5.1 DISCRIMINATION, RECOGNITION, and CLASSIFICATION

To appear in:

Handbook of Human Memory: Foundations and Applications

Volume I: Foundations

M.J. Kahana & A.D. Wagner, Eds.

Michael L. Mack

University of Toronto

Thomas J. Palmeri

Vanderbilt University

June 19, 2020

ABSTRACT

Retrieval processes provide ways of using memory. We describe three important ways of using memory, with a focus on using memory to make decisions about visual objects.[1] *Discrimination* is deciding if an object is distinct from another object experienced at the same time or moments ago, *recognition* is deciding if an object is the same as an object experienced some time in the past, and *classification* is deciding what kind of object something is. All three require comparing the representation of a currently perceived object with representations retrieved from memory to drive a decision process. We discuss the component mechanisms for discrimination, recognition, and classification as formally instantiated in computational models and discuss relationships between model mechanisms and brain mechanisms as revealed by neuropsychology and brain imaging.

KEYWORDS

dimensions, features, similarity, exemplars, prototypes, decision making, evidence accumulation models, computational modeling, dissociations, fMRI

[1] While we focus on memory retrieval processes given visual objects, much of what we discuss can likely be generalized, at least to some degree, to visual patterns and visual scenes, to other sensory modalities, like audition and somatosensation, and to certain multisensory percepts; we cannot say whether any of what we discuss generalizes to gustation or olfaction.

1. INTRODUCTION

Imagine your morning routine ends with a walk through the park on your way to the lecture hall for class. One morning, as you reach the park's central esplanade, a happy dog barrels toward you stopping just short of crossing your path. You can *discriminate* one dog from another dog: Is that the same dog that moments ago ran around a park bench to reach you? You can *recognize* the dog: Is that the dog you saw in this same park last week? You can *classify* the dog: You know it is a dog. You also know it is an animal. Is this dog, in particular, a Border Collie or another kind of herding dog?

These thoughts are examples of common types of memory decisions we make about our everyday experiences (Figure 5.1.1). They all involve some kind of comparison of a current perceived experience – in this chapter we focus on visual experiences with objects – with some form of representation of past experiences retrieved from memory that provide evidence to drive a decision process. *Discrimination*[2] is deciding whether the current object is distinct from an object experienced moments ago. *Recognition*[3] is deciding whether the current object is the same

[2] Here we focus on deciding whether one object and another object are the same or different (or more generally, whether one stimulus and another stimulus are the same or different), discriminating one object from another, sometimes classically characterized as an AX discrimination task. An ABX discrimination (or AXB discrimination) task presents three objects and asks the observer if X is the same as A or the same as B; even more complex variants of discrimination tasks have been used. Whereas discrimination involves deciding whether two objects are the same or different, "detection" involves deciding whether an object is present or absent (in a background of internal or external noise). Relations between discrimination and other concepts in psychology (e.g., classical and operant conditioning) are beyond the scope of this chapter and this volume.

[3] "Recognition" typically has a specific meaning in the memory literature as deciding whether a current experience (here, an object) is *old* (seen before) or *new* (not). In the vision science literature, by contrast, "object recognition" often refers broadly to the array of processes involved in creating a visual representation of an object (Palmeri & Gauthier, 2004; Palmeri & Tarr, 2008). Here we use "recognition" exclusively in the sense of memory and refer to perceptual processes involved in creating a representation of an object using other terms.

as one experienced some time in the past. *Classification* is deciding what kind of object something is. Kinds can range in abstraction from the highly specific identification of a particular individual (*my dog Max*), to a subordinate type (*Yellow Labrador Retriever)*, to a basic kind (a dog), to a superordinate class (*an animal*, *a living thing*).

On the surface, these seem like very different memory decisions dependent on very different retrieval processes. Discrimination retrieves a specific immediate past experience whereas recognition can involve a sense of familiarity with a host of past experiences.[4] Discrimination involves a specific individual whereas classification involves generalization over a broad category. Recognition is memory based on episodic experiences whereas classification reflects semantic knowledge about a category. Indeed, as we review in the following sections, certain theories of discrimination, recognition, and classification have viewed these as distinct memory processes relying on distinct forms of memory representation. But when performance in these memory tasks is formalized in computational models that instantiate memory representations and mechanistic processes in mathematics and simulation, we can see many more similarities in representations and processes than might be apparent otherwise.

Discrimination, recognition, and classification are all broad topics for discussion. Other chapters in this volume discuss recognition memory (e.g., Dennis & Osth, Chapter 5.2; Yonelinas, Chapter 5.6) and entire volumes have been devoted to classification and concepts (e.g., Murphy, 2004) and formal models of classification (e.g., Pothos & Wills, 2011). We focus on formal models to illustrate the representations and component processes involved in discrimination, recognition, and classification in precise mechanistic detail, to describe ways in

[4] Of course, recognition can involve the recollection of a specific past experience as well as an overall sense of familiarity.

which these might share aspects of representation and process, and to highlight the inferential power of computational modeling approaches (e.g., Farrell & Lewandowsky, 2018; Hintzman, 1990). We highlight specific examples of neural evidence using a model-based cognitive neuroscience approach (e.g., Forstmann & Wagenmakers, 2015; Palmeri, Love, & Turner, 2017; Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017), with other chapters in this volume discussing the neural basis of memory more generally (e.g., Davachi, Chapter 1.3; Helfrich, Knight & D'Esposito, Chapter 1.5; Montaldi, Chapter 5.8).

2. MODELING DISCRIMINATION, RECOGNITION, AND CLASSIFICATION

2.1 Components of Models

At a minimum, any model of discrimination, recognition, or classification requires at least three key components to be specified: the perceptual representation of an object[5], the memory representations that this perceptual representation is to be compared to, and the decision process that uses the evidence from the comparison to determine same versus different (discrimination), old versus new (recognition), one category versus other categories (classification).

2.1.1 Perceptual Representations. Many models assume that objects are represented as multidimensional arrays of discrete features or continuous dimensions (Figure 5.1.2). For example, the dimensions of an object representation might correspond to its shape (square vs. triangle), color (black vs. white), and size (large vs. small) (e.g., Nosofsky, Gluck, Palmeri,

[5] Recall that our focus is on discrimination, recognition, and classification of *visual objects*.

McKinley, & Glauthier, 1994); so [1 2 2] would then correspond to the perceptual representation of a small white square.

In a model, the values along particular dimensions for particular objects might simply correspond to the direct physical manifestation of objects realized in an experiment or by previous psychophysical studies that carefully map physical properties of objects onto psychological representations.

Alternatively, the values along dimensions for particular objects could be determined by techniques like multidimensionsional scaling (MDS) (e.g., Shepard, 1980; Nosofsky, 1992a), where a matrix of similarity ratings between all possible pairs of objects are obtained and are used to construct a multidimensional psychological space, with individual objects represented as points in that space, with objects positioned within that space so that distances between objects are proportional to their judged dissimilarity. MDS can confirm that physical manipulations of dimensions correspond to expected psychological dimensions of simple objects (e.g., Nosofsky, 1991; Nosofsky & Palmeri, 1997) or can be used to reveal multidimensional representations for more complex objects (e.g., Nosofsky, Sanders, & McDaniel, 2018; Palmeri & Nosofsky, 2001).

Multidimensional arrays can also be simulated to capture the statistical similarity structure of objects used in an experiment without any particular simulated array meant to correspond to any specific item from an experiment. For example, multidimensional arrays simulating individual objects could be samples from a multivariate normal distribution (e.g., Ross, Deroche, & Palmeri, 2014). Multidimensional random samples could be drawn in such a way to reflect statistically experimental factors like between-item similarity, within- and between-category similarity (e.g., Hintzman, 1986; Nosofsky, 1988), and the relative frequency of individual features (e.g., Shiffrin & Steyvers, 1997). In a simulation, while a model may not

predict at the level of individual items, it may predict at the level of factors like relative item similarity, category typicality, category similarity, and frequency that are reflected in the sampled multidimensional item arrays.

All of the above simulate perceptual dimensions without instantiating a model of perception itself. While people in an experiment view images of objects, the models view multidimensional arrays, not images. Researchers have also explored using models of object recognition and computer vision as a perceptual front end to create within the simulation perceptual representations directly from the same images used in experiments with human participants (e.g., Annis, Gauthier, & Palmeri, 2020; Mack & Palmeri, 2010; Ross et al., 2014; Sanders & Nosofsky, 2018).

2.1.2 Memory Representations. Certainly the simplest model assumption is that an experienced perceptual representation gets stored as a complete memory representation. While many failures of memory can be attributed to failures of retrieval, many models also assume that storage is imperfect and that there is some parameterized probability that features in the perceptual representation are encoded as part of a memory representation (e.g., Hintzman, 1988; Shiffrin & Steyver, 1997). Models can also assume that memory representations decay in strength over time (e.g., Estes, 1994; Nosofsky & Palmeri, 1997) and that the quality of memory representations can vary across individuals due to brain damage on one extreme (e.g., Nosofsky & Zaki, 1998) to cases of domain expertise on the other extreme (e.g., Annis & Palmeri, 2019).

