to the other category A exemplars, and perceptual learning will increase their response. If this occurs, the observer could perform well in this task by responding “yes” to any stimulus that elicits a strong feeling of visual familiarity. Recent fMRI studies of subjects in prototype distortion tasks show learning related changes

in visual cortex (Reber, Stark, & Squire, 1998), and are thus consistent with this hypothesis.

Table 2 summarizes the neural implications of the current neuropsychological and neuroimaging data. Note that Table 2 is consistent with current theories about the neurobiological bases of memory – in particular, that the basal ganglia are important in procedural memory, and the medial temporal lobes are critical for declarative memory. Despite the arguments and evidence in support of the Table 2 conclusions, however, much more work is needed before Table 2 can be considered more than speculative.

The Explicit System

Several recent neurospsychological theories agree with some of the same conclusions drawn in Table 2. For example, Figure 11 describes a recent neurobiological model of the explicit system(Ashby et al., 1998, Ashby, Isen, & Turken, 1999). The key structures are the anterior cingulate, the prefrontal cortex, and the head of the caudate nucleus. Figure 11 shows the model during a trial of the rule-based category learning task illustrated in Figure 7. Various salient explicit rules reverberate in working memory loops between prefrontal cortex (PFC) and thalamus (Alexander, Delong, & Strick, 1986). In Figure 11, one such loop maintains the representation of a rule focusing on the shape of the symbols and one loop maintains a rule focusing on symbol number. An excitatory projection from the PFC to the head of the caudate nucleus prevents the globus pallidus from interrupting these loops. The anterior cingulate selects new explicit rules to load into working memory, and the head of the caudate nucleus mediates the switch from one active loop to another (facilitated by dopamine projections from the ventral tegmental area and the substantia nigra).

The Figure 11 model is consistent with the neuroimaging data described in the previous section, and it accounts for the rule-based category learning deficits described in Table 1. First, of course, it is obvious that the model predicts that patients with lesions of the prefrontal cortex will be impaired on rule-based category learning tasks. It also predicts that the deficits seen in Parkinson's disease are due to dysfunction in the head of the caudate nucleus. Postmortem autopsy reveals that damage to the head of the caudate is especially severe in Parkinson's disease (van Domburg & ten Donkelaar, 1991), so the model predicts that this group should show widespread and profound deficits on rule-based categorization tasks. The neuropsychological evidence strongly supports this prediction (e.g., on the WCST; Brown & Marsden, 1988; Cools et al., 1984). In fact, the model described in Figure 11 predicts that, because of its reciprocal connection to the prefrontal cortex, many of the well documented "frontal-like" symptoms of Parkinson's disease might actually be due to damage in the head of the caudate nucleus.

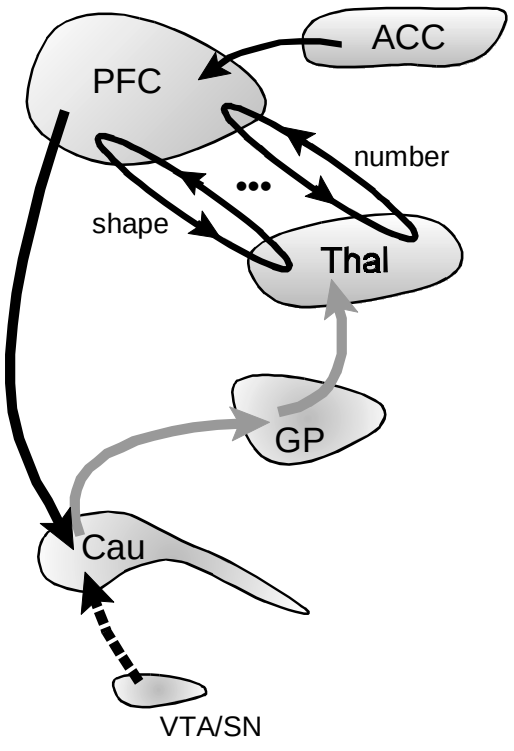

Figure 11. A model of the explicit category learning system. Black projections are excitatory, gray projections are inhibitory, and dashed projections are dopaminergic. PFC = prefrontal cortex, ACC = anterior cingulate cortex, Thal = thalamus, GP = globus pallidus, Cau = caudate nucleus, VTA = ventral tegmental area, SN = substantia nigra.

