Chapter 6

Visual Object Perception and Long-term Memory

Thomas J. Palmeri
Vanderbilt University

Michael J. Tarr
Brown University

> . . . they took twenty-seven eight-by-ten color glossy photographs with circles and arrows and a paragraph on the back of each one explaining what each one was to be used as evidence against us.
>
> —from *Alice's Restaurant,* by Arlo Guthrie

6.1

The late cognitive psychologist and memory theorist, Robert G. Crowder (MJT's colleague for 6 years), was fond of saying "memory *is* perception." What he meant was that memory is not a box in which things—objects, meanings, etc.—are stored. To Bob, memory was intrinsic to and a consequence of information processing, whether it be perceptual, linguistic, or cognitive in nature. In this context, it seems natural that our chapter discusses both object perception and object memory. At the same time, we acknowledge that in classic information processing flowcharts, as well as in the organization of most introductory textbooks, object perception precedes sensory or iconic memory (Chapter 2), which precedes short-term or working memory (Chapter 3), which precedes long-term memory (e.g., Atkinson & Shiffrin, 1968). By such accounts, visual perception is a modular, encapsulated *input* system, whereas memory is a cognitive box in which you put information (c.f., Fodor, 1983). But, as we shall see, contemporary research has demonstrated far closer links between object perception and object memory than anticipated by classic approaches to perception and cognition (Palmeri & Gauthier, 2004). Indeed, we are of the same mind as Bob: Drawing a clear demarcation between perception and memory is misguided.

We begin with a discussion of how objects are perceived and come to be represented over experience. And we describe the perceptual nature of the particular information stored in long-term memory (LTM) that allows us to recognize, identify, categorize, and perform perceptual skills on visual objects. These topics forge natural links to other chapters in this volume. To what extent do visual working memory and LTM have similar representational formats (Chapter 3)? How closely tied is visual working memory to visual LTM (Chapter 3)? What are the relationships between objects and scenes (Chapter 5)? What are the relationships between visual memory for objects and mental imagery (Chapter 9)?

Two themes weave their way throughout this chapter. One theme concerns the role of abstraction in perception and memory. Our everyday experience suggests that familiar objects can be recognized effortlessly under dramatically different viewing conditions, including changes in viewing position, object pose, and object configuration. According to this intuition, vision abstracts an invariant object representation that is removed from the particulars of specific experiences with that object. Everyday experience also suggests that our conceptual knowledge about objects is abstract. According to this intuition, although our experiences are specific, our knowledge is abstracted from those experiences. The ability to abstract from particular experiences is clear and is a hallmark of human perception and cognition. But this ability to abstract does not necessarily imply that object representations and object knowledge are themselves abstract in nature or that these abstractions are themselves amodal (Barsalou, 1999; Barsalou, Simmons, Barbey, & Wilson, 2003).

The second theme is how to carve up perception and memory into functional systems. A basic modus operandi of cognitive science is to "carve things up at the joints" (c.f., Fodor, 1983). This issue emerges in discussions of how to parcelize working memory (Chapter 3). And, in the context of object perception and object memory, are there domain-specific, informationally encapsulated subsystems for recognizing certain kinds of objects, remembering certain qualities about objects, or performing certain kinds of tasks on objects (Fig. 6–1)? These questions are particularly germane when interrelating theory and behavior with evidence from neurophysiology, functional brain imaging, and neuropsychological studies (Chapters 8 and 9).

6.2 VISUAL OBJECT PERCEPTION

How do we know that an object is the same object we have seen before? Or, at least, that it is of the same kind we have seen before? At first pass, this appears to be a trivial problem. One of us (TJP) can remember a meeting many years ago with an Associate Dean soon after being hired to talk about future research plans. The Dean simply could not understand why studying how people recognize objects could ever be a viable research problem. What could be simpler, he said. You just open your eyes and you see what's there [or as Terry Pratchett said in *Men at Arms* (1993): "How? He recognized him, of course. That's how you know who people are. You look at them and you say . . . that's him.

Figure 6–1. Phrenologists assumed that the mind was composed of numerous distinct innate faculties (secretiveness, benevolence, conjugality, self-esteem) and that each of these faculties had a unique location in the brain. Some contemporary accounts of brain organization localize function according to particular kinds of objects (faces, places, body parts) and particular kinds of memory (explicit memory, semantic memory, habits). Although some have rejected localization accounts entirely as a new form of phrenology (Uttal, 2001), we instead argue that localization of function should be characterized in terms of the representations and processes underlying the computational mechanisms of visual cognition.

That's called re-cog-nit-ion."]. That we all share such naïve intuitions belies the tremendous computational challenges facing our visual system with every glance of the world around us and ignores the fact that about one-half of our cortex is related to vision. The dynamic, ever-changing world conspires to present a dramatically different stimulus to our eyes even though the very same physical object may be present in front of us. Not only do we overcome such variation, but our perception of the world appears stable. Three-dimensional objects seem stable as we move around, as objects move around, and as the lighting changes (Fig. 6–2). But how does the visual system allow us to perceive this stability when the two-dimensional images falling on our retinae are changing so dramatically?

Answers to these questions are rooted in our visual memory for objects; that is, how they are represented with respect to such variation. This is true regardless of which form memory takes: I can ask you whether you saw this particular object recently (working memory), whether you have ever seen this object before (recognition memory), what category this object belongs to (classification), or to do something with an object (procedural task). All of these tasks require you to compare a representation of the perceived input with a representation encoded in memory. Classic approaches to cognitive science have often assumed that this comparison is amodal, and some contemporary approaches assume significant abstraction from any previous experience. Even so, understanding visual memory for objects requires understanding not only the representations and processes underlying those memories, but also understanding the inputs to memory—the perceptual representations. At the same time, more contemporary theories of visual memory assume that this comparison—the process of object recognition—is inherently perceptual. That is, visual memories for objects are part and parcel of the perception of those same objects, and object recognition is accomplished by comparing two *perceptual* representations.

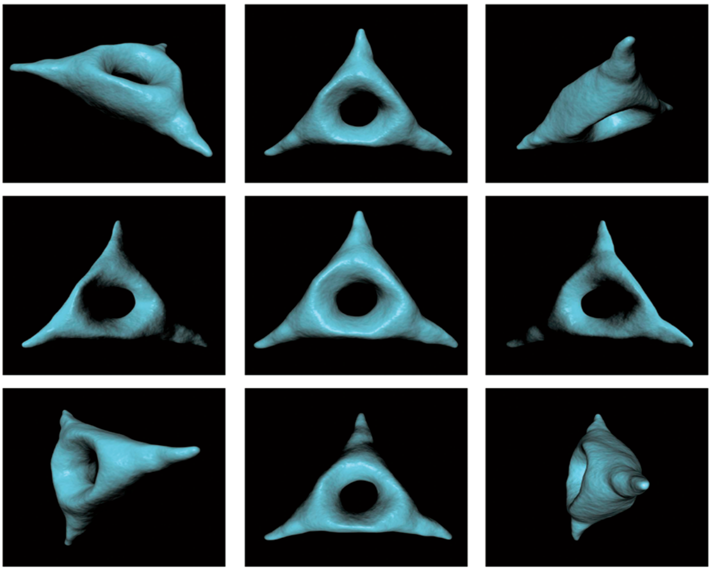

Figure 6–2. This figure illustrates the dramatic variability in viewed images of the same object when subjected to rotation along various axes and changes in lighting direction.

Under this view, the nature of the perceptual representations of objects created by the visual system place strong constraints on the nature of representations for objects in visual memory. Logically, memory representations abstracted from visual experience can be no less abstract than the visual representations derived from those visual experiences. In other words, if high-level visual representations have normalized away many of the perceptual details of visual experience, then memory representations will be void of those perceptual details as well. Of course, the converse need not be true: If visual representations retain some details, then memory representations might still be more abstract, and need not reflect those perceptual details. Such is the more standard view within both the cognitive and high-level vision communities (Biederman, 1987). In contrast, we propose that memory representations of objects do indeed retain perceptual details (see Chapter 5). And, in fact, many of those detailed memory representations are the very representations that underlie visual object perception itself.

6.3 THE PROBLEM

Very young children are fond of pointing to two similar objects and declaring "same thing!" So-called *basic-level* recognition involves categorizing visually similar yet distinct objects as members of the same class. Thus, one form of invariance requires our visual systems to perform a many-to-one mapping between individual exemplars and object categories. At the same time, individual exemplars of three-dimensional objects rarely appear the same from one moment to the next. Variation in the two-dimensional images falling on our retinae arises from almost any change in viewing conditions, including changes in position, object pose, lighting, or object configuration. We never really see the same object, or at least the same retinal image of an object, twice. This form of invariance requires our visual systems to perform a many-to-one mapping between individual *views* of objects and their unique identities.

Almost all solutions to the problem of vision begin by generally characterizing visual processing as a form of dimensionality reduction. The retinal representation has extremely high dimensionality in that each of the 120 million or so photoreceptors can independently encode a different (albeit highly local) aspect of the visual scene. The visual system transforms this high-dimensional stimulus representation into a low-dimensional representation (at least relative to the dimensionality of the retinal stimulation) that is used to recognize or categorize objects. Different theories propose different solutions to the problem of creating a low-dimensional object representation. Theories differ rather markedly in the form of visual representation and, in particular, how great a dimensionality reduction is assumed. In turn, such assumptions themselves are based on the assumptions each theory makes about the goals of vision.

6.3.1 Structural-description Theories

One early and influential class of models assumed that the fundamental goal of vision was to *reconstruct* the three-dimensional structure of objects and their spatial relationships (Marr & Nishihara, 1978). The appeal of such an approach is that many sources of variance are "partialed out" as a consequence of the reconstruction process. For example, different images arising from changes in lighting are mapped into a single shape, and different images arising from changes in viewpoint are mapped into a single three-dimensional object representation. This same reconstruction process achieves theoretically optimal dimensionality reduction: mapping the high-dimensional image array arriving at our eyes into a low-dimensional scene composed of objects and surfaces. One of the most intuitive proposals for constructing such representations, originally put forth by Marr and Nishihara (1978) and elaborated by Biederman (1987), assumes that every given object can be described in terms of generic three-dimensional components ("primitives") and their spatial relationships. The key idea is that the recovered three-dimensional *structural description* will be invariant over both class variation and viewing conditions, thereby directly addressing the twin challenges facing vision. That is, different views of an object *and* different exemplars within an object class will all map to the same configuration of three-dimensional primitives. This approach assumes the primary goal of vision is basic-level recognition without respect to image characteristics arising from lighting, viewpoint, and other variables. In this context, structural-description models—if achievable—are near-optimal.

Beyond the popularization of the study of object recognition, Biederman's contribution was to realize that structural-description models are far more likely to succeed if the mapping from images to primitives is precisely defined. Marr and Nishihara assumed an unrealized computational process that relied on dividing an object into parts, finding the major axis of each part, and, finally, deriving a cross-section capturing the three-dimensional appearance of a part with respect to its axis (although this is somewhat of an oversimplification, it serves to illustrate the basic principles of their theory). For, example, a three-dimensional cylinder might be described as a straight axis with a circular cross-section. This method for describing object parts leaves a great deal to the imagination: How are axes found, and how is the cross-section derived? How well does this description generalize from one exemplar to another? How consistent is this process over image variation? The concern is that, although the intent is dimensionality reduction, the actual mapping may be inefficient, with slight variations in axes or cross-sections leading to different representations.