While memory for individual episodic experiences is commonly modeled as the storage of individual episodic representational arrays in memory (e.g., Hintzman, 1988; Shiffrin & Steyver, 1997), models vary in how knowledge about categories and other forms of semantic

knowledge are represented in memory. Models that are *instance-based* or *exemplar-based* assume that abstract knowledge for purposes of classification emerges from retrieval of the very same episodic experiences that underlie recognition memory (e.g., Hintzman, 1986, 1988; Jacoby & Brooks, 1984; Logan, 1988; Nosofsky, 1988). Other models assume that knowledge about a category is inherently abstract and that memory representations responsible for classification involve the abstraction of a prototype (e.g., Posner & Keele, 1968; Smith & Minda, 2002), that classification involves learning simple rules combined with memory for exceptions to those rules (e.g., Nosofsky, Palmeri, & McKinley, 1994; Palmeri & Nosofsky, 1995; Sakamoto & Love, 2004), or that semantic knowledge is stored in more complete representational arrays in memory from those that encode incomplete episodic experiences (e.g., Shiffrin & Steyvers, 1997).

Finally, especially in the case of models of recognition memory, it is common to assume that a representational array, whether for a probe item or a stored memory, includes not only the perceptual features of the object but also features associated with the context in which an object was experienced (e.g., Hintzman, 1988; Murdock, 1982; Shiffrin & Steyvers, 1997; see also Dennis & Osth, Chapter 5.2; Manning, Chapter 5.12).

2.1.3 Similarity. To discriminate, recognize, or classify an object, its perceptual representation needs to be matched with retrieved memory representations. Common to many models of discrimination, recognition, and classification is that this match is based on similarity.[6]

[6] There are models and theories of classification that are not based on similarity, or at least not entirely, and are dependent on more abstract knowledge representations like rules (e.g., Nosofsky, Palmeri, & McKinley, 1994), causal relations (e.g., Rehder, 2003), or theories (e.g., Murphy, 2004; Murphy & Medin, 1985). We focus on similarity-based models of classification

If objects are represented as multidimensional arrays of discrete features or continuous dimensions, there are many natural ways to formalize mathematically the similarity between two representational arrays. If representational arrays are vectors of unit length (or if the length is proportional to the strength rather than the content of the representation), then cosine similarity, which is the dot product of the two vectors normalized by their length, is a possible measure. If the vectors contain discrete features, then the contrast model (Tversky, 1977), which is a weighted sum of both common and distinct features, is a possible measure (see also Nosofsky, 1991). A match can also be based on likelihoods as part of a Bayesian decision process (e.g., Anderson, 1990; Nosofsky, 1990; Shiffrin & Steyvers, 1997).

If objects are represented as points in a multidimensional psychological space, then a simple, and arguably most common, assumption (Figure 5.1.2) is that similarity, s_{ij}, between two objects, i and j, is a decreasing function of distance, d_{ij},

$$d_{ij} = \left(\sum_{m=1}^{M} w_m |i_m - j_m|^r \right)^{1/r}$$

where M is the number of dimensions in the representational array, i_m is the value of object i along dimension m, and r reflects the distance metric. The familiar Euclidean distance metric results when r=2, and a city-block distance metric results when r=1; it is common to assume the Euclidean metric for integral dimensions and city-block for separable dimensions (Garner, 1974; Shepard, 1987).[7] Dimensions are weighted, w_m, according to their relative diagnosticity for the

to highlight the relationships between these models and models of discrimination and recognition.

[7] As the name suggests, integral dimensions are often said to be perceived unitarily, as part of an integrated whole, such as the hue, saturation, and brightness of colors. Separable dimensions, by contrast, are often said to be perceived, attended, and processed independently, such as the shape and color of objects.

current task demands (Nosofsky, 1984; see also Carroll & Wish, 1974; Kruschke, 1992; Lambert, 2000); the importance of these weights will be highlighted later.

Similarity is a decreasing function of distance, with the most common general form being

$$s_{ij} = exp\left(d_{ij}^{q}\right)$$

where q=1 gives an exponential function and q=2 a Gaussian function. Shepard (1987) described the exponential (q=1) as a law of generalization based both on empirical data and Bayesian principles; cases where a Gaussian (q=2) provide a better account than an exponential (e.g., Nosofsky, 1986) may reflect the presence of sensory-perceptual noise (e.g., Kahana & Sekuler, 2002) when stimulus differences approach just-noticeable differences (e.g., Ennis, 1988).

2.1.4 Evidence for Discrimination, Recognition, and Classification. For a given task, we can define the evidence (E) in favor of one decision over other decisions as a function of the similarity between the representation of the probe object p and representations in memory.

For discrimination, we are asking if probe p is the same as item j that was just experienced. So the evidence for a "same" response ($E_{same|p}$) can be defined most simply[8] as

$$E_{same|p} = s_{pj}$$

For recognition memory, it is common to assume that the evidence for an "old" response ($E_{old|p}$) is based on the overall familiarity of probe p, formalized mathematically in its simplest form as the summed similarity between p and all items in memory (with a match of context cues

[8] Most accounts of discrimination bypass the important question of how the right memory representation of the first object is retrieved so it can be compared with the second object, or there is the tacit assumption that the context is so overwhelming that memories from objects on other discrimination trials do not intrude (but see Cohen & Nosofsky, 2000).

used to limit the relevant stored memories to a particular list or other such experimental context defining the appropriate recognition decision)

$$E_{old|p} = \sum_{k} s_{pk}$$

where *k* indexes all instances in memory. Evidence of this sort is an example of a *global matching* model of recognition (e.g., Dennis & Osth, Chapter 5.2; Gillund & Shiffrin, 1984; Hintzman, 1988). Such familiarity-based matching process is often, but not always, one of the components of dual-process models of recognition memory (e.g., Yonelinas, Chapter 5.6).

For classification, evidence that probe *p* belongs to Category *A* ($E_{A|p}$) depends on the similarity between probe *p* and the memory representation of Category *A*.[9] If knowledge about a category is represented in terms of stored exemplars, we can compute the evidence based on the nearest neighbor to probe *p* (most similar stored exemplar of Category *A* to probe *p*), average similarity of probe *p* to stored exemplars of Category *A* (e.g., Reed, 1972), or summed similarity of probe *p* to stored exemplars of Category *A*. Summed similarity is the most common (and most successful) variant assumed by the exemplar-based *Generalized Context Model* (GCM) (Nosofsky, 1984, 1986; see also Medin & Schaffer, 1978)

$$E_{A|p} = \sum_{k \in A} s_{pk}$$

where *k* indexes all instances of Category *A* in memory. Alternatively, if knowledge about a category is represented in terms of an abstracted prototype, then the evidence that probe *p* is a member of Category *A* is simply given by the similarity between probe *p* and the stored prototype for Category *A*, P_A.

[9] Again, we are focusing on similarity-based models of classification here. Classification based on abstract rules or knowledge would involve very different mechanisms.

2.1.5 Decision Rules. The evidence is used to make a decision, responding "same" or "different" in a discrimination task, "old" or "new" in a recognition task, "Category A" or "Category B" in a classification task.

In the case of discrimination and recognition, a simple deterministic decision rule would be to respond "same" (or "old") if the evidence (E_{same} or E_{old}) is greater than some criterion (k_{same} or k_{old})

$$E > k$$

and respond "different" (or "new") otherwise. Of course, the criterion (or the evidence) could be noisy, in which case the decision rule would be

$$E > k + \varepsilon$$

where ε is a sample from a normal distribution with mean zero and standard deviation σ. Decisions about discrimination and recognition are based on whether the match (discrimination) or familiarity (recognition) are sufficiently strong (relative to some criterion).

In the case of classification, a relative decision rule is needed instead. An object p is classified as a member of Category A if its summed similarity to exemplars of Category A (E_A) is greater than its summed similarity to exemplars of Category B (E_B), again allowing for there to be noise in the decision rule (or evidence)

$$E_A > E_B + \varepsilon$$

There can also be response bias (β) (see Nosofsky, 1991) irrespective of the relative evidence

$$E_A > E_B + \beta + \varepsilon$$

In the extreme, as β grows larger and larger, the likelihood of ever responding Category A grows smaller and smaller. Of course, discrimination and recognition decisions can also be

biased, but, mathematically, adding a bias term would be non-identifiable (any bias parameter would be additive with the criterion parameter).

In the case of deterministic decision rules, the probability (or frequency) of responding ("same", "old", "Category A") would need to be determined by Monte Carlo simulation (or by a mathematical derivation of the long-run probability given the specified deterministic decision rule).

Models can also be specified using a probabilistic decision rule. For discrimination, the probability of deciding "same" given a probe p is

$$P_{same}(p) = \frac{E_{same|p}}{E_{same|p} + k_{same}}$$

For recognition, the probability of deciding "old" given a probe p is

$$P_{old}(p) = \frac{E_{old|p}}{E_{old|p} + k_{old}}$$

For classification, the probability of deciding "Category A" given a probe p is

$$P_A(p) = \frac{\beta_A E_{A|p}}{\beta_A E_{A|p} + \beta_B E_{B|p}}$$

Again, while the response biases (β_A and β_B) are explicit for classification models (because of the relative decision rule), they are implicit (not uniquely identifiable) within the values of the criteria for discrimination and recognition models.