The Procedural Learning System

Figure 12 shows the circuit of a putative procedural memory-based category learning system (proposed by Ashby et al.,1998; Ashby & Waldron, 1999). The key structure in this model is the caudate nucleus, a major input structure within the basal ganglia. In primates, all of extrastriate visual cortex projects directly to the tail of the caudate nucleus, with about 10,000 visual cortical cells converging on each caudate cell (Wilson, 1995). Cells in the tail of the caudate (i.e., medium spiny cells) then project to prefrontal and premotor cortex (via the globus pallidus and thalamus; e.g., Alexander et al., 1986). The model assumes that, through a procedural learning process, each caudate unit learns to associate a category label, or perhaps an abstract motor program, with a large group of visual cortical cells (i.e., all that project to it). This learning is thought to be facilitated by a reward mediated dopamine signal from the substantia nigra (pars compacta) (e.g., Wickens, 1993).

Lesions of the tail of the caudate, in both rats and monkeys, impair the ability of the animal to associate one motor response with one visual stimulus and a different response with some other stimulus (e.g., vertical versus horizontal lines; McDonald & White, 1993, 1994; Packard et al., 1989; Packard & McGaugh, 1992). For example, in one study, rats with lesions in the tail of the caudate could not learn to discriminate between safe and unsafe platforms in the Morris water maze when the safe platform was marked with horizontal lines and the unsafe platform was marked with vertical lines (Packard & McGaugh, 1992). The same animals learned normally, however, when the cues signaling which platform was safe were spatial. Because the tail of the caudate nucleus is not a classic visual area, it is unlikely that these animals have an impaired ability to perceive the stimuli. Rather, it seems more likely that their deficit is in learning the appropriate stimulus-response associations. The Figure 12 model predicts that this same type of stimulus-response association learning mediates performance in the information-integration category learning tasks described in Figures 8 and 9 .

The Figure 12 model accounts for the category learning deficits of Parkinson's and Huntington's disease patients in information-integration tasks because both of these populations suffer from caudate dysfunction. It also explains why frontal patients and medial temporal lobe amnesiacs are relatively normal in these tasks – that is, because neither prefrontal cortex nor medial temporal lobe structures play a prominent role in the Figure 12 model.

The model shown in Figure 12 is strictly a model of visual category learning. However, it is feasible that a similar system exists in the other modalities, since they almost all also project directly to the basal ganglia, and then indirectly to frontal cortical areas (again via the globus pallidus and the thalamus; e.g., Chudler, Sugiyama, & Dong, 1995). The main difference is in where within the basal ganglia they initially project. For example, auditory cortex projects directly to the body of the caudate (i.e., rather than to the

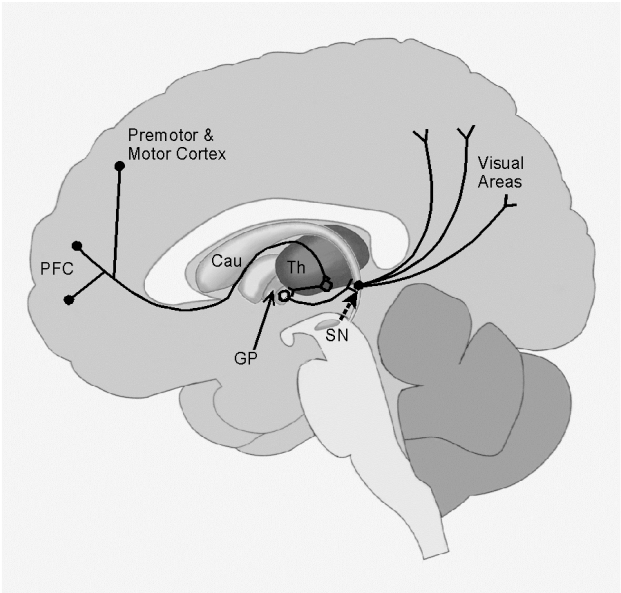

Figure 12. A procedural-memory-based category learning system. Excitatory projections end in solid circles, inhibitory projections end in open circles, and dopaminergic projections are dashed. PFC = prefrontal cortex, Cau = caudate nucleus, GP = globus pallidus, and Th = thalamus.

tail; Arnalud, Jeantet, Arsaut, & Demotes-Mainard, 1996).

The Perceptual Representation and Medial Temporal Lobe Category Learning Systems

No one has yet proposed a detailed category learning model that is based on the perceptual representation memory system. However, as noted above, based on work in the memory literature, it seems likely that such a category learning system would be based in sensory cortex (Curran & Schacter, 1996; Schacter, 1994).