To address such concerns, Biederman (1987) based his recognition-by-components (RBC) theory on a small set of *qualitative* three-dimensional primitives known as "Geons" (Fig. 6–3). Two innovations are included in RBC. First, primitives are recovered by attending to configurations of "viewpoint invariant properties" in the two-dimensional image. For example, a brick (one type of Geon) might be inferred when one encounters two sets of three parallel lines, several L junctions, several arrow junctions, and a Y junction (Fig. 6–4). Notice that this description avoids quantitative specifics about object parts: Many different brick-like parts from many different viewpoints will exhibit this configuration of image features and be reconstructed simply as a Geon brick in RBC's vocabulary (Fig. 6–4). Second, the entire repertoire of Geons numbers

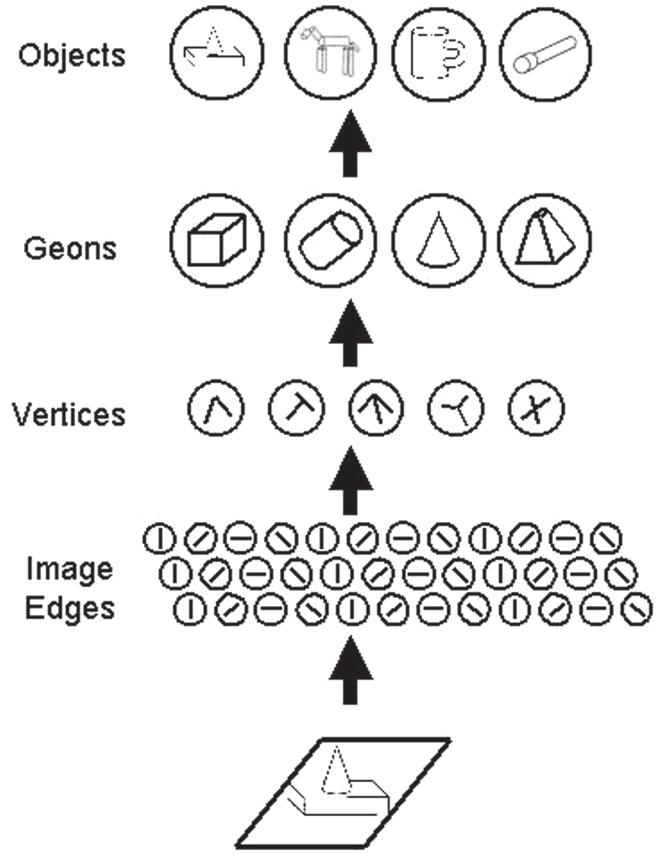

Figure 6–3. Recognition-by-components (Biederman, 1987; Hummel & Biederman, 1992) assumes that a retinal image is initially described in terms of its edges. A variety of nonaccidental primitive features are extracted from this edge description, such as L junctions, Y junctions, and other properties. Combinations of various viewpoint invariant primitives signal the presence of one of the small number of geometric icons (Geons). Viewpoint invariant object recognition involves recognizing the particular combination and relative configuration of the viewpoint-invariant Geon representations extracted from a complex object.

about 35 distinct primitives. Biederman's thesis is that any basic-level, visually defined object category may be uniquely represented by a small subset of these primitives in a particular spatial configuration. For example, a wide variety of birds are made up of roughly the same parts—head, body, wings, etc.—the assumption being that, across different birds (the exception being highly visually dissimilar birds such as penguins), the image projections of these parts will yield the same Geons in the same Geon configuration, that is, a single visual representation for many different birds. Thus, RBC provides a more satisfying

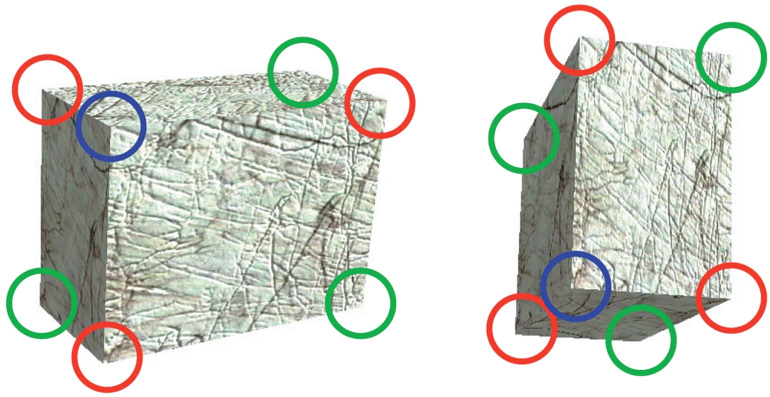

Figure 6–4. Illustration of various primitives that could be extracted from a three-dimensional brick as seen from two different viewpoints, including L junctions, arrow junctions, and Y junctions. Adapted with permission from Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience* 2: 1019–1025. Macmillan Publishers Ltd.

(less filling) approach to structural-description models: The inferential perception mechanisms for reconstruction are well-specified, and the mapping from high-to-low dimensionality is inherent in the end representation.

One of the most salient characteristics of structural-description models is their invariance over image variation. In particular, it is often assumed that changes in three-dimensional viewing position provide one of the strongest tests of theories of object recognition. Structural-description models, and RBC in particular, posit viewpoint invariance. That is, the same representation is derived, irrespective of prior experience, over a wide range of viewing conditions (although with opaque objects different configurations of Geons may be visible across large changes in viewpoint). The behavioral implication of this is that recognition performance should be independent of the particular viewpoint from which the object is seen (Biederman & Gerhardstein, 1993). This prediction is also consistent with our intuitions: Our recognition of familiar objects from unfamiliar viewpoints feels effortless.

At the same time, the prediction of viewpoint invariance seems at odds with the idea that we do remember a great deal of what we see, including the particular appearance of individual objects from specific vantage points. Indeed, a classic study in cognitive science demonstrated that our memories for objects are better at "canonical" viewpoints as compared to others (Palmer, Rosch, & Chase, 1981). If we have learned nothing over the past half century, it is that we should not always trust our conscious intuitions: What seems effortless may actually be an effortful, albeit unconscious, process. For the past 15 years or so, something of a cottage industry has arisen for testing these assumptions. More specifically, many different labs have attempted to devise psychophysical tests of the viewpoint-invariance assumption assessing, when, if ever, objects are recognized in a viewpoint-invariant manner (e.g., with equivalent error rates and response times for both familiar and unfamiliar views of an object (Bülthoff & Edelman, 1992; Humphrey & Khan, 1992; Jolicoeur, 1985; Lawson & Humphreys, 1996; Poggio & Edelman, 1990; Tarr, 1995; Tarr et al., 1998; Tarr & Pinker, 1989). The conclusion is . . . it depends. Certainly, there are limited conditions under which viewpoint invariance is achieved immediately (Biederman & Gerhardstein, 1993; Tarr & Bülthoff, 1998; Tarr, Kersten, & Bülthoff, 1998). However, the vast majority of the time, viewpoint invariance is only attainable with experience. More specifically, numerous studies have found that if observers learn to recognize novel objects from specific viewpoints, they are both faster and more accurate at recognizing these same objects from those familiar viewpoints relative to unfamiliar viewpoints (Bülthoff & Edelman, 1992; Tarr, 1995; Tarr & Pinker, 1989). Recognition performance at unfamiliar viewpoints is systematically related to those views that are familiar, with observers taking progressively more time and being progressively less accurate as the distance between the unfamiliar and the familiar increases. Consequently, viewpoint invariance seems to be achieved by learning about the appearance of objects from multiple viewpoints, not by deriving structural descriptions. Human object recognition seems to rely on multiple *views*, where each view encodes the appearance of an object under specific viewing conditions, including viewpoint, pose, configuration, and lighting (Tarr, Kersten, & Bülthoff, 1998), and a collection of such views constitutes the long-term visual representation of a given object.

6.3.2 Image-based Theories

Over the past decade, image-based theories have become popular as an alternative to structural-description models. These theories are based in part on the already-mentioned empirical findings regarding viewpoint invariance and in part on different assumptions regarding the goals of vision (Edelman, 1999; Shepard, 1994). Rather than assuming that we reconstruct the three-dimensional world, image-based approaches typically stress generalization from past to present experience (Shepard, 1994). Consider that we are highly unlikely to ever experience the same situation twice. Because similar objects often give rise to similar consequences, survival demands that we recognize these similarities (Shepard, 1987). One possible solution is for visual perception to create a faithful representation of each object that preserves its shape and three-dimensional structure (Marr, 1982). Similar objects should have similar or, as in the case of RBC, identical, mental representations. But an alternative solution is to create representations that preserve the similarity structure between objects without necessarily representing three-dimensional object structure explicitly (Edelman, 1997, 1999). As mentioned, image-based theories assume that objects are represented in terms of their similarity to collections of views that are instantiated in memory. Physically similar objects in the world, viewed under similar conditions, will all be similar to the same sets of views, allowing for generalization to occur, without any explicit representation of three-dimensional shape. At least for purposes of object recognition, representation of three-dimensional shape may not be necessary.

Particularly in the context of a volume on visual memory, we cannot understate one fundamental difference between structural-description and image-based theories. Structural-description theories assume a fixed processing architecture and a fixed set of primitives that construct object representations, irrespective of visual experience; particular configurations of primitives (say Geons) must be learned in order to categorize birds from dogs, but the primitives themselves (say, Geons) are not shaped by experience. By contrast, image-based theories assume that visual experience plays a significant role in shaping our visual behavior throughout a lifetime. Stable object perception is achieved by deploying our astonishing capacities for remembering particular experiences with particular objects under particular viewing conditions. We do encode a great deal of what we see as it originally appears. Object perception *is* visual memory.

But if we represent three-dimensional objects as collections of specific views, how do we manage to attain view invariance? One clue may be found in the systematic pattern of performance seen for the recognition of familiar objects in unfamiliar viewpoints. According to one view (Tarr & Pinker, 1989), this pattern is a consequence of mental rotation (Shepard & Metzler, 1971) or a continuous alignment process (Ullman, 1989) to transform unfamiliar viewpoints to familiar views in visual memory, with familiar viewpoints being recognized without the need for any transformation. The strongest evidence favoring this interpretation is the nearly identical linear reaction time pattern across viewpoint obtained for the same objects in naming and left- and right-handedness discrimination tasks (Tarr & Pinker, 1989). However, in an example of how neuroimaging can inform us regarding cognitive processes, Gauthier et al. (2002) found that entirely different brain systems exhibited viewpoint-dependent activity for recognition tasks and mental rotation tasks. Consistent with current thinking on the "division of labor" in the primate visual system (Goodale & Milner, 1992), the recognition of objects in unfamiliar viewpoints preferentially recruited the fusiform region along the ventral pathway, whereas handedness discriminations recruited the superior parietal lobe along the dorsal pathway (Gauthier et al., 2002). Thus, the computational mechanism underlying viewpoint-dependent recognition behavior seems to be based on "evidence accumulation" across neural subunits coding for different features of the object (e.g., Perret, Oram, & Ashbridge, 1998) and not, as suggested by Tarr and Pinker (1989), on the continuous transformation process of mental rotation (assumed to be isomorphic with physical rotations).

View invariance might be achieved by generalizing according to the similarity relationships between perceptual representations and stored views, without a need for any explicit image transformation (Poggio & Edelman, 1990; Riesenhuber & Poggio, 1999, 2002); see Figure 6–5 for one example. Indeed, the predictions of image-based models are consistent with the patterns of interpolation between learned views and limited extrapolation beyond learned views seen experimentally (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992). One of the appealing aspects of using similarity as a means to invariance is that the same mechanisms can account for how we generalize across both viewing

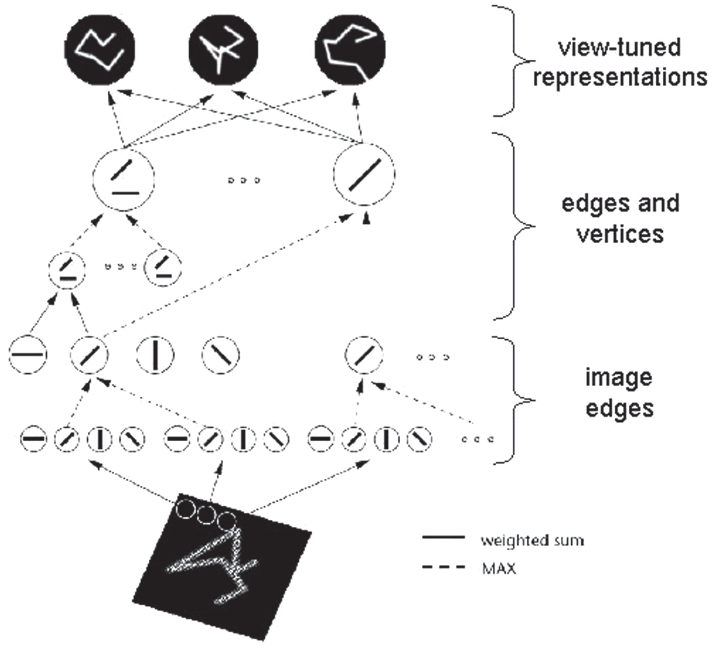

Figure 6–5. Sketch of one image-based model of object recognition. A hierarchy of representational layers consisting of weighted sum and max integration rules achieves representations that are scale- and translation-invariant. Objects are ultimately represented according to their similarity to learned view-tuned representations. Adapted with permission from Riesenhuber & Poggio, 1999, 2000.

and category variation. Specifically, invariance over viewing conditions can be achieved by encoding multiple views of *individual* objects. Invariance over object shape can be achieved by encoding multiple views of *multiple* objects. Both statements distill down to a memory-based explanation in which we remember a great deal about what we see. Thus, given a sufficient number of views per an object or class, viewpoint-invariant recognition is possible (Poggio & Edelman, 1990). Likewise, given a sufficient number of exemplars per a category, object categorization, even for new exemplars, is possible, and entirely novel objects can be represented in a distributed fashion according to their similarity to a relatively small number of views of known objects (Edelman, 1999).