And whereas discrimination and recognition are inherently two-choice decisions, classification can be multi-choice. In that case, the decision rule for classification can be extended to

$$P(p) = \frac{\beta_A E_{A|p}}{\sum_{K \in R} \beta_K E_{K|p}}$$

where K is the set of possible categories under consideration (see also Palmeri, 1997).

A range of decision rules from probabilistic to deterministic can be formalized by raising the evidence to a power (Ashby & Maddox, 1993)

$$P(p) = \frac{(E_{A|p})^{\gamma}}{(E_{A|p})^{\gamma} + (E_{B|p})^{\gamma}}$$

(leaving out the bias terms for simplicity) with $\gamma = 1$ giving the probabilistic decision rule and $\gamma \to \infty$ giving a deterministic decision rule without noise. Ashby and Maddox (1993) demonstrated formally the mathematical relationships between the probabilistic decision rule and deterministic decision rule.[10] While there are cases where humans (and animals) engage in probability matching in ways that suggest a purely probabilistic decision rule, in most cases of object classification, humans (and animals) tend to respond more deterministically than predicted by a purely probabilistic decision rule.

These deterministic and probabilistic decision rules only predict response probabilities. To extend these frameworks to response times as well as response probabilities, the evidence is used to drive an accumulation process (see Heathcote, Trueblood, & Starns, Chapter 5.10). The first such model of classification was the *Exemplar-based Random Walk* (EBRW) model (Figure 5.1.3; Nosofsky & Palmeri, 1997, 2015; Palmeri, 1997), which combined elements of the exemplar-based GCM model of classification (Nosofsky 1984, 1986), the instance theory of automaticity (Logan, 1988), and a random walk model of decision making (e.g., Busemeyer, 1985; Link, 1992; Ratcliff, 1978).

[10] While theoretically the form of a category representation (exemplar vs. prototype) and the nature of the decision rule (deterministic vs. probabilistic) are conceptually independent, it has been shown that considering both factors is critical for contrasting alternative models. For example, Nosofsky and Zaki (2002) showed that certain instantiations of a prototype model are sufficiently flexible to allow for a range of probabilistic vs. deterministic decision rules to be embedded within them non-identifiably whereas an exemplar model must have the probabilistic vs. deterministic nature of the decision rule expressed explicitly.

Time enters into EBRW in two ways. First, following instance theory, repetitions of the same item are assumed to create new memory representations of that item. When a probe *p* is presented, instances race to be retrieved from memory. Based on instance theory, more repetitions of the same item in memory mean that the winner of the race for retrieval will get faster with more experience. This property allows instance theory, and hence EBRW, to predict speed-ups with experience (see also Palmeri & Cottrell, 2009; Palmeri, Wong, & Gauthier, 2004). One thing that distinguishes EBRW from instance theory is that memory retrieval rates depend on the similarity between probe *p* and stored exemplar *j*. If retrieval times are exponentially distributed, then it can be shown that the probabilities of retrieving exemplar *j* (relative to all stored exemplars of Category *A* and Category *B*) is given by

$$P(retrieve\ j) = \frac{s_{pj}}{\sum_{k \in A} s_{pk} + \sum_{k \in B} s_{pk}}$$

and the probability of retrieving any exemplar from category *A* is simply

$$P(retrieve\ A) = \frac{\sum_{k \in A} s_{pk}}{\sum_{k \in A} s_{pk} + \sum_{k \in B} s_{pk}}$$

which is the GCM.

Rather than base a decision on a single retrieval (like GCM), EBRW assumes that each retrieval drives a random walk decision process.[11] Like other evidence accumulation models (Heathcote et al., Chapter 5.10), EBRW assumes that evidence is accumulated towards an upper boundary associated with a Category *A* decision and a lower boundary associated with a Category B decision (see Palmeri, 1997, for a generalization to multiple categories). This accumulation over time is the second way time enters the EBRW. If a Category *A* exemplar wins

[11] As the time for each step and the size of each step go to zero (appropriately) in the limit, a discrete random walk approaches a continuous diffusion process (Feller, 1968).

the retrieval race, the random walk takes a step towards the *A* threshold (θ_A), if a Category *B* exemplar wins, it takes a step towards the *B* threshold (θ_B). Which threshold is hit first determines both what response is made (*A* or *B*) and when it is made (like other evidence accumulation models, EBRW also assumes non-decision time associated with perceptual processing and motor execution).

Interestingly, if the response thresholds are set equal to one another ($\theta_A = \theta_B = \theta_*$), it can be shown (Nosofsky & Palmeri, 1997) that predicted response probabilities from EBRW are

$$P(A|p) = \frac{(\sum_{k\in A} s_{pk})^{\theta_*}}{(\sum_{k\in A} s_{pk})^{\theta_*} + (\sum_{k\in B} s_{pk})^{\theta_*}}$$

which is identical to the γ decision rule outlined earlier.

While the EBRW was originally developed to account for response probabilities and response times in classification tasks, its principles have been extended to account for a range of speeded discrimination and recognition paradigms as well (e.g., Annis et al., 2020; Annis & Palmeri, 2019; Mack & Palmeri, 2010; Nosofsky, Cao, Cox, & Shiffrin, 2014; Nosofsky & Palmeri, 2015). Further extensions of these models consider the dynamics of how object features are sampled and how these representational dynamics influence the time-course of discrimination, recognition, and classification (e.g., Cohen & Nosofsky, 2003; Cox & Criss, 2019; Lamberts, 2000).

2.2 Relations Between Discrimination, Recognition, and Classification

Having outlined a formal mathematical framework for modeling discrimination, recognition, and classification in the same language, we now highlight a few examples of work examining relationships between these memory retrieval processes. As noted elsewhere, the

literature on each of these processes is vast, so the following can only hope to be illustrative rather than exhaustive.

2.2.1 Relations Between Identification and Classification. Classification involves knowledge in memory generalized over a broad category of objects, for example deciding that some object is a kind of thing called a "dog" not a "cat". Identification involves knowledge in memory about a specific individual object, for example deciding that some object is an individual named "Brownie" not "Max" or "Chelsea". Identification is like discrimination (why we discuss identification here) in that they both involve a specific individual whereas classification involves generalization over a class.

An early theory of identification bears important similarities to the class of theories described earlier. The *Similarity-Choice Model* (SCM) of identification (Luce, 1963; Shepard, 1957) assumes that identification confusions, the probability of identifying stimulus *i* with the label associated with stimulus *j*, is given by

$$P(S_i) = \frac{\beta_j s_{ij}}{\sum_k \beta_k s_{ik}}$$

The evidence that stimulus *i* should be given the label associated with stimulus *j* is given by the similarity between *i* and *j*, just like the evidence for a same-different discrimination described earlier; the primary difference is that in identification, relative evidence is required to decide on a particular label to apply to an object, whereas, in discrimination, absolute evidence is required to decide if an object is the same or different as a previous object.

One early hypothesis assumed that similarity between objects was invariant over task demands. In that case, the object similarities that would govern identification of a unique individual should also govern classification as a member of a broad category. In its simplest

form, the identification confusions of objects within a category should simply sum together to predict the probability that an object is classified as a member of that category. But this *mapping hypothesis* (Nosofsky, 1986) between identification and classification fails (e.g., Shepard, Hovland, & Jenkins, 1961; see also Nosofsky, 1984). Classification performance cannot be predicted from identification performance if object similarities are assumed to be invariant across tasks. Indeed, Shepard et al. (1961) suggested from these failures that classification must involve some form of abstract knowledge learning, such as rule formation (see also Nosofsky et al., 1994), whereas identification involves some form of stimulus-response association learning.

Nosofsky (1984, 1986) instead allowed for similarities to vary across tasks in a principled way. While the locations of objects in multidimensional space would remain invariant across tasks, the weights applied to psychological dimensions, the w_m terms included in the distance metric described earlier, could vary across tasks (see also Kruschke, 1992). In particular, dimensions that were diagnostic for classification (or identification or recognition) would receive large weights while dimensions non-diagnostic for the task would receive small weights, thereby stretching the psychological space along diagnostic dimensions and shrinking the space along non-diagnostic dimensions. When attention weights are added to the modeling framework, classification can indeed be predicted from identification (Nosofsky, 1984, 1986, 1987). The same object representations, memory representations, and kinds of processes can govern identification and classification.

2.2.2 Relations Between Recognition and Classification. Many classic theories argued that recognition and classification depend on distinct memory systems. Recognition is *episodic* while classification is *semantic* (Tulving, 1972, 2002). Recognition is *declarative* while classification

is *non-declarative* (Squire & Zola, 1996). Some of these claims are based on neuropsychological and (more recently) brain imaging results (but see Section 3 of this Chapter), but certain empirical results also suggested dissociations between recognition and classification.