In cognitive psychology, one of the most popular and influential theories of category learning is exemplar theory (Brooks, 1978; Estes, 1986b; Medin & Schaffer, 1978; Nosofsky, 1986), which assumes that categorization decisions are made by accessing memory representations of previously seen exemplars. Although most exemplar theorists have not taken a strong stand about the neural basis by which these memory representations are encoded, those who have assume that the medial temporal lobes are heavily involved (e.g., Pickering, 1997). Despite the popularity of exemplar theory within cognitive psychology however, the most convincing direct neuropsychological evidence in support of a key role of the medial temporal lobes in category learning remains the late-training deficit identified in Table 1 (Knowlton et al., 1994). Even so, this finding is not without controversy, since a recent neuroimaging study found suppression of medial temporal lobe activity in this same task (Poldrack, Prabhakaran et al., 1999). Although many neurobiological models of hippocampal function have been proposed, there have been only a few attempts to apply these models to category learning (Gluck, Oliver, & Myers, 1996; Pickering, 1997).

Summary

Although hotly debated, the question of whether human category learning is mediated by one or several category learning systems is currently unresolved. Recent neuropsychological and neuroimaging data support the hypothesis that different memory systems may participate in different types of category learning tasks, but there is little current data that allow stronger conclusions to be drawn.

IV. Conclusions

The debate as to whether learning and memory is mediated by one or several distinct systems is being waged in many areas of cognitive psychology. Although the setting of these debates differs – from memory to function learning to discrimination learning to category learning – there are a number of common themes that tie all these debates together. First, the methodologies that are most appropriate for testing between single and multiple systems are the same no matter what the domain. For example, the fixed-point property and double dissociations are powerful tools that can (and should) be used in any area trying to resolve this issue. Second, regardless of the field, it is unrealistic to expect any single study to resolve the single versus multiple systems debate. Instead, it is imperative that all available evidence be evaluated simultaneously. For example, given three data sets that all seemingly point toward multiple systems, it is not valuable to show that there exists three different single-system models that are each consistent with one set of data. The important question is really: does the single model that best accounts for all three data sets simultaneously postulate one or multiple systems of learning and memory? Third, all fields engaged in the single versus multiple systems debate should look seriously toward cognitive neuroscience as a way to add more constraints to the existing models, and as a

mechanism for building bridges to other related areas of cognitive psychology.

In our view, however it is resolved, the single versus multiple systems debate is likely to prove a valuable experience for whatever field engages it. The benefit of asking whether there are single or multiple systems of learning and memory is that this question organizes new research efforts, it encourages collecting data of a qualitatively different nature than has been collected in the past, and it also immediately ties the field in question to the memory literature and a variety of other seemingly disparate literatures. The one danger that must be resisted is engaging in endless debate about what constitutes a system. One's definition of system will obviously affect how the single versus multiple systems question is answered, but the process of asking, and all its associated benefits, is far more important than the answer itself.

V. References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9*, 357-381.

Arnalud, 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.

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Ell, S. W. (under review). The neurobiological basis of category learning.

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

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response time with multidimensional stimuli. *Perception & Psychophysics, 55*, 11-27.

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. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review,106*, 529-550.

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 between identification, similarity, and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General, 121*, 385-393.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception & 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 & 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. (1997). Stimulus Categorization. In M. H. Birnbaum (Ed.), *Handbook of perception & cognition: Judgment, decision making, and measurement* (Vol. 3). New York: Academic Press.

Ashby, F. G., Queller, S., & Berretty, P. T. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*, 1178-1199.

Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154-179.

Ashby, F. G. & Waldron, E. M. (1999). The nature of implicit categorization. *Psychonomic Bulletin & Review, 6*, 363-378.

Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Paulesu, E., Frackowaik, R, S. J., & Dolan, R. J. (1993). Investigatrions of the functional anatomy of attention using the Stroop test. *Neuropsychologia, 33*, 907-922.

Bourne, L. E. (1966). *Human conceptual behavior*. Boston: Allyn and Bacon.

Brooks, L. (1978) Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.) *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

Brown, R. G. & Marsden, C. D. (1988). Internal versus external cures and the control of attention in Parkinson's disease. *Brain, 111*, 323-345.

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

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40*, 227-259.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 638-648.

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, N. J. & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesics: Dissociation of knowing how and knowing that. *Science, 210*, 207-210.

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.

Corkin, S. (1965). Tactually-guided maze learning in man: effect of unilateral cortical excision and bilateral hippocampal lesions. *Neuropsychologia, 3*, 339-351.