At the same time, this memory-based account seems to miss a fundamental fact about human vision: We are incredibly good at generalizing from a *small number* of examples. Thus, exposure to a single view of an object or a single exemplar of a category is often sufficient to allow us to recognize that object across many different sources of variance, including identifying novel members of that category. Although some of this "heavy lifting" might be accomplished through view-based mechanisms, some forms of generalization also appear to require structural models (e.g., articulation; Barenholtz & Tarr, submitted). So, is the structural-description account right after all? In some sense yes.[1] An image-based account relying on *undifferentiated*, template-like representations would have difficulty generalizing across many types of image variation (Hummel, 1998). In contrast, structural-description models, including those proposed by Biederman and others, can readily generalize across both viewing conditions and members of a class. The caveat here is that this is true for *any* model relying on compositional structure, including models that use image-based features (e.g., Zhang & Cottrell, 2005); that is, any model that includes spatial relations between reusable features or parts (Bienenstock & Geman, 1995). Thus, the take-home message is not that structural-description models are right and image-based models are wrong (or vice versa), but, as discussed in the next section, elements of both approaches are likely be incorporated into a viable theory of object recognition (Barenholtz & Tarr, 2007). Structural-description models teach us that parts or features and their spatial relations are important. Image-based models teach us that specific visual memories are important. In combination, we can think of long-term visual memories as collections of spatially related image features that are matched to percepts on the basis of similarity within a low-dimensional (relative to images) image feature space (the nature of the features still being an open question). Dimensionality reduction is realized by moving from image pixels to image features and from a spatially undifferentiated image to spatial relations between features. At the same time, specificity is preserved in the spatial relations between features encoding local properties of the image. By preserving meaningful similarity relationships, yet reducing overall dimensionality, this architecture enables generalization from small numbers of examples. In contrast, a qualitative structural-description model (e.g., RBC) ignores meaningful similarity relationships by reducing dimensionality to the point at which many exemplars or many views are simply the *same* representation. Conversely, a template model breaks meaningful similarity relationships by preserving too much dimensionality to the point at which each exemplar or view is a *different* representation.

At this point, you might be asking, exactly how do you define "meaningful similarity"? Consider that a single exemplar has a similarity relationship with other members of the same category (which is why categories arise in the first place). Likewise, a single view has a similarity relationship with other views of the same object. These particular relationships are representationally meaningful and should be present in visual memory. Moreover, they do important work in explaining why, before we have learned many exemplars or views, we are able to recognize new instances of a category with few exemplars or familiar objects

1 Behavioral data on view specificity (e.g., Tarr et al., 1998) speaks to the nature of the features used in object representations, for example, arguing for image-based features rather than Geons. However, these same data are agnostic as to whether features participate in structural descriptions or only exist in more template-like forms.

in completely novel views. These generalization processes, unlike the invariance conferred by multiple instances in memory, take *more* time and produce *more* errors as the similarity between the known and the unknown increases. In support of such representational assumptions, it is well established that objects learned at one view are more poorly and more slowly recognized at new views (Tarr, 1995; Tarr & Pinker, 1990) and that individual object-selective neurons tend to preferentially respond to specific object views (Logothetis & Pauls, 1995; Perrett et al., 1985). This sort of view-tuning may appear puzzling when considered at the single neuron level: If objects are represented by individual neurons tuned to specific views, how can any sort of invariance be achieved? The answer lies, of course, in considering populations of neurons as the actual neural code for objects. Individual neurons may code—from a familiar viewpoint—the complex features or parts of which objects are composed; that is, instantiating the representational architecture outlined earlier.

Consistent with this approach, Perrett et al. (1998) proposed that recognition then takes the form of an *accumulation of evidence* across all neurons selective for some aspect of a given object—a variation on classic stochastic accumulation-of-evidence models (Nosofsky & Palmeri, 1997; Ratcliff, 1978; Ratcliff & Smith, 2004; Smith, 2000; Smith & Ratcliff, 2004). Such models are achieving new prominence in explaining the neural bases of perceptual decision making across a variety of domains (Boucher, Palmeri, Logan, & Schall, 2006; Gold & Shadlen, 2001; Roitman & Shadlen, 2002; Schall, 2001, 2004). Critically, these models implement similarity relationships as a function of their pooled neural responses. For example, during recognition of a novel object view, the particular rate of accumulation will depend on the similarity between visible features in the present viewpoint and the view-specific features for which individual neurons are tuned (Perrett, Oram, & Ashbridge, 1998). Across a population of object-selective neurons, sufficient neural evidence (summed neural activity) will accumulate more slowly when the current appearance of an object is dissimilar from its learned appearance (Fig. 6–6). In contrast, when an object's appearance is close to a previously experienced view, evidence across the appropriate neural population will accumulate more rapidly. Thus, systematic behavioral changes in recognition performance with changes in viewpoint may be explained as a consequence of how similarity is computed between new object perceptual representations and their previously learned neural representations and how evidence is accumulated over time for a perceptual decision.

In these models, recognition amounts to reaching a threshold of sufficient evidence across a neural population. Unfamiliar views of objects will require more time to reach threshold, but will be successfully recognized given some similarity between an input and known viewpoints. Unfamiliar exemplars within a familiar class can likewise be recognized given some similarity (Tarr & Gauthier, 1998) with known exemplars from within that class. Consistent with the idea that view and category generalization rely on common mechanisms, one behavioral implication is that familiarity with individual objects should facilitate the viewpoint-dependent recognition of other, visually similar objects,

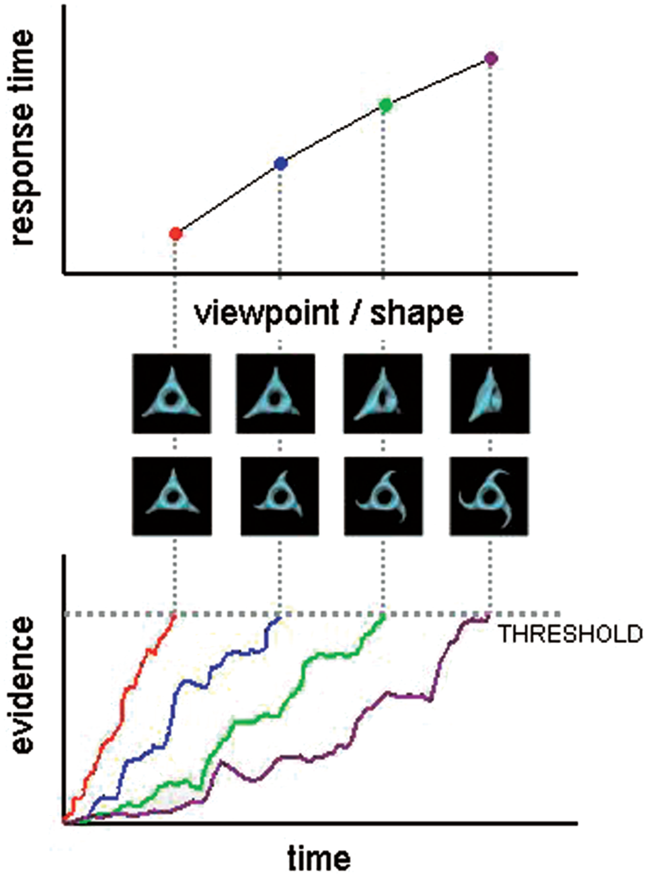

Figure 6–6. A broad class of models of perceptual decision making assume that evidence (*y-axis*) accumulates over time (*x-axis*). A response is made when an evidence threshold is reached (*bottom panel*). Response time is that time at which the threshold is reached (*top panel*). In the case of making a recognition or categorization decision, the rate of accumulation of evidence depends on the similarity between the perceived object and the stored memory representation of the object to be recognized or the class of objects to be categorized. Similarity can vary with either viewpoint or physical shape or both.

as borne out by several studies (Edelman, 1995, 1999; Tarr & Gauthier, 1998). Whether the same mechanism can account for all forms of object invariance remains unknown, although it seems possible that configuration and lighting variation present unique challenges that may require the inclusion of distinct forms of structural information (e.g., Bienenstock & Geman, 1995). Finally, as discussed later in this chapter, accumulation of evidence based on similarity to stored exemplars has also been proposed as a solution to the more general problem of categorization (Nosofsky & Palmeri, 1997). Thus, mechanisms based on similar (sic) computational principles seem to underlie many cognitive processes.

6.3.3 Hybrid Theories

As discussed earlier, one of the key differences between structural-description and image-based theories is the compositional nature of object representations. Under the cartoon view of the world, structural descriptions represent objects in terms of viewpoint-independent three-dimensional parts and their spatial relations (Biederman, 1987), and views represent objects in terms of a holistic image of the entire object (Edelman, 1997). However, both intuition and empirical evidence (Garner, 1974; Stankiewicz, 2002; Tversky, 1977) suggest that we represent complex objects in a compositional manner— objects are decomposable into parts. Yet these same intuitions and other empirical evidence (Hayward & Tarr, 1997; Tarr, Williams, Hayward, & Gauthier, 1998) suggest that these parts are not simple three-dimensional volumes. Is there a way to marry the best qualities of image-based theories with the compositional representations of structural-description theories?

One recent approach proposed by Ullman, Vidal-Naquet, and Sali (2002) measured mutual information between features and basic-level categories to discover the image features that were most informative for classification (see also Schyns & Rodet, 1997). They showed that features of "intermediate complexity" were best for basic-level classification. For faces, these features included what we would generally call the "parts" of a face such as the eyes or the nose; for cars, these included "parts" such as a wheel or the drivers' side window (Fig. 6–7).

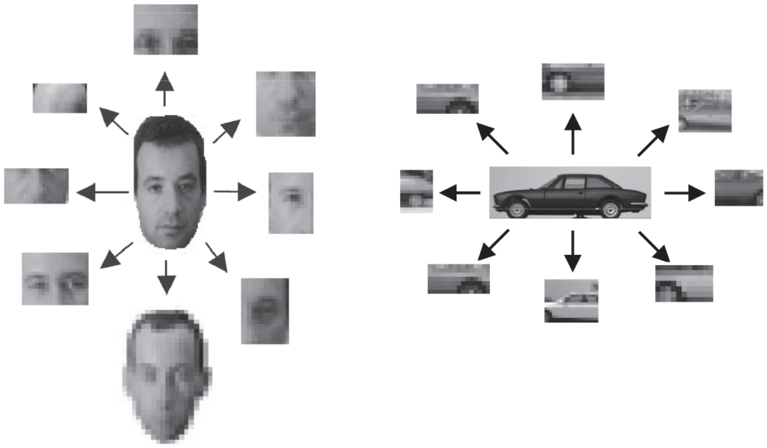

Figure 6–7. Image-based visual features of intermediate complexity maximize delivered information with respect to a basic-level category of objects. The figure shows examples of face fragments and car fragments. Adapted with permission from Ullman, Vidal-Naquet, & Sali, 2002. Zhang and Cottrell (2005) found somewhat larger and more complex image-based visual features for subordinate identification. Adapted with permission from Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5:682–687. Macmillan Publishers Ltd.

For other classes, "parts" are likely to include a wide variety of features, including, depending on the class, non–shape-based properties such as color or texture. Although there is as yet no direct evidence, it is tempting to speculate about the relationship of such "ad-hoc" features to the observed feature-selective responses of neurons in TEO (K. Tanaka, 1996, 2002). What is intriguing is that selective responses for individual neurons are elicited by somewhat odd patterns that do not correspond to what we might typically think of as distinct object parts (Fig. 6–8). Indeed, they appear to be ad-hoc and of intermediate complexity. This correspondence is less surprising if we consider that the features incorporated into the model proposed by Ullman et al. were found using an algorithm that operated on raw images without any intervention from a human teacher. These features emerged because they provided maximal information for the basic-level classification of those images. It is also important to emphasize that these are *viewpoint-dependent image-based features*, not anything like Geons or other volumetric primitives. Moreover, spatial relations between these parts are not explicitly encoded; rather, the local context is preserved for each image-based feature, and local features overlap, enabling an implicit representation of configural information.

To be clear, Ullman et al. (2002) proposed a solution to basic-level classification (classifying an object as a face or a car), not to more subordinate-level classification (classifying an object as Steve Luck or as a Porsche Boxster). Recently, Zhang and Cottrell (2005) extended the Ullman et al. approach to discover the image features possessing maximal informativeness for subordinate-level classification. What they found was that these image features were larger

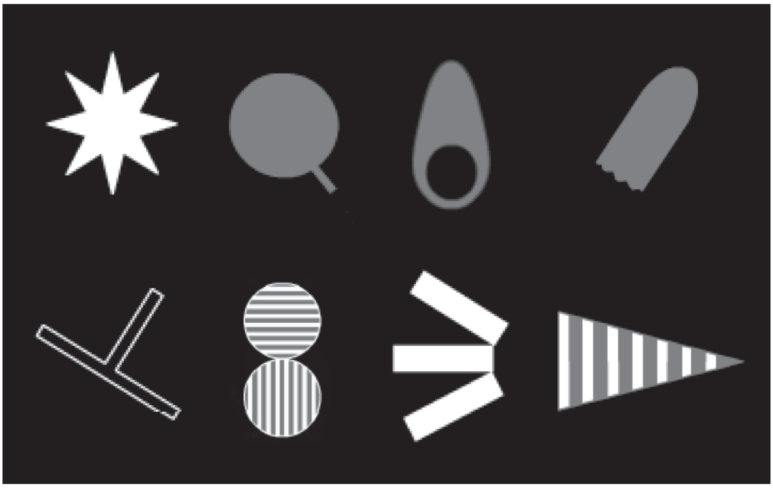

Figure 6–8. Some examples of ad hoc "features" that are preferred by certain cells in IT cortex of macaque monkeys. Adapted with permission from work by Keiji Tanaka and colleagues (e. g., K. Tanaka, 1996, 2002; see also http://www. brain. riken. go. jp/labs/cbms/tanaka. html).

and more complex than the features Ullman et al. reported for basic-level classification. For example, for face individuation, these features included an eye and a nose, or an eye with part of the nose and mouth. Thus, it is possible that accounting for configural/holistic effects, particularly as seen in face and expert-level object recognition, requires assembling *hierarchies* of features, not simply relating them in a single level of spatial relations (Gauthier & Tarr, 2002; Maurer, Grand, & Mondloch, 2002). Note that these maximally informative "parts" were still not the entire faces themselves. Thus, this approach to image-based representation is different from encoding complete views of objects for subordinate-level classification (Edelman, 1999). Here, incomplete yet complex image-based features were not only sufficient to successfully identify objects at the subordinate level, but provided the maximal information to support such classification.