In experiments where participants learn novel objects belonging to novel categories and are tested both on their classification of old and new objects and their recognition memory for old and new objects, there is often little correlation between classification confidence and recognition memory performance (e.g., Anderson, Kline, & Beasley, 1979; Hayes-Roth & Hayes-Roth, 1977; Metcalfe & Fisher, 1986). Intuitively, it seems reasonable to presume that if memory for specific exemplars drove both classification and recognition that objects classified with high confidence should also be well recognized. A failure to find a correlation between classification and recognition seems problematic for a single-system exemplar model.

Yet exemplar models account quite naturally for these observed cases of a lack of correlation (e.g., Nosofsky, 1988, 1991, 1992b). Even though according to models like the GCM, both classification and recognition are assumed to be based on similarity to the same stored exemplars in memory, classification depends on the relative summed similarity to exemplars in one category versus another category whereas recognition depends the absolute summed similarity to all exemplars in memory. An object could be similar to many stored exemplars, causing it to be recognized as "old", but those similar exemplars could belong to different categories, making classification confidence low; alternatively, an object could be dissimilar to most stored exemplars, causing it to be recognized as "new", but it could be similar to only exemplars in one category, making classification confidence quite high.

2.2.3 Effects of Learning and Expertise on Discrimination, Recognition, and Classification. Learning, experience, and expertise affect perceptual representations that aid discrimination, enhance perceptual and memory representations that aid recognition, and build representations that enable increases in performance at classification at various levels of abstraction (e.g., Gauthier, Tarr, & Bub, 2009; Palmeri et al., 2004).

Early in learning to classify objects, people may be given explicit rules (e.g., Palmeri & Nosofsky, 1995; Palmeri, 1997) or generate rules on their own (e.g., Ashby, Alfonso-Reese, & Waldron, 1998; Nosofsky et al., 2004). Imperfect category rules can be supplemented with exceptions stored in memory (e.g., Davis, Love, & Preston, 2012; Nosofsky et al., 2004; Nosofsky & Palmeri, 1998; Sakamoto & Love, 2004). According to instance theory (Logan, 1988) and EBRW (Nosfsky & Palmeri, 2015; Palmeri, 1997), performance in a task like classification involves a race between applying a categorization rule and retrieval of memories for past experiences performing the classification. Early in learning, the rule tends to dominate performance because it wins the race. Assuming that memory storage is obligatory (Logan, 1988) and that additional stored memories speed up the memory retrieval process, later performance, after some experience, is based on memory retrieval (Palmeri, 1997; see also Johansen & Palmeri, 2002), leading to automaticity and expertise (Palmeri et al. 2004; see also Healy, Proctor, & Kole, Chapter 11.4).

Often accompanying classification learning of new object categories are changes to the perceptual representations of objects that belong to those learned categories as revealed by changes in discrimination performance. Beyond the kind of dimensional stretching of diagnostic psychological dimensions during object classification, as reflected by changes to the w_m parameter in the formula for the distance metric, there can often be longer-term changes in

perceptual discriminability along those diagnostic dimensions. As a consequence of learning novel categories, people show increased perceptual discrimination for basic form and color dimensions (e.g., Goldstone, 1994) as well as along more complex dimensions of complex objects (e.g., Folstein, Gauthier, & Palmeri, 2012; Goldstone & Steyvers, 2001) that are relevant to learned categories.

And finally, learning and expertise with object classification can affect recognition memory for objects in an expert domain. Real-world perceptual expertise in a domain, say for birds or cars or medical images, is often accompanied by increases in visual short-term memory (e.g., Curby, Glazek, & Gauthier, 2009) and long-term memory (e.g., Evans, Cohen, Tambouret, Horowitz, Kreindel, & Wolfe, 2011; Herzmann & Curran, 2011). Annis and Palmeri (2019) observed increases in short-term and long-term memory for birds as a function of perceptual expertise at classifying birds and via an extension of EBRW showed that these increases in memory performance were best explained by a combination of an increase in the quality of the perceptual representations of birds and the memory strength for bird memories as a function of real-world expertise (see also Healy, Proctor, & Kole, Chapter 11.4).

2.2.4 Other Theoretical Perspectives. In the above discussion, we have intentionally focused on one particular family tree of computational models (SCM, GCM, EBRW) because they allowed us to illustrate potential theoretical connections between discrimination, recognition, and classification. Of course, there are other theoretical perspectives in this active area of research. Some past debates, for example between exemplar models (Nosofsky & Smith, 1992) and decision boundary models, have given rise to more nuanced contrasts between exemplar (McKinley & Nosofsky, 1995) and exemplar-like models (e.g., Ashby & Waldron, 1999) that

make similar predictions about behavior but differ in their relations to neural processes (e.g., Ashby & Rosedahl, 2017). Object classification can be based on rules rather than similarity (e.g., Ashby et al., 1998; Erickson & Kruschke, 1998), perhaps supplemented by similarity to stored exemplars, especially those that represent exceptions to classification rules (e.g., Nosofsky et al., 1994; Palmeri & Nosofsky, 1995). Or can be based on representations that can approximate rules, prototypes, or exemplars in a more flexible manner (e.g., Love, Medin, & Gureckis, 2004), or can reflect a combination of rules and exemplars whose combination might vary as a function of experience and expertise (e.g., Johansen & Palmeri, 2002; Palmeri et al., 2004). To the extent that representations supporting classification are more abstract than experienced exemplars, then multiple memory systems could be implicated in recognition and classification.

3. NEURAL EVIDENCE

The modeling framework detailed above highlighted one particular theoretical view that there is a common mechanistic substrate for discrimination, classification, and recognition through flexible retrieval of memory representations. Early patient and neuroimaging work, some of which we note below, supported a different view that there are multiple distinct learning and memory systems supported by separable neural mechanisms. More recent work leveraging sophisticated model-based cognitive neuroscience approaches has demonstrated common neural substrates for component processes involved in discrimination, recognition, and classification. Rather than a comprehensive review of the rich literature on related neuroscience research, we highlight these studies that integrate computational and neuroscientific methods.

3.1 Neural Representations, Similarity, and Attention

The computational framework we describe is based on feature-based representations of experiences and similarity processes that compare current experiences to stored representations to make discrimination, recognition, and classification decisions. Neural evidence for these mechanisms has existed for as long as the field has studied the brain. For example, seminal neurophysiological studies in monkeys showed that firing rates for neurons in inferotemporal (IT) cortex exhibit sensitivity to features diagnostic for newly-learned classification (Sigala & Logothetis, 2002; Freedman et al., 2003) while retaining coding for item-specific features (Op de Beeck et al., 2001). Similarly, early studies leveraging functional magnetic resonance imaging (fMRI) in humans demonstrated that activation of brain regions along the ventral visual pathway discriminate between categories of visual content (Kanwisher et al., 1997; Epstein & Kanwisher, 1998; Gauthier et al., 1999) and do so according to a gradient of feature similarity (e.g., Gauthier et al., 1997b).

The advent of multivariate techniques that enable researchers to quantify patterns of neural activity across cell populations and multiple voxels provided the opportunity to more precisely assess the content of neural representations when making different types of decisions. For example, real-world categories of visual content are distinctly represented in activation patterns across ventral visual cortex (e.g., Haxby et al., 2001; Connolly et al., 2012; Kriegeskorte et al., 2008) and the medial temporal lobe (e.g. Liang et al., 2013; LaRocque et al., 2013). Work specifically targeting the contribution of visual and semantic features to neural representations for naturalistic and novel visual objects has converged on two central findings: 1) the more features two objects share, the more similar their neural representations, and 2) there exists a gradient of feature type along the ventral visual pathway with more visual features dominating

representations in occipital and posterior temporal regions and more conceptual features dominating medial and anterior temporal regions (Martin et al., 2018; Clarke & Tyler, 2014; Erez et al., 2016; Davis & Poldrack 2014). Interestingly, item- and category-specific patterns of activation elicited when viewing or encoding visual content are also present during memory retrieval (Chadwick et al., 2016; Kuhl, Rissman, & Wagner, 2012; Mack & Preston, 2016; Polyn, Natu, Cohen, & Norman, 2005; Staresina, Henson, Kriegeskorte, & Alink, 2012; Tompary, Duncan, & Davachi, 2016), a finding consistent with the proposal for a common neural substrate underlying classification and recognition.

Model-based cognitive neuroscience approaches (e.g., Forstmann & Wagenmakers, 2015; Palmeri, Love, & Turner, 2017; Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017), in which latent representations and processes from computational models are leveraged to interrogate brain function, have recently significantly advanced our understanding of the neural underpinnings of discrimination, recognition, and classification. One such study (Mack, Preston, & Love, 2013) looked to neural representations of fMRI activity present during classification as evidence to adjudicate between exemplar- and prototype-based models of classification. The logic followed that the evidence guiding classification decisions ($E_{A|p}$) would be evident in neural activation patterns. Given the differences in the nature of their underlying representations, exemplar and prototype models predict unique signatures of classification evidence. Thus, the better correspondence between neural activation and a model's latent quantity of classification evidence, the better that model formalizes the representations and processes of classification. This approach showed that in a classic classification task using what is known as the "5-4" category structure (Medin & Schaffer, 1978), almost all participants were better fit by a classification model using exemplar representations rather than prototype representations.