Crick, F. & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci., 2*, 2263-275.

Crick, F. & Koch, C. (1995a). Are we aware of neural activity in primary visual cortex? *Nature, 375*, 121-123.

Crick, F. & Koch, C. (1998). Cousciousness and neuroscience. *Cerebral Cortex, 8*, 97-107.

Curran, T. & Schacter, D. L. (1996). Memory: Cognitive neuropsychological aspects. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral Neurology and Neuropsychology* (pp. 463-471). New York: McGraw-Hill.

Curran, T. & Schacter, D.L. (1997). Implicit memory: What must theories of amnesia explain? *Memory, 5*, 37-48.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review, 95*, 91-101.

Elliott, R. & Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *Neuroimage, 8*, 17-29.

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

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

Estes, W. K. (1986b). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115*, 155-174.

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

Falmagne, J. C. (1968). Note on a simple fixed-point property of binary mixtures. *British Journal of Mathematical & Statistical Psychology, 21*, 131-132.

Falmagne, J. C. & Theios, J. (1969). On attention and memory in reaction time experiments. *Acta Psychologia, 30*, 319-323.

Filoteo, J. V., Maddox, W. T., & Davis, J. (submitted). A possible role of the striatum in linear and nonlinear categorization rule learning: Evidence from patients with Huntington's disease. Manuscript under review.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (in press). Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society*.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar.* New York: McGraw-Hill.

Fuster, J. M. (1989). *The Prefrontal Cortex* (2nd Edition). New York: Raven Press.

Gaffan, D. (1974). Recognition impaired and association intact in the memory of monkeys after transsection of the fornix. *Journal of Comparative and Physiological Psychology, 86*, 1110-1109.

Gloor, P. (1997). *The Temporal Lobe and Limbic System*. New York: Oxford University Press.

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.

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

Goldman-Rakic, P. S. (1987). Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. in *Handbook of Physiology* (Plum, F. & Mountcastle, V., eds.), pp. 373-417, American Physiological Society.

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron, 14*, 477-485.

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology, 33*, 154-207.

Grant, D. A. & Berg, E. A. (1948). Behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psycholog, 38*, 404-411.

Hayes, N. A. & Broadbent, D. (1988). Two modes of learning for interactive tasks. *Cognition, 28*, 249-276.

Higham, P. A., Vokey, J. R., & Pritchard, J. L. (2000). Beyond dissociation logic: Evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General, 129*, 457-470.

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

Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: A theory. *Behavioral Biology, 12*, 421-442.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psycholog, 101*, 116-122.

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

Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 444-458.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the

abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 418-439.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs, 28*, 1-86.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

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.

Janowsky, J. S., Kritchevsky, A. P., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103,* 548-560.

Kendler, T. S. & Kendler, H. H. (1962). Inferential behavior in children as a function of age and subgoal constancy. *Journal of Experimental Psychology, 64,* 460-466.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (in press). Decision making and the evolution of memory: Multiple systems, multiple functions. *Psychological Review.*

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., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1,* 106-120.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273,* 1399-1402.

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 & Co.

Kolb, B. & Whishaw, I. Q. (1990). *Fundamentals of Human Neuropsychology* (3rd Ed.). New York: W. H. Freeman & Company.

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.

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

Leng, N. R. & Parkin, A. J. (1988). Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology, 27,* 359-362.

Lupker, S. J. & Theios, J. (1977). Further tests of a two-state model for choice reaction times. *Journal of Experimental Psychology: Human Perception & Performance, 3,* 496-504.

Maddox, W. T. & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics, 53,* 49-70.

Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional categorization. *Perception & Psychophysics, 60,* 620-637.

Maddox, W. T., & Filoteo, J. V. (in press). Striatal contribution to category learning: Quantitative modeling of simple linear and complex non-linear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society.*

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.

McCloskey, M. (1993). Theory and evidence in cognitive neuropsychology: A "radical" response to

Robertson, Knight, Rafal, and Shimamura (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 718-734.

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.

Medin, D. L., Alton, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 37-50.

Medin, D. L. & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68-85.

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. (1997). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19,* 242-279.

Meier, B., & Perrig, W. J. (2000). Low reliability of perceptual priming: Consequences for the interpretation of functional dissociations between explicit and implicit memory. *The Quarterly Journal of Experimental Psychology, 53A,* 211-233.