As alluded to earlier, hybrid theories suggest a compositional aspect to the representation of objects in terms of ad-hoc features (parts). They also reflect the fact that we know more about objects than just their shape. We can remember an object's color, position, orientation, or size, and can use such dimensions to determine an object's identity or category if those dimensions prove diagnostic for those perceptual decisions (Kruschke, 1992; Naor-Raz, Tarr, & Kersten, 2003; Nosofsky, 1998). So, the perceptual representation of a complex object may consist of a collection of image-based parts, color, orientation, location, and other independent or semi-independent perceptual dimensions (Ashby & Townsend, 1986) and their spatial relations. Such information is recruited for a given task based on the difficulty of the discrimination at hand—that is, the degree to which particular features are stored or retrieved from working or long-term visual memory is modulated by task complexity, not object complexity per se.

Finally, many theories (and most experiments) of object recognition live in a world with just a single object at a time (but see Mozer, 1991; Mozer & Sitton, 1998). This assumption is typically justified by invoking early attentional processes that select one object for high-level visual processing (Treisman & Gelade, 1980). Thus, the object perception problem is often reduced to the recognition of decontextualized, static objects. Yet, natural vision systems excel at dynamic scene recognition; that is, the invariant recognition of not only objects but their entire context as well as their actions. Indeed, it is probably not possible or desirable to completely separate the problems of object and scene recognition. All levels of object recognition seem contextualized. Recognizing an object part is dramatically facilitated by considering it in the context of the whole object (Tanaka & Farah, 1993). Scene recognition can be impossible without considering constituent objects (but see Oliva, 2005), and object recognition itself is more effective if the nature of the scene has been established (Hollingworth, in press, 2006; Hollingworth & Henderson, 2002). Similarly, object recognition is enhanced by the inclusion of diagnostic dynamic information (Johanson, 1973; Vuong & Tarr, 2004). Thus, any architecture for object recognition and scene recognition should include dynamic information and processes that enable a compositional hierarchy of contexts to interact in a manner that aids interpretation.

6.3.4 Category-specific Visual Object Perception?

Although the question of invariance has often dominated thinking on visual object perception, recent neuroimaging results have focused on the perception and recognition of particular classes of object. One of the challenges to the human visual system is discriminating between objects at different levels of specificity, including, for some classes, individuals within a homogeneous set, the most salient example being face recognition. Following in the tradition of neuropsychology (Lissauer, 1890), the specific question addressed within this domain is often whether faces are "special" or not (Farah, Wilson, Drain, & Tanaka, 1998); that is, whether there exists a functional or neural module dedicated to face recognition. Although the form of this debate has varied, neuroimaging studies speak logically to the issue of neural modularity (Fodor, 1983) in that neuroimaging methods necessarily produce spatially localized neural responses associated with specific tasks—patterns that look temptingly like neural modules (recall Fig. 6–1). Of course, it is sometimes difficult to pin down what one means by a "module." Does module refer to an encapsulated cognitive function (which Fodor believed could only apply to perceptual systems, but others have extended to most cognitive abilities)? Or, does module refer to spatially localized brain regions that appear to subserve particular functions (independently of how such regions may interact with other regions)? Across multiple literatures, use of the term module is fast and loose, sometimes even to the point of absurdity (e.g., Fisher, Aron, & Brown, 2006; Beauregard & Paquette, 2006). Even worse, within the neuroimaging literature, there has been a tendency to associate functional localization with functional specialization, as if a localized peak of neural activity is equivalent to a discrete module dedicated to accomplishing a singular task.

With regard to a putative functional/neural module for face recognition, a large body of data show distinct regions of the visual system that appear to respond preferentially to faces. Neuroimaging studies using both positron emission tomography (Sergent, Ohta, & MacDonald, 1992) and functional magnetic resonance imaging (fMRI; Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Gore, & McCarthy, 1995) reveal a small region in the fusiform gyrus of the ventral-temporal lobe that is more active when we view faces as compared to other objects. One interpretation of this finding is that this brain area, dubbed the "fusiform face area" or FFA (Kanwisher, McDermott, & Chun, 1997), is a face-specific neural module. That is, its function is to perceive or recognize faces and only faces. An alternative explanation is that this and other forms of putatively face-specific processing (e.g., Farah, 1990; Yin, 1969) are actually by-products of our extensive experience, which makes us face experts (Diamond & Carey, 1986). Thus, the recognition of individual faces exhibits qualities that should be true for any domain of visual expertise for a homogeneous object class. Faces are processed this way by because of their social importance, but not as a result of anything intrinsic to them as visual objects.

Gauthier and colleagues (Gauthier & Brown, 2004) have explored these competing accounts using several different approaches. Experts have been created in the laboratory for novel objects called "Greebles" in order to measure the observed changes in behavioral (Gauthier & Tarr, 1997) and neural activity (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Rossion et al., 2000) with expertise. Similar comparisons in both behavior and neural activity have been made between novices and real-world experts (Gauthier, Skudlarski, Gore, & Anderson, 2000; Righi & Tarr, 2004; Tanaka & Curran, 2001). Several findings speak directly to the question "Are faces special?" First, Greeble experts, but not Greeble novices, show behavioral effects—notably configural processing—that are often taken as markers for specialized face processing (Gauthier & Tarr, 1997; Gauthier, Williams, Tarr, & Tanaka, 1998). Second, Greeble experts, but not Greeble novices, show category-selectivity for Greebles in the right fusiform gyrus (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Similarly, bird experts show category-selectivity for birds, but not cars, in the right fusiform, whereas car experts show category-selectivity for cars, but not birds (Gauthier, Skudlarski, Gore, & Anderson, 2000). Reinforcing the generality of this result, chess experts, but not chess novices, show category-selectivity in right fusiform for valid, but not invalid, chess game boards (Righi & Tarr, 2004). Third, across Greeble expertise training, subjects show a significant positive correlation between a behavioral measure of holistic processing (sensitivity to the presence of the correct parts for that object) and neural activity in the right fusiform (Gauthier & Tarr, 2002). Similarly, bird and car experts show a significant correlation between their relative expertise measured behaviorally (birds minus cars) and neural activity in the right fusiform (Gauthier, Skudlarski, Gore, & Anderson, 2000), and years of experience playing chess correlates significantly with localized fusiform responses (Righi & Tarr, 2004). Fourth, the N170 potential (as measured by event-related potentials) shows face-like modulation in both Greeble (Rossion et al., 2000) and bird or dog experts (Tanaka & Curran, 2001).

These and other findings (e.g., Gauthier, Curby, Skudlarski, & Epstein, in press; Tarr & Gauthier, 2000) suggest that putatively face-specific effects may be obtained with nonface objects, but only when subjects are experts for the nonface-object domain. Thus, the answer to the question "Are faces special?" is yes and no. There is no doubt that faces are special in terms of their centrality to social interaction. On the one hand, it could be that this social importance necessitates built-in special-purpose brain circuits devoted to face recognition. But on the other hand, it could be that social importance has more indirect effects in that people develop expertise with faces from the repeated interactions with faces and the demands for individual-level recognition of faces. Some data supporting the latter argument come from studies using both Greeble and extant experts in domains as diverse as cars, birds, and chess. Across these domains, we find a pattern of behavioral and neural effects consistent with those seen for face recognition. In particular, category-selective activation in the fusiform gyrus has, of late, been taken as the hallmark of face specificity. Gauthier, Tarr, and others see similar selectivity for many other object domains, particularly when subjects are experts. Of course, this analysis only addresses the question of spatial specialization: "Is a particular piece of neural real estate dedicated to face processing?" (unlikely given current data) and raises the more meaningful question "What are the computational principles underlying processing in this brain region?" (we don't know at present).

Recent arguments based on finer resolution imaging or other methods for assessing spatial overlap between selective regions in the fusiform for faces and nonface objects miss this point (see http://web.mit.edu/bcs/nklab/expertise.shtml). From a theoretical perspective, even if convincing evidence existed that the microstructure of the brain regions recruited by faces and nonface objects of expertise were non- or partially overlapping, this would not demonstrate that these regions were functionally distinct. Indeed, good evidence already suggests that category-selective regions for different object categories are not functionally separable and that the representations of faces and different objects are both distributed and overlapping (Haxby et al., 2001). Moreover, adjacent, overlapping regions in visual cortex often show selective tuning for particular stimulus properties, but common underlying computational principles—one example being orientation columns in V1 (Kamitani & Tong, 2005). From an empirical point of view, the two studies addressing the question of overlap both used stimuli that were outside of the domain of expertise being tested, for example, antique cars shown to modern car experts (Grill-Spector, Knouf, & Kanwisher, 2004; Rhodes, Byatt, Michie, & Puce, 2004). Thus, it is unlikely that any strong effect of expertise could have ever been obtained under these conditions, let alone evaluated in terms of its relationship to face processing.

6.4 VISUAL LONG-TERM MEMORY

What do we remember about objects, what do we know about objects, and what do we do with objects? According to classic cognitive theories, memory, knowledge, and skills are abstract. Memory is poor because only the gist, particularly the semantic content, is retained. Knowledge is abstract because knowledge representations—abstract rules, schemas, or prototypes—are abstracted from experience (see Chapter 5 for a discussion of memory for gist in scene memory). Cognitive skills generalize (the sine qua non of skilled behavior is generalization) because they are not tied to any specific instances of prior skilled action. But, alternatively, in much the same way that invariant visual object perception can arise from specific views of objects, abstract memory, knowledge, and skills can arise from specific perceptual experiences (Barsalou, 1999; Palmeri, Wong, & Gauthier, 2004). Much of visual perception is based on the context provided by visual memories. Much of visual cognition is similarly grounded.

Contemporary neuropsychological theories often posit different kinds of memory (with lots of circles and arrows)—declarative versus procedure, episodic versus semantic, perceptual versus habit, explicit versus implicit, and other such dichotomies—that are subserved by functionally independent systems (Fig. 6–9). As with the study of category-selectivity, neuropsychology and neuroimaging methods play naturally to this sort of (relatively simplistic) modular theorizing, providing evidence for system-specific, dedicated brain regions, each with a unique set of representational and processing assumptions.

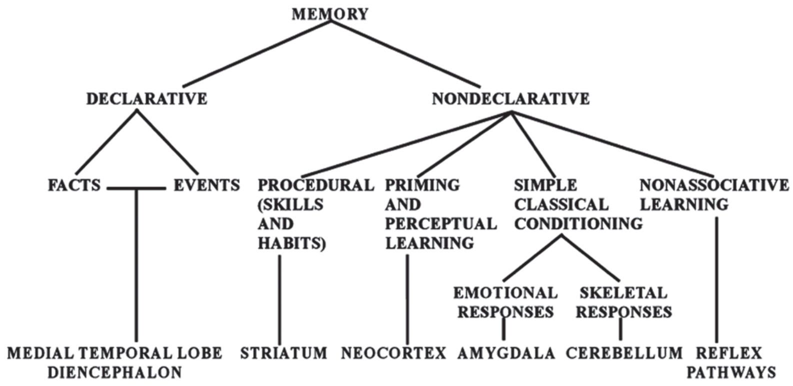

Figure 6–9. A well-known taxonomic hierarchy of long-term memory systems (from Squire, 2004). The taxonomy divides memory into conscious declarative (hippocampal-dependent) memory and unconscious nondeclarative (nonhippocampal-dependent) memory. There is little debate that particular brain structures are especially important for particular kinds of memory. Debate centers around whether those brain structures are important because those structures are the systems responsible for particular kinds of memory tasks or because those structures carry out processing that is especially important for certain kinds of memory tasks under certain conditions. Adapted with permission from Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, USA*. 93, 13515–13522.

Much in the same way that functional specialization in visual cortex may be explained by alternative organizational principles, functional specialization for visual memory and learning may be organized around the kinds of representations and processes recruited by particular tasks and not around the tasks themselves (Palmeri & Flanery, 2002; Roediger, Buckner, & McDermott, 1999).