These findings were recently extended (Bowman & Zeithamova, 2018) with a classification task defined by category structures that encourage more abstract or prototype-like representation (Zeithamova, Maddox, & Schnyer, 2008), finding stronger evidence in neural activation for classification evidence as predicted by a prototype model. Interestingly, both studies demonstrated that attention to different stimulus features as predicted by the computational models was consistent with the similarity between neural representation in multiple brain regions associated with visual processing (lateral occipital and parietal cortices), memory (hippocampus and medial temporal lobe cortex), and decision making (ventromedial and rostrolateral prefrontal cortex) (see also, Davis, Goldwater, & Giron, 2017). In other words, the similarity of activation patterns for different objects during classification behavior matches the attention-weighted similarity computation formalized in computational models. These highlighted studies offer compelling neural evidence for the model mechanisms outlined in the prior section, specifically multidimensional representations based on feature combinations, feature-based attention that distorts and shapes representations (e.g., Goldstone & Styvers, 2001; Nosofsky, 1986), and similarity-based processes (Shepard, 1987). And, collectively, this work supports the notion that the nature of memory representations underlying any given type of discrimination, classification, or recognition decision are adaptive with regard to prior experience and current task goals (Anderson, 1991; Love & Gureckis, 2007; Love, Medin, & Gureckis, 2004; Mack, Love, & Preston, 2016). In fact, emerging evidence from neural firing and synchrony in monkey lateral prefrontal cortex suggests both specific exemplar-like and generalized prototype-like representations are available through separable neural circuits during classification decisions (Wutz, Loonis, Roy, Donoghue, & Miller, 2018).

The library of cognitive neuroscience research on discrimination, classification, and recognition is vast and we have highlighted only a fraction of relevant findings. By focusing on both seminal studies that first established how neural evidence supports flexible decisions based on memory representations and more recent studies that integrate computational modeling with neuroscience methods, we hope to provide the interested reader with helpful entry points into the rich literature.

3.2 A Common Neural Framework

In the following sections, we specifically highlight cognitive neuroscience studies that have assessed the central proposal outlined above; namely, that discrimination, recognition, and classification share a core set of computational mechanisms and representations.

3.2.1 Discrimination and Classification. A fundamental prediction of common mechanisms and representations for discrimination and classification is that memory representations and similarity computations that support successful classification should impact discrimination performance. In terms of the model framework proposed here, the same representations driving decision evidence to classify an object ($E_{A|p}$) are at play in discriminating between that object and one seen just before it ($E_{same|p}$). Put simply, learning that two objects belong to different categories should change the perception of these two objects such that they appear more different even in perceptual tasks not related to classification. This increase in discrimination due to classification learning is known as *acquired distinctiveness* and has been a guiding hypothesis in many behavioral (e.g., Goldstone, 1994; Goldstone & Styvers, 2001; Notman, Sowden, & Özgen, 2005; Op de Beeck, Wagemans, & Vogels, 2003) and neural (e.g., De Baene, Ons,

Wagemans, & Vogels, 2008; Folstein, Palmeri, & Gauthier, 2013; Jiang et al., 2007; Li, Ostwald, Giese, & Kourtzi, 2007; Sigala & Logothetis, 2002) studies.

Although initial attempts to find neural evidence of acquired distinctiveness resulted in mixed findings, one notable study by Folstein and colleagues (2013) provided a compelling demonstration of a neural link between classification and discrimination. Participants first learned to classify complex visual objects, cars, carefully constructed from a morph space of 3D car models into two categories (Figure 5.1.4). After successfully learning to classify the cars, participants performed a visual discrimination task during fMRI scanning in which they determined whether or not two sequentially-presented cars were positioned in the same location on the screen. Critically, car pairs were composed of the same car or two different cars that varied along a feature dimension that was either relevant or irrelevant for the classification task. This paradigm was designed to reveal fMRI adaptation (Grill-Spector, Henson, & Martin, 2006), an effect in which the second presentation of a visual stimulus results in lower activation since it is engaging the same population of neurons. The logic followed that if classification shapes neural representations such that features relevant for classification are selectively enhanced with richer representations, pairs of objects that differ along these classification-relevant features will exhibit less or no adaptation. In contrast, object pairs that differ along irrelevant features will have more overlap in neural representation resulting in greater adaptation effects. Indeed, this was exactly what was found: regions within the ventral visual stream including fusiform gyrus and occipital regions showed greater sensitivity to cars varying along classification-relevant features. This neural acquired distinctiveness was mirrored in behavior—participants were better able to discriminate small variations in cars that differed along relevant vs. irrelevant features. These findings demonstrate an important link between the neural substrate of discrimination and

classification. Learning to classify visual objects changes their neural representations and these learning-related changes impact the neural representations and behavior during subsequent discrimination tasks.

3.2.3 Classification and Recognition. Some of the earliest work on the neuroscience of classification and recognition was conducted with individuals with amnesia (Knowlton & Squire, 1993). This work demonstrated that in spite of catastrophic memory loss due to damage to key memory structures in the medial temporal lobe, individuals with amnesia were capable of learning new classification tasks. The compelling results stood in the face of the formal theoretical relationship between recognition and classification (e.g., Nosofsky, 1988; Love & Gureckis, 2007) by positing distinct neural mechanisms (Squire & Zola, 1996), a theoretical stance supported by fMRI evidence that distinct brain regions are recruited in service of the two tasks (Reber, Gitelman, Parrish, & Marsel Mesulam, 2003). However, it has been demonstrated that global similarity models that leverage common representations and computations for classification and recognition can successfully account for patient data in both types of tasks (e.g., Love & Gureckis, 2007; Nosofsky & Zaki, 1998; Palmeri & Flanery, 2002). Moreover, neuroimaging research from the past decade suggests that in addition to its role in memory, the MTL is in fact a critical player in classification (Bowman & Zeithamova, 2018; Davis, Love, & Preston, 2012a, 2012b; Mack et al., 2016; Mack, Love, & Preston, 2018; Poldrack et al., 2001; Seger, Braunlich, Wehe, & Liu, 2015; Zeithamova et al., 2008). Here, we highlight two studies that target a common mechanistic account of classification and recognition through an integrated approach combining computation models with neuroimaging.

As we detail above in Section 2, decisions of classification and recognition rely on the same underlying evidence (i.e., summed similarity of the current stimulus to stored memory representations) compared to a task-specific criterion. Notably, this criterion is expected to qualitatively vary between classification and recognition (Nosofsky et al., 2012). For example, to make successful recognition decisions while minimizing false alarms to stimuli that are related but nonetheless novel, the current stimulus must match a specific prior experience to a high degree, thus necessitating a high criterion. In contrast, a classification decision relies on the relatively lax criterion of sufficient similarity to prior experiences with no need for exact matches. It follows that such differences in criterion may lead to different activation profiles across brain regions, in spite of the same computations and representations guiding decisions in the two tasks. Nosofsky, Little, and James (2012) tested this hypothesis by manipulating the criterion used by participants for recognition and classification during fMRI scanning. After matching criterion settings across the tasks, they found no discernable differences in brain activation. Moreover, an exemplar-based model of the two tasks accounted for both classification and recognition behavior. Finally, model-based predictions of criterion settings across the tasks corresponded with individual differences in BOLD activation in both the frontal eye field and anterior insula, two brain regions implicated in perceptual decision-making tasks. Thus, these findings support a common computational mechanism for classification and recognition with neural engagement that varies according to task demands.

A complementary study by Davis et al. (2014) examined a common role for similarity-based comparisons to memory representations in recognition and classification. Specifically, they hypothesized that if the same representations and similarity-based computations underlie decision evidence for classification and recognition, similarity between neural representations

elicited during these two tasks should predict recognition and classification behavior. By analyzing activation patterns in the medial temporal lobe during classification and recognition, it was demonstrated that global neural pattern similarity (i.e., the summed similarity between the activation pattern for the current stimulus and all other stimuli) predicted both confidence in recognition memory decisions (e.g., $E_{old|p}$) and a latent model measure of classification evidence (e.g., $E_{A|p}$). These findings suggest that not only is MTL function important for classification and recognition, MTL-based representations operate according to the formal principles of the computational framework that accounts for both classification and recognition.

4. FINAL COMMENTS

Discrimination, recognition, and classification span a wide range of behaviors with seemingly distinct psychological and neural mechanisms. However, converging behavioral and neural evidence motivated by a rich theoretical history, the highlights of which were covered in this chapter, suggests that these different types of decisions may arise from common representations and processes, or at least that a common set of computational principles may govern forms of representations and kinds of processes, even if those representations and processes may be distributed across different brain areas (Bowman & Zeithamova, 2018; Zeithamova et al., 2019). The central component of this integrative framework is the flexible retrieval of memory representations. Whether deciding if the object in front of you is the same as what you experienced just moments ago, is a specific kind of animal, or is familiar based on a host of past experiences, the same memory representations serve as evidence for your final decision. Looking forward, we expect that future research will elaborate on this common framework by targeting how factors of experience impact decisions across these different types

of flexible retrieval. For example, how does perceptual expertise that relies on extensive semantic knowledge (e.g., expert bird watchers) impact memory for experiences related and unrelated to that expertise (e.g., Annis & Palmeri, 2019)? And, what are the neural mechanisms that support flexible encoding and retrieval of memory representations across changing task goals (e.g., Bowman & Zeithamova, 2018; Davis et al., 2017; Mack et al., 2016; Mack et al., 2020).