Meulemans, T., Peigneux, P., & Van der Linden, M. (1998). Preserved artificial grammar learning in Parkinson's disease. *Brain & Cognition, 37,* 109-112.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge, MA: Harvard Press.

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.

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

Myung, I. J. (1994). Maximum entropy interpretation of decision bound and context models of categorization. *Journal of Mathematical Psychology, 38,* 335-365.

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87-108.

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

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception and Psychophysics, 45,* 279-290.

Nosofsky, R. M., & Alfonso-Reese, L. A. (1999). Effects of similarity and practice on speeded classification response times and accuracies: Further tests of an exemplar-retrieval model. *Memory & Cognition, 27,* 78-93.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282-304.

Nosofsky, R. M. & Johansen, M. K. (in press). Exemplar-based accounts of "multiple-system"

phenomena in perceptual categorization. *Psychonomic Bulletin & Review*.

Nosofsky, R. M. & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9*, 247-255.

O'Keefe, J. & Nadel, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.

Packard, M. G., Hirsh, 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.

Pickering, A. D. (1997). New approaches to the study of amnesic patients: What can a neurofunctional philosophy and neural network methods offer? *Memory, 5*, 255-300.

Poldrack, R. A., Prabhakaran, V. Seger, C. A., & Gabrieli, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*, 564-574.

Poldrack, R. A., Selco, S. L., Field, J. E., & Cohen, N. J. (1999). The relationship between skill learning and repetition priming: Experimental and computational analyses. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 208-235.

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., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304-308.

Posner, M. I. & Petersen, S. E. (1990). Attention systems in the human brain. *Annual Review of Neurosci*ence, *13*, 25-42.

Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J., Cunningham, J. M., Cox, R. W., Stein, E. A., & Binder, J. R. (1997) Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport, 8*, 1987-1993.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81*, 115-119.

Reber, P. J., Stark, C. E. L., and Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory, 5*, 420-428.

Redington, M. (2000). Not evidence for separable controlled and automatic influences in artificial grammar learning: Comment on Higham, Vokey, and Pritchard (2000). *Journal of Experimental Psychology: General*, *129*, 471-475.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 189-221.

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.

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

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology*. London: Academic Press.

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. (1974). Learning conceptual rules III: Processes contributing to rule

difficulty. *Memory and Cognition, 2*, 549-553.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501-518.

Schacter, D. L. (1994) Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems 1994* (pp. 233-268). Cambridge: MIT Press.

Shallice, T. (1988). *From neuropsychology to mental structure.* New York: Plenum Press.

Shaw, M. L. (1982). Attending to multiple sources of information. I: The integration of information in decision making. *Cognitive Psychology, 14*, 353-409.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of "dot-pattern" classification and recognition. *Journal of Experimental Psychology: General, 121*, 278-304.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

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

Smith, D. J. & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 3-27.

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.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), Attention and performance II, *Acta Psychologica, 30*, 276-315.

Teuber, H. L. (1955). Physiological psychology. *Annual Review of Psychology,* 6, 267-296.

Townsend, J. T. & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* New York: Cambridge University Press.

Tulving, E. & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*, 302-306.

van Domburg, P. H. M. F. & ten Donkelaar, H. J. (1991). *The Human Substantia Nigra and Ventral Tegmental Area,* Berlin. Springer-Verlag.

Van Orden, G. C., Pennington, B. F. , & Stone, G. O. (in press). What do double dissociations prove? *Cognitive Science.*

Waldron, E. M. & Ashby, F. G. (in press). The effects of concurrent task interference on category learning. *Psychonomic Bulletin & Review.*

Warrington, E. K., & Weiskrantz, L. (1970). Amnesic syndrome: Consolidation or retrieval? *Nature, 228*, 628-630.

Wickens, J. (1993). *A theory of the striatum*. New York: Pergamon Press.

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.

Wilson, C. J. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia* (pp. 29-50). Cambridge: Bradford.

Winocur, G. & Eskes, G. (1998). Prefrontal cortex and caudate nucleus in conditional associative learning: Dissociated effects of selective brain lesions in rats. *Behavioral Neuroscience, 112*, 89-101.

Yellott, J. I. (1969). Probability learning with noncontingent success. *Journal of Mathematical Psychology, 6*, 541-575.

Yellott, J. I. (1971). Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology, 8*, 159-199.

Zola-Morgan, S., Squire, L. R., & Mishkin, M. (1982). The neuroanatomy of amnesia: Amygdala-hippocampus versus temporal stem. *Science, 218*, 1337-1339.