6.5 VISUAL MEMORY FOR OBJECTS

Our everyday experience leads us to the conclusion that our visual memory seems quite poor. Rare cases of eidetic memory aside (Luria, 1987), most people feel that they have great difficulty remembering visual details, even over relatively short periods of time. Supporting this intuition, some early experimental studies of long-term visual memory for objects and scenes suggested that people remember only the gist (e.g., Brewer & Treyens, 1981), but not the specific details (but see Shepard, 1967). This conclusion is reinforced by recent "change blindness" studies that suggest that we remember very little of what we see from one moment to the next (but see Hollingworth, in press; Rensink, O'Regan, & Clark, 1997; Simons & Rensink, 2005; see Chapter 5 of this volume for a detailed discussion of the change blindness literature and its theoretical implications).

Such limitations are especially apparent in eyewitness testimony, in which witnesses are notoriously bad at recognizing or recalling visual details, and their visual memories for events can be significantly influenced by nonvisual, interfering information (Loftus, 2004; Wells & Loftus, 2003; Wells & Olson, 2003). These and related results have led some researchers to conclude that memory for objects and scenes is not perceptual, but instead reflects a semantic recoding of perceptual information guided by abstract schematic knowledge (e.g., Fodor, 1975; Newell & Simon, 1972; see Barsalou, 1999). Indeed, the ability of people to form visual images has either been rejected outright or has been characterized as an epiphenomenon of cognitive processing in which images arise as a by-product of accessing inherently nonvisual memories (Anderson, 1978; Pylyshyn, 1973, 1981).

Happily, the field does lurch forward over time. Many contemporary memory models assume distinct episodic representations that can, in principle, retain detailed information about specific perceptual experiences (Hintzman, 1986; Logan, 1988; Nosofsky, 1991). Why, then, does memory often seem so poor? And why do we seem to remember only the gist and not the details? After all, memory performance should be related to how well information is encoded, stored, retrieved, and used (see Chapter 5). In principle, the visual detail of visual memories is constrained at the upper limit by the visual details provided by the perceptual system. Clearly, if some of the visual information is not processed during a perceptual episode, then that information cannot be encoded into an enduring memory representation to be retrieved later. Both visual attention and eye movements can conspire to render part of an object or scene invisible or poorly visible and, hence, absent from memory (Rensink, O'Regan, & Clark, 1997; Simons, 1996; Simons & Rensink, 2005). In addition, some visual information may be normalized or explained away (Kersten, Mamassian, & Yuille, 2004) during visual processing, rendering those visual details for all intents and purposes invisible to the perceiver. If it's not perceived, it can't be remembered. If it's perceived poorly, it can be remembered poorly at best.

At the same time, many theories of memory assume that some probability is associated with whether particular visual properties are encoded into memory, accurately or at all (Hintzman, 1986; Hintzman, 1988; Shiffrin & Steyvers, 1997). Thus, failures to remember can also arise from failures to encode. The ability to remember visual details is also limited by how well the information is retained in memory. Although some theories have attributed memory failures to factors other than memory storage (Gillund & Shiffrin, 1984), many theories also assume that memory traces can change rather dramatically because of decay (Hintzman, 1986) or some form of consolidation (McClelland, McNaughton, & O'Reilly, 1995; Shiffrin & Steyvers, 1997) over time. At one extreme, individual memory traces remain highly distinct from one another (Gillund & Shiffrin, 1984; Hintzman, 1986; Raaijmakers & Shiffrin, 1981), and storage failures are due to memory decay. At the other extreme, all memories share largely the same representational substrate, which results in similar memory representations becoming physically and informationally indiscriminable from one another (e.g., McClelland & Rumelhart, 1985). In this case, storage failures are due to interference from new memories. Both kinds of memory may exist, with the hippocampus and associated structures especially involved in distinct memories for particular visual episodes and cortical areas maintaining more generalized memories in a distributed fashion (e.g., O'Reilly & Norman, 2002).

Finally, failures of visual memory may also emerge due to failures of memory retrieval. Visual details may be visible, they may be successfully encoded into memory, and they may be retained in memory. But retrieval cues may be insufficient to retrieve the relevant visual memory representations. Retrieval failure is probably one of the primary reasons for memory failure. Consider that failing to provide the right retrieval cues is a failure to reinstantiate the context with which the memories were first encoded. Most explicit recognition or recall tasks do not ask subjects to report whether they have *ever* seen that object before. Rather, they typically require someone to say whether or not they saw a particular object during some initial encoding session—whether earlier that day, a week earlier, or a year earlier—while tested in the same context or a different context (Mensink & Raaijmakers, 1988; Shiffrin & Steyvers, 1997). Both external context cues (the room, the time, the experimenter, and the like) and internal context cues (associations with other studied items) are needed to discriminate old studied objects from new lures.

Apparent visual memory failures during retrieval can arise based on the way memory retrieval takes place. A large class of memory models assumes that explicitly recognizing an old object as one you have seen before could arise from retrieving a specific memory that matches the probe or from a global familiarity based on the match between a retrieval probe and all memory representations. According this view, we remember the gist not because it's the gist that is stored, but rather because a probe cue retrieves a number of matching memories that are combined together (Hintzman, 1988; Shiffrin & Steyvers, 1997). That is, seemingly abstract memories are produced online during the act of memory retrieval (Barsalou, 1990, 1999) because what is retrieved is a blend of memories. What is common between these memories is what we would typically call the "gist"—those visual properties present across many instances stored in memory.

6.5.1 Explicit Versus Implicit Visual Memory

If contemporary memory research has taught us one thing, it is that memory is far more than explicitly recognizing or recalling past experiences. With our every action, we reveal memory through our performance. Perhaps the most well-known experimental example of this is perceptual repetition priming. People are faster and more accurate at identifying a visual object they have seen before. Repetition priming is highly specific to perceptual details (e.g., see Schacter, Chiu, & Ochsner, 1993), it has been reported for delays of over a year in the absence of explicit recollection (Cave, 1997; Kolers, 1976), and it is normal in individuals with explicit memory deficits (e.g., Squire, 1992). Such results have led researchers to divide memory into functionally independent memory systems for explicit memory and implicit memory (Schacter & Tulving, 1994). For example, Tulving and Schacter (1990) proposed "Perceptual Representation Systems" that stored perceptual memories to support repetition priming. Primarily based on reports of various neuropsychological dissociations between memory tasks, the number of functionally independent memory systems has ballooned to include separate systems for episodic memory, semantic memory, habit learning, perceptual learning, conditioning, and host of other memory tasks (Schacter, 2000; Squire & Zola, 1996; Squire, 2004).

The multiple memory systems approach delineates the many different independent memory systems, often tying such systems to particular brain structures, and providing, perhaps, some rationale for why evolution could have favored these particular divisions and not others. However, specific versions of this approach typically omit much discussion regarding the specific mechanisms underlying the myriad component memory systems. How are memories encoded, how are they represented, and what processes can be brought to bear on them?

Interestingly, many of these weaknesses parallel those seen in modularist approaches to visual object recognition (Palmeri & Gauthier, 2004). For example, much in the same way that modularist approaches to visual perception assign independent systems to particular kinds of objects, multiple memory systems approaches assign independent systems to particular kinds of tasks without explaining how they accomplish said tasks. The hippocampus is for explicit declarative memory used to recognize or recall specific visual experiences. The basal ganglia is for learning skills that might associate an object with a well-learned response. Bits of visual cortex are for perceptual learning that leads to priming. Although at first approximations these assertions are indisputable based on the neuropsychological and brain imaging evidence, saying that the hippocampus is a necessary neural substrate for explicit declarative memory is different from saying that it creates explicit declarative memories and explicit declarative memories only (Squire, Stark, & Clark, 2004). Indeed, the hippocampus appears to be involved more generally in creating configural representations (Chun & Phelps, 1999; Cohen & Eichenbaum, 1993; Meeter, Myers, & Gluck, 2005) or consolidating memories in cortical representations (O'Reilly & Rudy, 2001). That is, its role in conscious declarative memory may be a useful by-product of its more general role in binding together cues from multiple modalities into a single representation. The importance of representations created by the hippocampus is clearly measured in explicit declarative memory tasks, such as recall or recognition. Thus, explicit declarative memory deficits are most conspicuous in amnesics with hippocampal damage. At the same time, the role of the hippocampus can also be gauged by performance in appropriately designed implicit memory tasks (see Chapter 7). Based on such evidence, contemporary theories of memory associate specific representations and process roles to a network of interdependent memory systems, rather than assuming separable systems that are tied to particular tasks (e.g., see Meeter, Myers, & Gluck, 2005).

Returning to perceptual repetition priming as an important example of implicit visual memory, what are some alternative explanations for enhanced object recognition as a result of prior perceptual experience? Tulving and Schacter (1990) proposed that priming was mediated by a perceptual representation system independent of the memory system that supports explicit declarative memory. Previous experiences are stored in this memory module in order to prime later experiences. Why? Priming has some adaptive value, so we should store perceptual memories for the purposes of enhancing perceptual performance at later encounters. Alternative theories do not place such a clear demarcation between memories underlying priming and memories used for other purposes. For example, Rouder, Ratcliff, and McKoon (2000) proposed that priming effects are a by-product of how view-based memories are used during normal object recognition. In their model, priming is caused by a bias to interpret perceptual information as supporting familiar objects, as opposed to unfamiliar objects. They instantiated this hypothesis by simply adding biases to Poggio and Edelman's (1990) simple view-based model of object recognition. The added biases simply reflect the learned likelihood of seeing a given object again, thereby causing known objects to be identified more quickly when they are seen again later. Adding this simple psychological mechanism accounts for perceptual priming without the need to posit any additional implicit memory system above and beyond the use of object representations already incorporated into almost all models of object recognition. Again, memory *is* perception.

Some more standard memory theories also posit that perceptual priming and other forms of implicit memory are a normal by-product of memory. In these models, a contrast is often drawn between the visual features of represented objects and the features associated with its context. That is, explicit memory tasks are contextualized judgments about whether a given object appeared at a particular location at a specific time, not whether that object has ever been seen before (Gillund & Shiffrin, 1984; Hintzman, 1988). Loss of this contextual information, or a failure to encode such information due to brain damage, would result in failures of explicit memory with preserved implicit memory. Moreover, a subset of these theories have hypothesized that, over time, memory traces can become decontextualized through memory consolidation (Shiffrin & Steyvers, 1997). Put another way, they are not transferred to a different memory store, but become dissociated from the context under which they were learned as a natural product of how memory storage works.

Consistent with a more integrated approach, a number of computational memory models make no clear demarcation between memories for general semantic knowledge, memories for particular experiences, and implicit memories. At the core, the same visual memories are used to recall, recognize, categorize, identify, or do things with objects (Logan, 2002; Nosofsky, 1992). For example, repetition priming may rely on the same memories that underlie cognitive skills (Logan, 1990). At first blush, single-system memory models of this sort may seem rather out of touch with contemporary cognitive neuroscience research on memory, in which the *desiderata* seems to be as many distinct systems as possible. Yet, a recent fMRI study supports the view that important aspects of perceptual priming and skill learning may share similar neural loci. Specifically, Dobbins et al. (Dobbins, Schnyer, Verfaellie, & Schacter, 2004) contrasted a perceptual locus for repetition priming with a high-level response learning locus for repetition priming. They found that prefrontal cortical activity tracked repetition priming behavior, not activity in visual cortex. Thus, as suggested by Rouder et al. (2000), repetition priming effects may not reflect the creation of new perceptual representations, or even the short-term tuning of perceptual representations, but may instead reflect a bias to do things with objects that we did with them before, and to do so more quickly (Logan, 1990).

Perceptual Categorization and Visual Knowledge of Objects

As discussed earlier, one hallmark of visual cognition is generalization. Even very young children seem to know when two visually similar but different objects are members of the same category. One solution to the problem of generalizing from specific experiences is to create knowledge representations that are themselves abstract generalizations. According to early theories, conceptual knowledge is organized into abstract semantic networks or conceptual hierarchies (Anderson, 1976; Collins & Quillian, 1969) that link one kind of thing with another kind of thing through propositional structures. Knowledge is stored efficiently, so that object properties that are true of a superordinate category of objects are only stored at the most general level (and are inherited as needed). Only properties that are unique to subordinate categories or specific individuals are encoded at lower levels of the conceptual hierarchy (Smith, Shoben, & Rips, 1974). In this way, what we know about objects is abstracted away from our perceptual experiences. As such, objects are categorized as different kinds of things using abstract logical rules (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Bruner, Goodnow, & Austin, 1956; Johansen & Palmeri, 2002; Nosofsky, Palmeri, & McKinley, 1994) or by comparing an object with an abstract prototype or schema (Lakoff, 1987; Minda & Smith, 2001; Posner & Keele, 1968; Rosch, 1975). That is, class invariance is achieved through representations that are invariant over members of that class.