Finally, it is worth noting the value of computational modeling in arriving at the field's current understanding of discrimination, recognition, and classification. Only by formalizing the computations and representations implicit in verbal descriptions of these different decisions has it been possible to disconfirm some theories and strongly support others. Moreover, the recent boom in model-based and computationally-sophisticated neural approaches to investigating discrimination, recognition, and classification has significantly broadened the theoretical scope and impact of findings. Such an approach offers the unique opportunity to localize quantitatively-defined model representations and computations to specific brain regions (e.g., Davis et al., 2012a; Kragel et al., 2015; Mack et al., 2016) and to adjudicate among competing formal theories with brain measures (e.g., Davis et al., 2012b; Mack et al., 2013; Bowman & Zeithamova, 2018). Indeed, it is compelling and persuasive evidence to demonstrate that the neural representations of stimuli match one model's representational predictions more than another (e.g. Mack et al., 2013) or that trial-by-trial fluctuations in neural signal from a specific brain region correspond with the dynamics of a latent model component (e.g., Davis et al., 2012a). A promising avenue for future work will be a focus on how individuals or patients differ in terms of computational model predictions (e.g., differences in feature dimension attention

weights) and how these model-based quantities are linked to individual differences in neural function and representation.

This integrative approach also promises to motivate the development of more comprehensive mechanistic theories that bridge levels of analysis to account for both behavior and brain function. For example, one exciting new direction in the field (Gureckis & Love, 2007; Mack et al., 2018; Zeithamova et al., 2020) is a theoretical push to marry the computational framework we have highlighted here with the brain-based accounts of episodic memory such as the influential Complementary Learning Systems theory (McClelland et al., 1995; Norman & O'Reilly, 2003; Schapiro et al., 2017). This work extends the perspective of common functions and representations underlying classification and recognition behavior to formalize how functions of memory encoding (e.g., pattern separation in the hippocampus) may form distinct exemplar-like representations of our experiences. Such representations can then be flexibly retrieved for a variety of decisions through selective attention and goal-based strategies ascribed to cortex (e.g., Hutchinson et al., 2014; Mack et al., 2020) and potentially leveraged to support the formation of representations in other brain regions.

Looking forward, key questions remain: How is selective attention tuned to diagnostic feature dimensions and how are these weights applied across changing task demands? How does learning shape the nature of representations necessary for classification and what impact does this have on recognition? How does recent experience interact with prior knowledge in making flexible discrimination, classification, and recognition decisions? Future research that integrates computational theory with brain measures will undoubtedly answer these questions to shed light on the cognitive and neural dynamics underlying how we perceive, learn, and remember.

ACKNOWLEDGEMENTS

This work was supported in part by NSF grant SMA 1640681 (TJP) and NSERC Discovery Grant RGPIN-2017-06753 (MLM).

REFERENCES

Anderson, J.R. (1990). *The Adaptive Character of Thought*. Erlbaum.

Anderson, J.R., Kline, P.J., & Beasley, C.M. (1979). A general learning theory and its application to schema abstraction. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation*. New York: Academic Press.

Annis, J., Gauthier, I., & Palmeri, T.J. (2020). Combining convolutional neural networks and cognitive models to predict novel object recognition in humans. *Manuscript under revision*.

Annis, J., & Palmeri, T.J. (2019). Modeling memory dynamics in visual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Annis, J., & Palmeri, T.J. (2018). Bayesian statistical approaches to evaluating cognitive models. *Wiley Interdisciplinary Reviews in Cognitive Science*.

Ashby, F.G. (1992). *Multidimensional Models of Perception and Cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc..

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

Ashby, F.G., & Lee W.W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, 120, 150.

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. (2005). Human category learning. *Annual Review of Psychology*, 56, 149-178.

Ashby, F.G., & Rosedahl, L. (2017). A neural interpretation of exemplar theory. *Psychological Review*, 124, 472-482.

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

Bowman, C.R., & Zeithamova, D. (2018). Abstract Memory Representations in the Ventromedial Prefrontal Cortex and Hippocampus Support Concept Generalization. *The Journal of Neuroscience, 38*(10), 2605–2614.

Busemeyer, J.R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 538-564.

Carroll, J.D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D.H. Krantz, R.C. Atkinson, R.D. Luce, and P. Suppes (Eds.), *Contemporary Developments in Mathematical Psychology*, Vol. 2. San Francisco: W.H. Freeman.

Chadwick, M.J., Anjum, R. S., Kumaran, D., Schacter, D.L., Spiers, H.J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences*, *113*(36), 10180–10185. https://doi.org/10.1073/PNAS.1610686113

Clarke, A., & Tyler, L.K. (2014). Object-Specific Semantic Coding in Human Perirhinal Cortex. *Journal of Neuroscience*, *34*(14), 4766–4775.

Cohen, A.L., & Nosofsky, R.M. (2000). An exemplar-retrieval model of speeded same-different judgments. *Journal of Experimental Psychology: Human Perception & Performance*, 26(5), 1549-1569.

Cohen, A.L., & Nosofsky, R.M. (2003). An extension of the exemplar-based random-walk model to separable-dimension stimuli. *Journal of Mathematical Psychology*, 47(2), 150-165.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., … Haxby, J.V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*(8), 2608–2618.

Cox, G.E., & Criss, A.H. (2019). Similarity leads to correlated processing: A dynamic model of encoding and recognition of episodic associations. *Manuscript under review*.

Curby, K.M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 94-107.

Davis, T., Goldwater, M., & Giron, J. (2017). From concrete examples to abstract relations: The rostrolateral prefrontal cortex integrates novel examples into relational categories. *Cerebral Cortex, 27*(4), 2652–2670.

Davis, T., Love, B.C., & Preston, A.R. (2012a). Striatal and hippocampal entropy and recognition signals in category learning: Simultaneous processes revealed by model-based fMRI. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 821-839.

Davis, T., Love, B.C., & Preston, A.R. (2012b). Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, 22, 260-273.

Davis, T., & Poldrack, R.A. (2014). Quantifying the internal structure of categories using a neural typicality measure. *Cerebral Cortex*, *24*(7), 1720–1737.

De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning and Memory, 15*(9), 717–727.

Ennis, D.M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, 117(4), 408-411.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

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

Erez, J., Cusack, R., Kendall, W., & Barense, M.D. (2016). Conjunctive Coding of Complex Object Features. *Cerebral Cortex*, *26*(5), 2271–2282.

Evans, K.K., Cohen, M., Tambouret, R., Horowitz, T., Kreindel, E., & Wolfe, J.M. (2011). Does visual expertise improve visual recognition memory? *Attention, Perception, & Psychophysics*, 73(1), 30-35.

Farrell, S., & Lewandowsky, S. (2018). *Computational Modeling of Cognition and Behavior*. Cambridge University Press.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol. 1*. Wiley.

Folstein, J., Gauthier, I., & Palmeri, T.J. (2012). Not all morph spaces stretch alike: How category learning affects object perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 807-820.

Folstein, J., Palmeri, T.J., Gauthier, I (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, 23(4), 814-823.

Folstein, J., Palmeri, T.J., Van Gulick, A.B., & Gauthier, I. (2015). Category learning stretches neural representations in visual cortex. *Current Directions in Psychological Science*, 24, 17-23.

Forstmann, B.U., & Wagenmakers, E.-J. (2015). *An Introduction to Model-Based Cognitive Neuroscience*. Springer: New York.

Freedman, D.J., Riesenhuber, M., Poggio, T., & Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, *23*(12), 5235–5246.

Garner, W.R. (1974). *The Processing of Information and Structure*. Erlbaum.

Gauthier, I., Anderson, A.W., Tarr, M.J., Skudlarski, P., & Gore, J.C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Current Biology*, *7*(9), 645–651.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573.

Gauthier, I., Tarr, M.J., & Bub, D. (2009). *Perceptual Expertise: Bridging Brain and Behavior*. Oxford University Press.

Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191–197.

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

Goldstone, R.L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, 130, 116-139.

Green, D.A., & Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23.

Gureckis, T.M., James, T.W., & Nosofsky, R.M. (2011). Re-evaluating dissociations between implicit and explicit category learning: An event-related fMRI study. *Journal of Cognitive Neuroscience*, 23, 1697-1709.

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior*, 16, 321-328.

Haxby, J.V., Gobbini, M.I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430.

Herzmann, G., & Curran, T. (2011). Experts' memory: An ERP study of perceptual expertise effects on encoding and recognition. *Memory & Cognition*, 39(3), 412-432.

Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93(4), 411-428.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95(4), 528-551.

Hintzman, D.L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology*, 41, 109-139.

Hutchinson, J.B., Uncapher, M.R., Weiner, K.S., Bressler, D.W., Silver, M.A., Preston, A.R., & Wagner, A.D. (2014). Functional heterogeneity in posterior parietal cortex across attention and episodic memory retrieval. *Cerebral Cortex*, *24*(1), 49–66.

Jiang, X., Bradley, E., Rini, R., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity. *Neuron, 53*(6), 891–903.

Johansen, M.K., & Palmeri, T.J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, 45, 482-553.

Kahana, M.J., & Sekuler, R. (2002). Recognizing spatial patterns: A noisy exemplar approach. *Vision Research*, 42, 2177-2192.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Knowlton, B.J., & Squire, L.R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*(5140), 1747-1749.

Kragel, J.E., Morton, N.W., & Polyn, S.M. (2015). Neural activity in the medial temporal lobe reveals the fidelity of mental time travel. *Journal of Neuroscience*, *35*(7), 2914–2926.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P.A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*, *60*(6), 1126–1141.

Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22-44.

Kuhl, B.A., Rissman, J., & Wagner, A.D. (2012). Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia, 50*(4), 458–469.

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

LaRocque, K.F., Smith, M.E., Carr, V.A., Witthoft, N., Grill-Spector, K., & Wagner, A.D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *Journal of Neuroscience*, *33*(13), 5466–5474.

Lewandowsky, S., Palmeri, T.J., & Waldmann, M.R. (2012). Introduction to special section on theory and data in categorization: Integrating computational, behavioral, and cognitive neuroscience approaches. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 803-806.

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible Coding for Categorical Decisions in the Human Brain. *Journal of Neuroscience*, 27(45), 12321–12330.

Liang, J.C., Wagner, A.D., & Preston, A.R. (2013). Content Representation in the Human Medial Temporal Lobe. *Cerebral Cortex*, *23*(1), 80-96,

Link, S.W. (1992). *The Wave Theory of Difference and Similarity*. Earlbaum.

Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.

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

Love, B.C., & Gureckis, T.M. (2007). Models in search of a brain. *Cognitive, Affective, & Behavioral Neuroscience*, 7(2), 90–108.

Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 103-189), Wiley.

Mack, M.L., Love, B.C., & Preston, A.R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *Proceedings of the National Academy of Sciences*, 113(46), 13203–13208.

Mack, M.L., Love, B.C., & Preston, A.R. (2018). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*, *680*, 31–38.

Mack, M.L., & Palmeri, T.J. (2010). Modeling categorization of scenes containing consistent versus inconsistent objects. *Journal of Vision*, 10(3), 1-11.

Mack, M.L., & Palmeri, T.J. (2011). The timing of visual object categorization. *Frontiers in Psychology*.

Mack, M.L., Preston, A.R., & Love, B.C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology*, 23(20), 2023-2027.

Mack, M.L., & Preston, A.R. (2016). Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. *NeuroImage*, 127, 144–157.

Mack, M.L., Preston, A.R., Love, B.C. (2020). Ventromedial prefrontal cortex compression during learning. *Nature Communications*, 11, 46.

Martin, C.B., Douglas, D., Newsome, R.N., Man, L.L., & Barense, M.D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *ELife*, *7*.

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*(3), 419–457.

McKinley, S.C., & Nosofsky, R.M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, 21(1), 128-148.

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

Metcalfe, J., & Fisher, R.P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition*, 14, 164-173.

Murdock, B.B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609-626.

Murphy, G. (2004). *The Big Book of Concepts*. MIT press.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92(3), 289-316.

Norman, K., & O'Reilly, R. (2003). Modeling Hippocampal and Neocortical Contributions to Recognition Memory: A Complementary-Learning-Systems Approach. *Psychological Review*, *110*(4), 611–646.

Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104-114.

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(1), 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. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology*, 34(4), 393-418.

Nosofsky, R.M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23(1), 94-140.

Nosofsky, R.M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 3-27.

Nosofsky, R.M. (1992a). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43, 25-53.

Nosofsky, R.M. (1992b). Exemplar-based approach to relating categorization, identification, and recognition. In F.G. Ashby (Ed.), *Multidimensional Models of Perception and Cognitio*n. (pp. 363-393). Lawrence Erlbaum Associates, Inc.

Nosofsky, R.M., Cox, G.E., Cao, R., & Shiffrin, R.M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(6), 1524–1539.

Nosofsky, R.M., Gluck, M., Palmeri, T.J., McKinley, S.C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, 22, 352-369.

Nosofsky, R.M., Little, D.R., & James, T.W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences*, 109, 333-338.

Nosofsky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, 75, 97-129.

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

Nosofsky, R.M., & Palmeri, T.J. (2015). An exemplar-based random-walk model of categorization and recognition. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford Handbook of Computational and Mathematical Psychology* (pp. 142–164). New York: Oxford University Press.

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

Nosofsky, R.M., Sanders, C., & McDaniel, M. (2018). A formal psychological model of classification applied to natural-science category learning. *Current Directions in Psychological Science*, 27, 129-135.

Nosofsky, R.M., & Smith, J.K. (1992). Similarity, identification, and categorization: Comment on Ashby and Lee (1991). *Journal of Experimental Psychology: General*, 121(2), 237–245.

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

Nosofsky, R.M., & Zaki, S.R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(5), 924-940.

Notman, L.A., Sowden, P.T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition, 95*(2), 1–14.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*(12), 1244–1252.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2003). The Effect of Category Learning on the Representation of Shape: Dimensions Can Be Biased but not Differentiated. Journal of Experimental Psychology: General, 132(4), 491–511.

Palmeri, T.J. (2014). An exemplar of model-based cognitive neuroscience. *Trends in Cognitive Science*, 18(2), 67-69.

Palmeri, T.J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 324-354.

Palmeri, T.J., & Cottrell, G. (2009). Modeling perceptual expertise. In I. Gauthier, M. Tarr, & D. Bub (Eds.), *Perceptual Expertise: Bridging Brain and Behavior*. Oxford University Press.

Palmeri, T.J., & Flanery, M.A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, 10, 526-530.

Palmeri, T.J., & Flanery, M.A. (2002). Memory systems and perceptual categorization. In B.H. Ross (Ed.), *The Psychology of Learning and Motivation* (Volume 41), Academic Press.

Palmeri, T.J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5, 291-303.

Palmeri, T.J., & Nosofsky, R.M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experiment Psychology: Learning, Memory, and Cognition*, 21, 548-568.

Palmeri, T.J., & Nosofsky, R.M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology*, 54, 197-235.

Palmeri, T.J., & Tarr, M. (2008). Visual object perception and long-term memory. In S. Luck & A. Hollingworth (Eds., pp. 163-207), *Visual Memory*. Oxford University Press.

Palmeri, T.J., Love, B.C., & Turner, B.M. (2017). Model-based cognitive neuroscience. *Journal of Mathematical Psychology*, 76, 59-64.

Palmeri, T.J., Wong, A.C.-N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Science*, 8, 378-386.

Poldrack, R.A., Clark, J., Pare-Blagoev, E.J., Shohamy, D., Moyano, J.C., Myers, C., & Gluck, M.A. (2001). Interactive memory systems in the human brain. *Nature*, 414(6863), 546-550.

Polyn, S.M., Natu, V.S., Cohen, J.D., & Norman, K.A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*, 1963–1966.

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

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

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347-356.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333-367.

Reber, P. J., Gitelman, D.R., Parrish, T.B., & Marsel Mesulam, M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, 15(4), 574–583.

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

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science*, 27(5), 709-748.

Richler, J.J., & Palmeri, T.J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews in Cognitive Science*, 5, 75-94.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.

Ross, D.A., Deroche, M., & Palmeri, T.J. (2014). Not just the norm: Exemplar-based models also predict face aftereffects. *Psychonomic Bulletin & Review*, 21, 47-70.

Rouder, J.N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, 133, 63-82.

Sanders, C.A., & Nosofsky, R.M. (2018). Using deep-learning representations of complex natural stimuli as input to psychological models of classification. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

Sakamoto, Y., & Love, B.C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, 133 (4), 534-553.

Seger, C. A., Braunlich, K., Wehe, H. S., & Liu, Z. (2015). Generalization in Category Learning: The Roles of Representational and Decisional Uncertainty. Journal of Neuroscience, 35(23), 8802–8812.

Schapiro, A.C., Turk-Browne, N.B., Botvinick, M.M., & Norman, K.A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711).

Shepard, R.N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22, 325-345.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210(4468), 390-398.

Shepard, R.N. (1987) Toward a universal law of generalization for psychological science. *Science*, 237(4820), 1317-1323.

Shepard, R.N., Hovland, C.I., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, 75(13), 1-42.

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM - retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4(2), 145-166.

Sigala, N., & Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320.

Squire, L.R., & Knowlton, B. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences, 92*, 12470-12474.

Squire, L.R., & Zola, S.M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences*, 93, 13515-13522.

Smith, J.D., & Minda, J.P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 800-811.

Staresina, B.P., Henson, R.N.A., Kriegeskorte, N., & Alink, A. (2012). Episodic reinstatement in the medial temporal lobe. *The Journal of Neuroscience*, 32(50), 18150–18156.