A sharp distinction between memory for specific visual experiences and abstract visual knowledge is also manifest in the classic distinction between semantic and episodic memory (see Squire & Schacter, 2002; Tulving, 1985, 1993). Some memory researchers have argued that good computational reasons exist for keeping specific memories separate from abstract knowledge. After all, if all we have are specific memories for particular objects, how could we ever know anything general that was true about members of a class? And, if all we have is general knowledge, how could we ever know anything about specific objects?

This approach is generally similar to Biederman's (1987) theory in that RBC proposes that both view and class invariance are achieved by constructing representations—prototypes—that are themselves invariant over views and class exemplars. Interestingly, in the same way that theorists have argued that we do not have need for view- and/or class-invariant object representations to attain view and class invariant object recognition (Bülthoff & Edelman, 1992; Poggio & Edelman, 1990; Tarr & Pinker, 1989), we may not need class-invariant category representations to achieve a basic-level classification over object categories (Palmeri, 1999; see Palmeri & Gauthier, 2004).

To be clear, visual memories for specific experiences with objects can support *both* the recognition of particular objects and general knowledge about classes of objects. Direct abstraction is not needed. What matters is how specific exemplar memories are used relative to one another. As noted earlier, explicit recollection is typically a contextualized decision. Did you see this object on that occasion? Explicit recognition uses memory retrieval cues that contain both information about the object and its context. Explicit recall uses retrieval cues with context alone. Thus, any inability to reinstantiate the original context, to encode the context, or a loss of information about the context in memory will lead to degradation of explicit memory. In contrast, questions about visual knowledge are decontextualized. An object is a member of a particular category across most, if not all, contexts. And, an object is associated with other objects, has particular properties or elicits certain behaviors across many different contexts. Thus, explicit memory for objects requires the integrity of particular visual memories, whereas classification of objects or general knowledge of objects, including object recognition, can utilize a panoply of visual memories across a variety of visual contexts.

Two key properties of these memory models are: (a) retrieval is similarity-based, and (b) decisions are based on the retrieval of multiple visual memory traces. Additionally, LTM is probed with a retrieval cue tailored to the particular memory task. For example, the retrieval cue for a recognition memory task would include features of the object and features of the context, whereas the retrieval cue for a categorization task could often include features of the object only. Memory traces are activated according to the similarity between the retrieval cue and the trace (Gillund & Shiffrin, 1984; Hintzman, 1986; Nosofsky, 1992). Using a process of decisional selective attention, matches or mismatches of certain visual features may be weighed more heavily if they are particularly diagnostic for the decision being made (Kruschke, 1992; Lamberts, 1998, 2000; Logan, 2002; Nosofsky, 1984, 1986). Because retrieval is similarity-based, objects that have never been seen before can be falsely recognized as previously seen objects during recognition if they are similar to studied objects (Nosofsky, 1991). Similarly, objects that are prototypical will be quickly and accurately categorized as category members because they are similar to many other category examples, whether or not they have been studied before (Busemeyer, Dewey, & Medin, 1984; Hintzman, 1986; Shin & Nosofsky, 1992). Of course, the decision rules underlying categorization and recognition decisions are different. Recognition is an absolute judgment of whether an object is sufficiently similar to objects that have been studied before. If so, the object is recognized as old. Categorization is a relative judgment about an object's similarity to known categories. As such, although recognition and categorization may depend on the same underlying memories, recognition and categorization judgments need not be tightly correlated (Nosofsky, 1992). Thus, your ability or inability to recognize a familiar object does not predict whether you will be able to categorize it correctly. Critically, this stochastic independence between recognition and categorization does not imply that these two processes are based on different memory systems, rather it is equally probable that they are simply based on different decisions rules.

One challenge to this unified approach comes from reported neuropsychological dissociations between recognition and categorization (Squire & Zola, 1996). Specifically, amnesics can learn novel visual categories but are impaired (Knowlton & Squire, 1993) or at chance (Squire & Knowlton, 1995) at recognition. At first blush, this seems like clear evidence in support of functionally independent memory systems for visual recognition memory and visual categorization. However, simply by assuming that amnesics have poorly discriminated memories relative to normal controls, perhaps because of failures of the damaged hippocampus to create new configural representations or to consolidate memories into cortical areas, models assuming the same memories for categorization and recognition predict the observed dissociation a priori (Kinder & Shanks, 2001; Nosofsky & Zaki, 1998; Palmeri & Flanery, 2002). To elaborate, most categorization tasks require broad generalization from learned examples to test examples, whereas recognition requires fine discrimination between old and new test items. Recognition is influenced significantly more by degradation than categorization (Palmeri & Flanery, 2002). On top of this, many of the perceptual categorization tasks that have been used in the neuropsychological literature may not rely on long-term memories for trained category exemplars whatsoever. Specifically, a number of the categorization tests that have been used to assess long-term category memory can be performed just as well whether people have studied category exemplars or not (Palmeri & Flanery, 1999; Zaki & Nosofsky, 2001, 2004). When these methodological flaws are addressed, individuals with memory impairments may indeed learn novel categories less well than normal controls (Zaki, Nosofsky, Jessup, & Unversagt, 2003). Although it is possible that visual memories supporting recognition and visual memories supporting categorization are functionally independent of one another, a more computationally tractable solution is that the same visual memory representations underlie a variety of memory tasks, with the information requirements of the particular task modulating performance.

6.5.3 Levels of Categorization

Objects can be categorized at multiple levels of abstraction, from identifying unique individuals by name to grouping together dissimilar objects as the same kind of thing. Discriminating between highly similar objects for purposes of identification and generalizing across many different objects for purposes of categorization appear to be competing goals that require different kinds of visual representations (Biederman, 1987; Logothetis & Sheinberg, 1996;

Marr, 1982). In fact, some structural-description theories (e.g., Biederman, 1987) view basic-level classification as a primary goal of visual perception. In such theories, the structural descriptions for different members of the same basic-level category are the same. At the same time, visual memories of a qualitatively different sort are needed in order to discriminate similar objects for purposes of identification or more subordinate levels of classification (Biederman, 1987). Such is the logic often used in arguing for dissociations between face and nonface object recognition (e.g., Farah, Wilson, Drain, & Tanaka, 1998).

Along similar lines, evidence from speeded perceptual categorization tasks have been used as evidence that basic-level classification is a stage of processing that precedes more subordinate or superordinate classification because basic-level classification is significantly faster (Grill-Spector & Kanwisher, N., 2003; Jolicoeur, Gluck, & Kosslyn, 1984). Typically, the fastest categorization task is basic-level categorization and is termed the "entry level" (Jolicoeur, Gluck, & Kosslyn, 1984) into conceptual knowledge. However, the entry level for a given individual can vary greatly with experience (Johnson & Mervis, 1997; Tanaka & Taylor, 1991). At the same time, fastest does not mean first (Palmeri, Wong, & Gauthier, 2004). In fact, a number of computational models of object recognition and perceptual categorization make no clear demarcation between identifying unique objects and categorizing objects as members of a class (Nosofsky, 1992; Riesenhuber & Poggio, 1999, 2000; Tjan, 2001). Specifically, identification and categorization are both evidence-based perceptual decisions. Identification may require more perceptual processing (Lamberts, 2000), but prior categorization is not necessary. Thus, the same visual memories that support invariance across changes in the image, for example, as generated by rotations in depth, also support access to objects at multiple levels of categorization, for example, recognition memory, identification, and categorization (Edelman, 1999). What varies is not our memories, but how such memories are used to make perceptual decisions that change from one task to another (Palmeri & Gauthier, 2004) (Fig. 6–10).

It should be emphasized that invariant performance can emerge from memory representations that do not themselves embody that invariance. For instance, viewpoint-invariant object recognition is enabled by comparing percepts with views in memory. Class-invariant object recognition is enabled by comparing an object with category exemplars in memory. Beyond the point at which abstract memory representations are not needed, it behooves us to spell out the mechanisms by which said invariances are achieved. For example, do we remember *all* views of objects we encounter, or is generalization good enough to allow encoding of only salient (Tarr & Kriegman, 2001) or frequently experienced views (Blanz, Tarr, & Bülthoff, 1999; Palmer, Rosch, & Chase, 1981)? Similarly, are all exemplars of all categories encoded, or are there "key" exemplars that help to delineate a given category? Although precise estimates of the capacity of visual memory are impossible, a surprising amount of detailed perceptual information may be encoded into visual memory (Hollingworth, 2004, 2005; Standing, 1973). And such memories may persist and influence

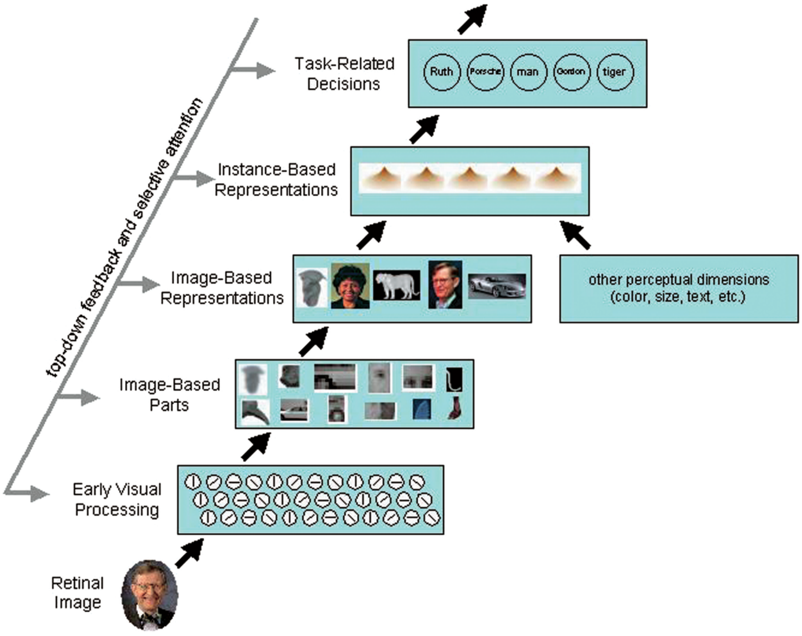

Figure 6–10. This illustration summarizes some of the elements of a class of image-based/instance-based/exemplar-based models of object recognition and perceptual categorization. Starting with the retinal image, a hierarchy of steps from low-level visual processing, to representations in terms of image-based parts (Ullman, Vidal-Naquet, & Sali, 2002), to representations of views and instances of objects (Riesenhuber & Poggio, 2000). The same representations of views and instances can be associated with perceptual decisions like a basic-level category (car or face), subordinate-level category (Honda Civic), or identity (Gordon Gee). An important component of many models is that selective attention can highlight aspects of a perceptual representation that are particularly diagnostic for a decision (e. g., Ahissar & Hochstein, 2004; Kruschke, 1992).

behavior for a long time period. One might argue that there is adaptive significance for the ability to encode a large amount of information in memory. This allows abstraction on-the-fly rather than requiring a prescient gatekeeper to decide what information might be necessary for survival at some later point in time (Barsalou, 1990). Along the same lines, instance-based models of object recognition and categorization have often been mathematically formalized, assuming that every view or every exemplar is stored in visual memory (Logan, 1988; Nosofsky, 1992). However, as discussed earlier, a more sparse encoding of views and exemplars that provides a nearly full but incomplete covering of the space of experienced instances supports recognition and categorization across changes in the image and exemplars within an object class (e.g., Bülthoff & Edelman, 1992; Rosseel, 2002). Moreover, detailed visual memories of specific perceptual experiences may support tasks beyond visual recognition and categorization. In particular, abstract conceptual knowledge that may appear amodal and abstracted from actual experience may in fact be grounded in perceptual knowledge (Barsalou, 1999; Martin, Ungerleider, & Haxby, 2000). In much the same way that abstractions of visual properties may be created on-the-fly, abstract conceptual properties may be "revealed" by mental simulations of perceptual knowledge (Barsalou, 1990).

Other isomorphisms seem to exist between visual and conceptual memories. For instance, in much the same way that evidence may differentially accumulate over time when recognizing particular views of an object (Perrett, Oram, & Ashbridge, 1998), evidence may differentially accumulate over time when categorizing different exemplars of a category. Nosofsky and Palmeri (1997) proposed a model of speeded categorization that combined representational elements of exemplar-based models of categorization and memory (Nosofsky, 1992), temporal assumptions of an instance-based model of automaticity (Logan, 1988), and stochastic evidence accumulation from random-walk models of perceptual decisions (Link, 1975; Luce, 1986; Ratcliff, 1978; Ratcliff & Rouder, 1998). In such models, the rate of accumulation depends on the similarity between the object to be categorized and the specific category exemplars in memory (Fig. 6–6). An object similar to many exemplars of a single category will be classified quickly and accurately. An object dissimilar to category exemplars will be classified more slowly, but perhaps accurately if only relatively similar to a single category. But an object similar to exemplars of different categories will be classified slowly and inaccurately because of the contradictory evidence.