Sternberg, S. (1966). High speed scanning in human memory. *Science*, 153, 652-654.

Tompary, A., Duncan, K., & Davachi, L. (2016). High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. *Hippocampus, 26*(8), 995–1007.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson, *Organization of Memory*. Academic Press.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53(1), 1-25.

Turner, B.M., Forstmann, B.U., Love, B., Palmeri, T.J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, 76, 65-79.

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

Wong, A.C.-N., Palmeri, T.J., & Gauthier I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert – but which type? *Psychological Science*, 20, 1108-1117.

Wong, A.C.-N., Palmeri, T.J., Rogers, B.P., Gore, J.C., & Gauthier, I. (2009). Beyond shape: How you learn about objects affects how they are represented in visual cortex. *PLoS One*, 4(12), e8405.

Wutz, A., Loonis, R., Roy, J. E., Donoghue, J.A., & Miller, E.K. (2018). Different Levels of Category Abstraction by Different Dynamics in Different Prefrontal Areas. *Neuron, 97*(3), 716-726.

Zaki, S.R. (2005). Is categorization really intact in amnesia? A meta-analysis. *Psychonomic Bulletin and Review*, 11, 1048-1054.

Zaki, S.R., & Nosofsky, R.M. (2007). A high-distortion enhancement effect in the prototype-learning paradigm: Dramatic effects of category learning during test. *Memory & Cognition*, 35, 2088-2096.

Zeithamova, D., Mack, M.L., Braunlich, K., Davis, T., Seger, C.A., Van Kesteren, M.T.R., Wutz, A. (2019). Brain mechanisms of concept learning. *Journal of Neuroscience, 39*(42), 8259-8266.

Zeithamova, D., Maddox, W.T., & Schnyer, D.M. (2008). Dissociable prototype learning systems: evidence from brain imaging and behavior. *The Journal of Neuroscience, 28*(49), 13194–13201.

FIGURE CAPTIONS

Figure 5.1.1: Example discrimination, recognition, and classification decisions for a single visual object. Photograph by Sebastian Lehmann.

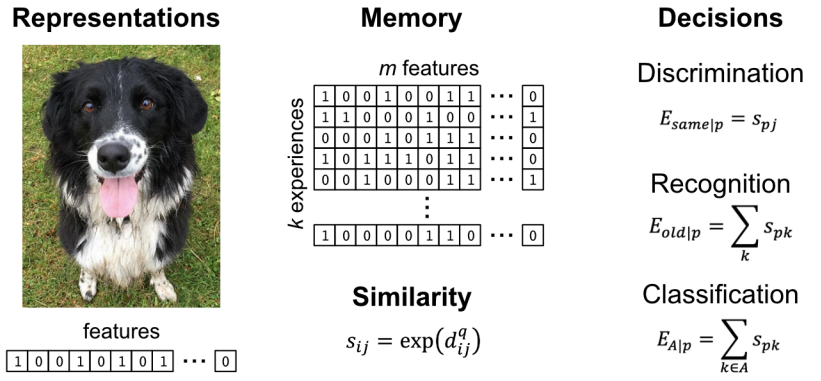
Figure 5.1.2: Am example modeling framework that accounts for discrimination, recognition, and classification decisions. Objects are composed of multiple features which are represented as vectors of feature values. In this depiction, features are assumed to be present or absent (i.e., a value of 1 or 0); however features can also have continuous values. These representations of experiences are stored in a memory matrix with each row representing a memory trace consisting of features from a prior experience. A similarity measure can be used to query the match between a current object, p, and a memory trace. Similarity is often formalized as the exponential of a Minkowski distance. These models components can be leveraged to make decisions about discrimination, recognition, and classification. Photograph by Anna Blumenthal.

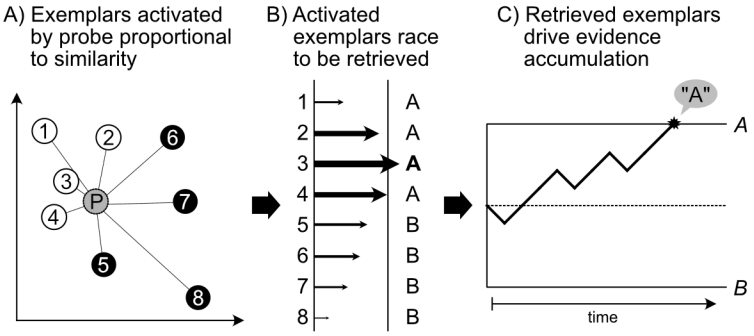
Figure 5.1.3: The EBRW model was the first classification model to formalize how similarity processes acting on multidimensional memory representations can account for both response probabilities and response times. A) In EBRW, a probe item (P) activates stored exemplars (1-8) proportional to their similarity. B) This similarity drives a race between activated exemplars to be retrieved from memory. C) Retrieved exemplars then drive an evidence accumulation process towards a decision. Figure adapted from Nosofsky et al. (2014), Figure 7.

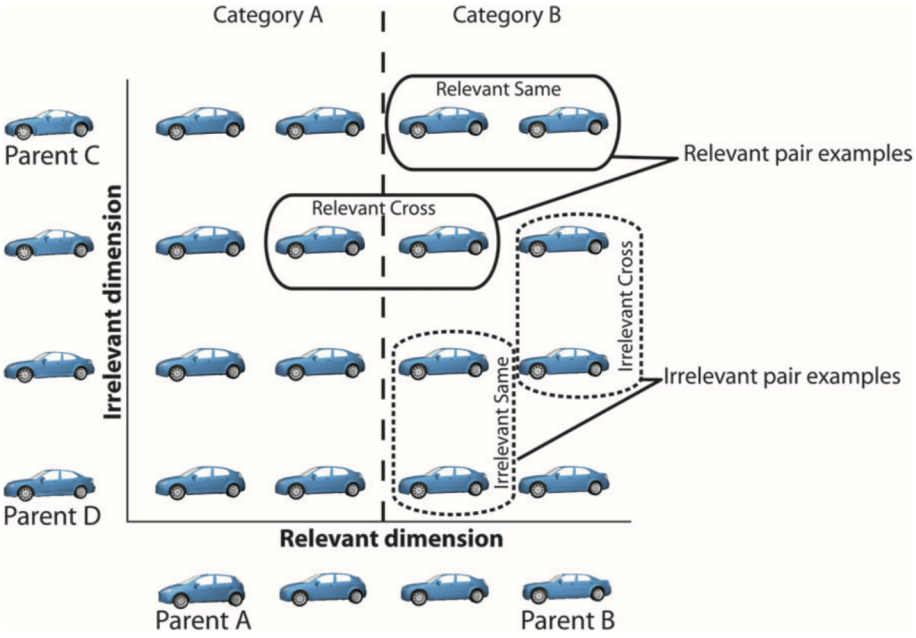
Figure 5.1.4: Illustration of the morph space from Folstein et al. (2013). Two sets of parent car models were first morphed to create two perceptual dimensions (parent A to B and parent C to

D). The full morph space was then created by factorially blending these two perceptual dimensions. Participants learned to classify the cars into two categories defined by the vertical dashed line. After classification learning, participants then performed an unrelated perceptual task in which pairs of morph cars were presented in sequence. Relevant pairs varied along the perceptual dimension relevant for the classification learning, irrelevant pairs varied along the other dimension. Folstein et al. (2013) found that relevant pairs were associated with a greater release from repetition suppression than irrelevant pairs in subregions of lateral occipital and ventral temporal cortex. These findings suggest that neural representations of the car stimuli were shaped by classification learning selectively to the dimension relevant for classification. Figure adapted from Folstein et al. (2013), Figure 1.


Discrimination
Same dog from
a moment ago?
Recognition
Did I see this
dog last week?
Classification
Is this dog a
Border Collie?


Representations
features
1 0 0 1 0 1 0 1 ⋯ 0
Memory
m features
k experiences
1 0 0 1 0 0 1 1 ⋯ 0
1 1 0 0 0 1 0 0 ⋯ 1
0 0 0 1 0 1 1 1 ⋯ 0
1 0 1 1 1 0 1 1 ⋯ 0
0 0 1 0 0 0 1 1 ⋯ 1
1 0 0 0 0 1 1 0 ⋯ 0
Similarity
$s_{ij} = \exp(d_{ij}^{q})$
Decisions
Discrimination
$E_{same|p} = s_{pj}$
Recognition
$E_{old|p} = \sum_{k} s_{pk}$
Classification
$E_{A|p} = \sum_{k \in A} s_{pk}$


A) Exemplars activated by probe proportional to similarity
B) Activated exemplars race to be retrieved
C) Retrieved exemplars drive evidence accumulation
P
1
2
3
4
5
6
7
8
A
A
A
A
B
B
B
B
"A"
A
B
time


Category A
Category B
Relevant Same
Relevant pair examples
Parent C
Relevant Cross
Irrelevant dimension
Irrelevant Cross
Irrelevant pair examples
Irrelevant Same
Parent D
Relevant dimension
Parent A
Parent B