6.5.4 Perceptual Expertise

Radiologists, ornithologists, firefighters, and other specialists are noted for their remarkable abilities at categorizing, identifying, and recognizing objects within their domain of expertise (Palmeri, Wong, & Gauthier, 2004). But understanding perceptual expertise is more than characterizing the behavior of individuals with idiosyncratic skills in highly specialized domains. Perceptual expertise may also explain some of the unique aspects of recognizing faces (Diamond & Carey, 1986; Gauthier & Tarr, 2002), words (McCandliss, Cohen, & Dehaene, 2003), and letters (McCandliss, Cohen, & Dehaene, 2003). The development of perceptual expertise involves a complex interplay of changes in visual perception, visual memory, visual categorization, and visual skills. Indeed, viewing perceptual expertise as the end-point of the normal learning trajectory, rather than an idiosyncratic skill, allows us to exploit studies of perceptual experts to understand the general principles as well as the limits of visual perception, memory, and learning.

Experts are fast (Tanaka & Taylor, 1991). They make fine perceptual discriminations and precise identifications with speeds that can astonish the novice observer. Experts also perceive differently within their domain of expertise (e.g., see Gauthier, Curran, Curby, & Collins, 2003; Goldstone, 2003; Myles-Worsley, Johnston, & Simons, 1988; Snowden, Davies, & Roling, 2000);

that is, "holistically" and more efficiently extracting discriminating information. What makes experts so fast? And what makes them apparently perceive objects in their domain so differently from novices? Have they developed qualitatively different ways of processing information? Have they created new ways of representing information? Or, have they discovered optimal ways of using the representations they had as novices (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Joyce & Cottrell, 2004; Palmeri, Wong, & Gauthier, 2004)?

As discussed in the previous section, novices categorize objects fastest at a basic (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) or entry level (Jolicoeur, Gluck, & Kosslyn, 1984). Expertise has been characterized as establishing a new entry level, such that objects are categorized as quickly at more subordinate levels (Johnson & Mervis, 1997; Tanaka & Taylor, 1991). But what does it mean to establish a new entry level? It could mean creating a new special-purpose module or perceptual routine (Jolicoeur, Gluck, & Kosslyn, 1984) for expert categorization at these subordinate levels. Positing the creation of such a module would indeed account for the automaticity, domain-specificity, and attentional inflexibility seen with perceptual experts, and it would link the development of perceptual expertise in novel domains with the purported modularity in domains such as face and letter recognition. Of course, this account begs questions of how a new module might be created, how this module might operate, and whether such a module is computationally necessary at all (Riesenhuber & Poggio, 1999, 2000; Tarr & Cheng, 2003). We have instead approached an understanding of the development of perceptual expertise by viewing it as the end-point of normal learning that underlies recognizing, categorizing, and remembering objects (Gauthier & Tarr, 2002; Joyce & Cottrell, 2004; Palmeri, Wong, & Gauthier, 2004), attempting to characterize the development of perceptual expertise within theories of normal object recognition, categorization, and memory.

An important step in becoming a perceptual expert is learning what aspects of an object class are relevant for a perceptual identification. This learning can get a head start when someone is given an explicit rule (Noelle & Cottrell, 1996). Such rules specify which features are important, explicitly guiding dimensional selective attention (Johansen & Palmeri, 2002), as well as how to combine this information to make a decision. Even when explicit rules are not provided, observers induce simple rules on their own (Ashby, Queller, & Berretty, 1999; Nosofsky & Palmeri, 1998; Nosofsky, Palmeri, & McKinley, 1994; Waldron & Ashby, 2001). But of course, in some domains, verbal labels cannot adequately convey the diagnostic perceptual qualities for the novice, making any explicit instruction a futile enterprise.

Moreover, the use of explicit rules does not seem to characterize expert behavior (Brooks, Norman, & Allen, 1991). Experts may or may not be able to articulate explicit rules to a novice—although in some cases they may simply "know" what things are and may be entirely unaware of whether there might exist a simple rule (Biederman & Shiffrar, 1987)—but they do not seem to use these rules, especially for making their rapid initial perceptual identifications. Perceptual experts make decisions automatically and implicitly. Taking the instance theory of automaticity as a theoretical starting point (Logan, 1988), we argue that this automaticity is largely grounded in the vast perceptual memories experts have acquired (Palmeri, 1997; Palmeri, Wong, & Gauthier, 2004). Experts are fast because memory retrieval is fast. Experts decisions are automatic because memory retrieval is automatic. Experts make difficult perceptual discriminations easily because they have performed similar perceptual discriminations before. Experts show relatively limited generalization because memory retrieval is based on perceptual similarity (Gauthier & Tarr, 1997; Palmeri, 1997; Tanaka, Curran, & Sheinberg, 2005). The development of expertise often entails a shift from rules to visual memories because memory retrieval becomes more efficient than rule use as perceptual memories are strengthened over learning (Johansen & Palmeri, 2002; Palmeri, 1997, 2001; Rickard, 1997).

These perceptual memories should not be confused with simple templates for several reasons. First, retrieval is similarity-based, allowing generalization to novel objects (Poggio & Bizzi, 2004; Shepard, 1987). Second, decisions are not based on retrieving a single perceptual memory, but on retrieving an ensemble of similar perceptual memories (Gauthier & Palmeri, 2002; Poggio & Bizzi, 2004). Third, retrieval of these perceptual memories is not based on raw similarity, but selective attention mechanisms serve to weight diagnostic dimensions over nondiagnostic dimensions in determining similarity (Kruschke, 1992; Lamberts, 1998; Nosofsky, 1984, 1986).

As stated earlier, initial stages of learning involve figuring out which parts of objects are more important than others for making perceptual identification and categorizations. Although this learning can get a boost from an explicit rule that might be supplied (Medin & Smith, 1981; Palmeri & Nosofsky, 1995), more fine-tuned learning involves more implicit trial-to-trial adjustment of selective attention to particular dimensions (Gauthier & Palmeri, 2002; Kruschke, 1992; Lee & Navarro, 2002; Nosofsky, Gluck, Palmeri, McKinley, & et al., 1994; Nosofsky & Kruschke, 1992). Indeed, learning to selectively attend to the right representations may characterize a significant amount of perceptual expertise and perceptual learning (Dosher & Lu, 1999; Petrov, Dosher, & Lu, 2005). Ahissar and Hochstein (2004) argued that "what typically limits naïve performance is the accessibility of task-relevant information rather than the absence of such information within neuronal representations." According to their theory, this selection process works from the top down, so that easy-to-learn problems are those that require selecting relatively high-level representations whereas difficult problems require selecting low-level representations. In this context, many classic category-learning problems are easy tasks (in that they can be learned in a few hundred training trials) because they require learning to selectively attend to highly salient parts or dimensions of an object. By contrast, many perceptual learning problems are hard tasks (in that they require many days of training) because they may require learning to selectively attend to visual processing channels early in the visual stream (Petrov, Dosher, & Lu). Although, selective attention mechanisms have been given a short shrift in some discussions of perceptual expertise and perceptual learning (Schyns, Goldstone, & Thibaut, 1998), they likely play a critical role, especially during initial learning (see also de Beeck, Wagemans, & Vogels, 2003; Palmeri, 1998; Petrov, Dosher, & Lu, 2005).

That said, the development of perceptual expertise requires creating new representations as well as selecting from existing representations. Of course, forming perceptual memories is creating new representations. The initial creation of these exemplar memories may be mediated by hippocampal areas thought to be involved in creating novel configural representations (Gluck, Meeter, & Myers, 2003; Meeter, Myers, & Gluck, 2005; O'Reilly & Norman, 2002). At the same time, exemplar memories may be insufficient to explain all of the perceptual effects manifest in perceptual experts (Palmeri & Gauthier, 2004; Palmeri, Wong, & Gauthier, 2004). Instead, an important aspect of the development of perceptual expertise may also involve the creation of new image-based part representations (Zhang & Cottrell, 2005). These image-based parts can support the perception, categorization, and memory for learned objects, but can also efficiently support perception and memory for new objects (Gauthier & Tarr, 2002). But creating a perceptual expert can take a long time (Gauthier, Williams, Tarr, & Tanaka, 1998). To the extent that this lower-level learning involves cortical updating, this learning will be far slower than the kind of rapid memory formation seen for particular exemplars. Mirroring the top-down progression for selective attention posited by Ahissar and Hochstein (2004), the creation of new representations likely takes place in a top-down manner, with exemplar memories being formed rapidly but image-based part memories taking more time. A critical aspect of learning probably involves creating the right perceptual building blocks, but arguably creating those building blocks may require a great deal of training (but cf. Schyns, Goldstone, & Thibaut, 1998).

6.6 CONCLUSION

Our goal in this chapter was to review and synthesize recent thinking in two domains: visual object perception and visual memory. For whatever reasons, as with much of the larger discipline of cognitive science, there has been a tendency toward compartmentalization. It is almost as if the field's conceptualization of the mind and brain as a collection of modular processing systems is reflected in how the field itself has become organized. Unfortunately, such divisions are often more matters of convenience (both in creating theory and in choosing our domains of study). As such, it is important to consider how nominally separable processes relate to one another. In the case of memory and object perception, this examination is more than cursory. We argue that memory and perception are intimately related and in essence two sides of the same coin. That is, memory arises as a consequence of object perception and, conversely, object recognition tasks are effectively memorial processes.

ACKNOWLEDGMENTS

This work was partially supported by a grant from the James S. McDonnell Foundation, NSF grant HSD-DHBS05, and by the Temporal Dynamics of Learning Center (NSF Science of Learning Center SBE-0542013).

REFERENCES

Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences* 8:457–464.

Anderson, J. R. (1976). *Language, Memory, and Thought*. Hillsdale, NJ: Erlbaum.

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

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

Ashby, F. G., Queller, S., and Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception and Psychophysics* 61:1178–1199.

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

Atkinson, R. C., and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *The Psychology of Learning and Memory, Vol. 2*, eds. K. W. Spence and J. T. Spence. New York: Academic Press.

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In *Content and Process Specificity in the Effects of Prior Experiences*, eds. T. K. Srull and R. S. Wyer, Jr., 61–88. Hillsdale, NJ: Lawrence Erlbaum.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences* 22:577–660.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences* 7:84–91.

Barenholtz, E., and Tarr, M. J. (2007). Reconsidering the role of structure in vision. In *Categories in Use. Series: The Psychology of Learning and Motivation*, Vol. 47, eds. A. Markman and B. Ross, 157–180. San Diego: Academic Press.

Barenholtz, E., and Tarr, M. J. (submitted). Visual judgment of similarity across shape transformations.

Beauregard, M., and Paquette, V. (2006). Neural correlates of a mystical experience in Carmelite nuns. *Neuroscience Letters* 405:186–190.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review* 94:115–147.

Biederman, I., and Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance* 19:1162–1182.

Biederman, I., and Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:640–645.

Bienenstock, E., and Geman, S. (1995). Compositionality in Neural Systems. In *The Handbook of Brain Theory and Neural Networks*, ed. M. A. Arbib, 223–226. Cambridge, MA: MIT Press.

Blanz, V., Tarr, M. J., and Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception* 28:575–600.

Boucher, L., Palmeri, T. J., Logan, G. D., and Schall, J. D. (2006). Interactive race model of countermanding saccades. Manuscript under review.

Brewer, W. F., and Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology* 13:207–230.

Brooks, L. R., Norman, G. R., and Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General* 120:278–287.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. (1956). *A Study of Thinking*. Oxford, England: Wiley.

Bülthoff, H. H., and Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences* 89:60–64.

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

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science* 8:322–325.

Chun, M. M., and Phelps, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience* 2:844–847.

Cohen, N.J., Eichenbaum, H. (1993). *Memory, Amnesia, and the Hippocampal System*. Cambridge, MIT Press.

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

de Beeck, H. O., Wagemans, J., and 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* 132:491–511.

Diamond, R., and Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General* 115:107–117.

Dobbins, I. G., Schnyer, D. M., Verfaellie, M., and Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature* 428:316–319.

Dosher, B. A., and Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research* 39:3197–3221.

Edelman, S. (1995). Representation, similarity, and the chorus of prototypes. *Minds and Machines* 5:45–68.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Science* 1:296–304.

Edelman, S. (1999). *Representation and Recognition in Vision*. Cambridge, MA: MIT Press.

Edelman, S., and Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research* 32:2385–2400.

Farah, M. J. (1990). *Visual Agnosia: Disorders of Object Recognition and What They Tell Us about Normal Vision*. Cambridge, MA: MIT Press.

Farah, M. J., Wilson, K. D., Drain, M., and Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review* 105:482–498.

Fisher, H. E., Aron, A., and Brown, L. L. (2006). Romantic love: A mammalian brain system for mate choice. *Philosophical Transactions of the Royal Society of London. B: Biological Sciences* 361:2173–2186.

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press.

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

Gauthier, I., and Brown, D. D. (2004). The Perceptual Expertise Network: Innovation on collaboration [Electronic Version]. *APA Online: Psychological Science Agenda* from http://www.apa.org/science/psa/sb-gauthier.html.

Gauthier, I., Curby, K. M., Skudlarski, P., and Epstein, R. A. (in press). Individual differences in FFA activity suggest independent processing at different spatial scales. *Cognitive and Affective Behavioral Neuroscience.*

Gauthier, I., Curran, T., Curby, K. M., and Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience* 6:428–432.

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron* 34:161–171.

Gauthier, I., and Palmeri, T. J. (2002). Visual neurons: Categorization-based selectivity. *Current Biology* 12:R282–284.

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

Gauthier, I., and Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research* 37:1673–1682.

Gauthier, I., and Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance* 28:431–446.

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

Gauthier, I., Williams, P., Tarr, M. J., and Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research* 38:2401–2428.

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

Gluck, M. A., Meeter, M., and Myers, C. E. (2003). Computational models of the hippocampal region: Linking incremental learning and episodic memory. *Trends in Cognitive Sciences* 7:269–276.

Gold, J. I., and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences* 5:10–16.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In *Perceptual Organization in Vision: Behavioral and Neural Perspectives*, eds. R. Kimchi, M. Behrmann and C. R. Olson, 233–280. Mahwah, NJ: Lawrence Erlbaum Associates.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences* 15:20–25.

Grill-Spector, K., Kanwisher, N. (2003). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science* 16:152–160.

Grill-Spector, K., Knouf, N., and Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience* 7(5):555–562.

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

Hayward, W. G., and Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance* 23:1511–1521.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93:411–428.
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* 95:528–551.
Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance* 30:519–537.
Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:396–411.
Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:58–69.
Hollingworth, A. (in press). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition: Special Issue on Visual Search and Attention*.
Hollingworth, A., and Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance* 28:113–136.
Hummel, J. E. (1998). Where view-based theories break down: The role of structure in shape perception and object recognition. In *Cognitive Dynamics: Conceptual Change in Humans and Machines*, eds. E. Dietrich and A. Markman. Cambridge, MA: MIT Press.
Hummel, J. E., and Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review* 99(3):480–517.
Humphrey, G. K., and Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology* 46:170–190.
Johansen, M. K., and Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology* 45:482–553.
Johanson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics* 14:201–211.
Johnson, K. E., and Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General* 126:248–277.
Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory and Cognition* 13:289–303.
Jolicoeur, P., Gluck, M. A., and Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology,* 16, 243–275.
Joyce, C., and Cottrell, G. W. (2004). *Solving the visual expertise mystery*. Paper presented at the Proceedings of the Eighth Neural Computation and Psychology Workshop.
Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8:679–685.
Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* 17:4302–4311.
Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology* 55:271–304.
Kinder, A., and Shanks, D. R. (2001). Amnesia and the declarative/nondeclarative distinction: A recurrent network model of classification, recognition, and repetition priming. *Journal of Cognitive Neuroscience* 13:648–669.

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

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory* 2:554–565.

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

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:695–711.

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

Lawson, R., and Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance* 22:395–416.

Lee, M. D., and Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin and Review* 9:43–58.

Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology* 12:114–135.

Lissauer, H. (1890). Ein fall von seelenblindheit nebst einem Beitrage zur Theori derselben. *Archiv fur Psychiatrie und Nervenkrankheiten* 21:222–270.

Loftus, E. F. (2004). Memories of things unseen. *Current Directions in Psychological Science* 13:145–147.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review* 95(4):492–527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology* 22:1–35.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review* 109:376–400.

Logothetis, N. K., and Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex* 5:270–288.

Logothetis, N. K., and Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience* 19:577–621.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

Luria, A. R. (1987). *The Mind of a Mnemonist*. Cambridge, MA: Harvard.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Marr, D., and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 200:269–294.

Martin, A., Ungerleider, L. G., and Haxby, J. V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of objects. In *The New Cognitive Neurosciences*, ed. M. S. Gazzaniga. Cambridge: MIT Press.

Maurer, D., Grand, R. L., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science* 6:255–260.

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences* 7:293–299.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102:419–457.

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

Medin, D. L., and Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory* 7:241–253.

Meeter, M., Myers, C. E., and Gluck, M. A. (2005). Integrating Incremental Learning and Episodic Memory Models of the Hippocampal Region. *Psychological Review* 112:560–585.

Mensink, G. J., and Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review* 95:434–455.

Minda, J. P., and Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:775–799.

Mozer, M. C. (1991). *The Perception of Multiple Objects: A Connectionist Approach*. Cambridge, MA: MIT Press.

Mozer, M. C., and Sitton, M. (1998). Computational modeling of spatial attention. In *Attention*, ed. H. Pashler, 341–393). Hove, England: Psychology Press.

Myles-Worsley, M., Johnston, W. A., and Simons, M. A. (1988). The Influence of Expertise on X-Ray Image Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:553–557.

Naor-Raz, G., Tarr, M. J., and Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception* 32:667–680.

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Upper Saddle River, NJ: Prentice-Hall.

Noelle, D. C., and Cottrell, G. W. (1996). *Modeling interference effects in instructed category learning*. Paper presented at the 18th Annual Conference Of The Cognitive Science Society, La Jolla, CA.

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

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115:39–61.

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. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In *Multidimensional Models of Perception and Cognition*, ed. F. G. Ashby, 363–393. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nosofsky, R. M. (1998). Optimal performance and exemplar models of classification. In *Rational Models of Cognition*, eds. M. Oaksford and N. Chater, 218–247. New York: Oxford University Press.

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

Nosofsky, R. M., and Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. In *The Psychology of Learning and Motivation Vol. 28*, ed. D. L. Medin, 207–250). San Diego: Academic Press.

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

Nosofsky, R. M., and Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin and Review* 5:345–369.

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

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

O'Reilly, R. C., and Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences* 6:505–510.

O'Reilly, R. C., and Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review* 108:311–345.

Oliva, A. (2005). Gist of a Scene. In *Neurobiology of Attention*, eds. L. Itti, G. Rees and J. Tsotsos, 251–256. San Diego: Academic Press.

Palmer, S., Rosch, E., and Chase, P. (1981). Canonical perspective and the perception of objects. In *Attention and Performance IX*, eds. J. Long and A. Baddeley, 135–151). Hillsdale, NJ: Lawrence Erlbaum.

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. (1998). Formal models and feature creation. *Behavioral and Brain Sciences* 21:33–34.

Palmeri, T. J. (1999). Learning categories at different hierarchical levels: A comparison of category learning models. *Psychonomic Bulletin and Review* 6:495–503.

Palmeri, T. J. (2001). The time course of perceptual categorization. In *Similarity and Categorization*, eds. U. Hahn and M. Ramscar, 193–224. New York: Oxford University Press.

Palmeri, T. J., and 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., and Flanery, M. A. (2002). Memory systems and perceptual categorization. In *The Psychology of Learning and Motivation Vol. 41*, ed. B. H. Ross, 141–189. San Diego: Academic Press.

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

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

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

Perrett, D. I., Oram, M. W., and Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition* 67:111–145.

Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 223(1232):293–317.

Petrov, A. A., Dosher, B. A., and Lu, Z. L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review* 112:715–743.

Poggio, T., and Bizzi, E. (2004). Generalization in vision and motor control. *Nature* 431:768–774.

Poggio, T., and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature* 343:263–266.

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

Pratchett, T. (1993). *Men at Arms*: New York: Harper Collins.

Puce, A., Allison, T., Gore, J. C., McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology* 74:1192–1200.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin* 80:1–24.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review* 88:16–45.

Raaijmakers, J. G., and Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review* 88:93–134.

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

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

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

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science* 8:368–373.

Rhodes, G., Byatt, G., Michie, P. T., and Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *Journal of Cognitive Neuroscience* 16:189–203.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General* 126:288–311.

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience* 2:1019–1025.

Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. *Nature Neuroscience* 3(Suppl):1199–1204.

Riesenhuber, M., and Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology* 12:162–168.

Righi, G., and Tarr, M. J. (2004). Are chess experts any different from face, bird, or Greeble experts? *Journal of Vision* 4:504.

Roediger, H. L., III, Buckner, R. L., and McDermott, K. B. (1999). Components of processing. In *Memory: Systems, Process, or Function?* eds. J. K. Foster and M. Jelici, 31–65. Oxford, England: Oxford University Press.

Roltman, J. D., and Shadlen, M. D. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience* 22:9475–9489.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104:192–233.

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

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology* 46:178–210.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuroreport* 11(1):69–74.

Rouder, J. N., Ratcliff, R., and McKoon, G. (2000). A neural network model of implicit memory for object recognition. *Psychological Science* 11:13–19.

Schacter, D. L. (2000). Memory: Memory systems. In *Encyclopedia of Psychology Vol. 5*, ed. A. E. Kazdin, 169–172. Washington, DC: American Psychological Association.

Schacter, D. L., Chiu, C.-Y. P., and Ochsner, K. N. (1993). Implicit memory: A selective review. *Annual Review of Neuroscience* 16:153–182.

Schacter, D. L., and Tulving, E. (1994). *Memory Systems*. MIT Press.

Schall, J. D. (2001). Neural Basis of Deciding, Choosing, and Acting. *Nature Reviews Neuroscience* 2:33–42.

Schall, J. D. (2004). On building a bridge between brain and behaviour. *Annual Review of Psychology* 55:23–50.

Schyns, P. G., Goldstone, R. L., and Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences* 21:1–54.

Schyns, P. G., and Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:681–696.

Sergent, J., Ohta, S., and MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain* 115(1):15–36.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior* 6:156–163.

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

Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review* 1:2–28.

Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171:701–703.

Shiffrin, R. M., Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review* 4:145–166.

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

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science* 7:301–305.

Simons, D. J., and Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences* 9:16–20.

Smith, E. E., Shoben, E. J., and Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review* 81:214–241.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology* 44:408–463.

Smith, P. L., and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences* 27:161–168.

Snowden, P. T., Davies, I. R. L., and Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: A functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance* 26:379–390.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* 99:195–231.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory* 82:171–177.

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

Squire, L. R., and Schacter, D. L. (2002). *Neuropsychology of memory (3rd ed.)*. New York, NY: Guilford Press.

Squire, L. R., Stark, C. E. L., and Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience* 27:279–306.

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

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology* 25:207–222.

Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance* 28:913–932.

Tanaka, J. W., and Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science* 12:43–47.

Tanaka, J. W., Curran, T., and Sheinberg, D. L. (2005). The training and transfer of real world perceptual expertise. *Psychological Science* 16(2):145–151.

Tanaka, J. W., and Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology* 46A:225–245.

Tanaka, J. W., and Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 23:457–482.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience* 19:109–139.

Tanaka, K. (2002). Neuronal representation of object images and effects of learning. In *Perceptual Learning*, eds. M. Fahle and T. Poggio, 67–82. Cambridge, MA: MIT Press.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review* 2:55–82.

Tarr, M. J., and Bülthoff, H. H. (1998). Image-based object recognition in man, monkey, and machine. *Cognition* 67:1–20.

Tarr, M. J., and Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Science* 7:23–30.

Tarr, M. J., and Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition* 67:73–110.

Tarr, M. J., and Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience* 3:764–769.

Tarr, M. J., Kersten, D., and Bülthoff, H. H. (1998). Why the visual recognition system might encode the effects of illumination. *Vision Research* 38:2259–2275.

Tarr, M. J., and Kriegman, D. J. (2001). What defines a view? *Vision Research* 41:1981–2004.

Tarr, M. J., and Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology* 21:233–282.

Tarr, M. J., and Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science* 1:253–256.

Tarr, M. J., Williams, P., Hayward, W. G., and Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience* 1:275–277.

Tjan, B. (2001). Adaptive object representation with hierarchiacally-distributed memory sites. *Advances in Neural Information Processing Systems* 13:66–72.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology* 12:97–136.

Tulving, E. (1985). How many memory systems are there? *American Psychologist* 40:385–398.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science* 2:67–70.

Tulving, E., and Schacter, D. L. (1990). Priming and human memory systems. *Science* 247:301–306.

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

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition* 32:193–254.

Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience* 5:682–687.

Vuong, Q. C., and Tarr, M. J. (2004). Rotation direction affects object recognition. *Vision Research* 44:1717–1730.

Waldron, E. M., and Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin and Review* 8:168–176.

Wells, G. L., and Loftus, E. F. (2003). Eyewitness memory for people and events. In *Handbook of Psychology: Forensic Psychology, Vol. 11*, ed. A. M. Goldstein, 149–160. Hoboken, NJ: John Wiley and Sons, Inc.

Wells, G. L., and Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology* 54:277–295.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology* 81:141–145.

Zaki, S. R., and Nosofsky, R. M. (2001). A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognitive, Affective and Behavioral Neuroscience* 1:344–359.

Zaki, S. R., and Nosofsky, R. M. (2004). False prototype enhancement effects in dot pattern categorization. *Memory and Cognition* 32:390–398.

Zaki, S. R., Nosofsky, R. M., Jessup, N. M., and Unversagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society* 9:394–406.

Zhang, L., and Cottrell, G. W. (2005). Holistic processing develops because it is good. In *Proceedings of the 27th Annual Cognitive Science Conference*, eds. B. G. Bara, L. Barsalou and M. Bucciarelli. Mahwah: Lawrence Erlbaum.