7

Modeling Perceptual Expertise

Thomas J. Palmeri and Garrison W. Cottrell

> *To have one's hunches about how a simple combination of processes will behave repeated dashed by one's own computer program is a humbling experience that no experimental psychologist should miss. Surprises are likely when the model has properties that are inherently difficult to understand, such as variability, parallelism, and nonlinearity—all, undoubtedly, properties of the brain.*
>
> Hintzman, 1990

INTRODUCTION

In this chapter we delineate what we believe to be the important characteristics of perceptual expertise that a complete model should try to capture, motivate why computational models are important for any complete understanding of perceptual expertise, and then describe several models that have been constructed to account for visual object processing, perceptual categorization, and face processing. Models are evaluated in terms of their ability to account for the phenomena of perceptual expertise. A challenge in developing a comprehensive computational model of perceptual expertise is that the range of empirical phenomena, many of which are described in the various chapters in this volume, are at the intersection of so many fundamental areas of perception and cognition. This implies that any complete understanding of the various facets of perceptual expertise requires a theoretical coupling across a number of traditionally distinct areas of visual perception and visual cognition.

Radiologists, ornithologists, birders, firefighters, and other specialists are noted for their remarkable ability to rapidly recognize, categorize, and identify objects and events in their domain of expertise. Understanding the unique abilities of experts can certainly have important real-world implications for enhancing the development of expertise in the workplace. However, we believe that understanding perceptual expertise has implications beyond simply characterizing the behavior of individuals with idiosyncratic skills in highly specialized domains (Palmeri et al., 2004). Mechanisms of perceptual expertise may also explain some of the unique aspects of everyday domains such as recognizing faces (Diamond & Carey, 1986; Gauthier & Tarr, 2002), words (McCandliss, Cohen, & Dehaene, 2003), or letters (Wong & Gauthier, 2006). We view perceptual expertise as the logical end point of the normal trajectory of learning, rather than an idiosyncratic skill. This allows us to exploit studies of experts to understand the general principles and limits of human learning and plasticity. Furthermore, viewing faces, words, and letters

9780195309607_0197-0244_Gauthier_GAUT_Ch07 10/7/2009 16:59 Page:197

as domains of perceptual expertise may yield new insights into how abnormal brain development or brain damage can lead to the perceptual and cognitive deficits seen in autism, dyslexia, prosopagnosia, and other conditions, and may lead to breakthroughs in education and treatment.

This theoretical view of perceptual expertise is mirrored in our approach to developing computational models. A comprehensive computational theory of the development of perceptual expertise remains elusive. However, viewing perceptual expertise as the end point of the trajectory of normal learning suggests that we should ultimately look to various computational models from literatures such as object recognition, face recognition, perceptual categorization, automaticity, and skill learning as theoretical starting points. Indeed, we believe that the development of perceptual expertise should be explored first within the context of extant models of normal visual cognition. Of course, this is a hypothesis, not an axiom. We may ultimately discover that specialized domains of expertise require specialized domain-specific computational models. However, to this point, we have not needed to make that assumption.

We begin by delineating what we believe to be the core phenomena of perceptual expertise (largely taken from Palmeri et al., 2004) that a comprehensive model should account for. We then briefly review some general issues in modeling, followed by consideration of models of object processing, perceptual categorization, and face processing as models of perceptual expertise. Of course, no model can account for all of the characteristics of perceptual expertise (although significant progress has been made). At the end of the chapter, we will briefly discuss ways in which some of these models might be theoretically integrated.

THE CORE FEATURES OF PERCEPTUAL EXPERTISE

There are a number of behavioral and neural characteristics that distinguish novices and experts, many of which are discussed in other chapters in this volume. It goes without saying that experts know more than novices about their domain of expertise. They can verbalize more properties, describe more relationships, make more inferences, and so forth (e.g., Ericsson et al., 2006; Kim & Ahn, 2002; Murphy & Wright, 1984; Johnson & Mervis, 1997). This is, after all, what makes them experts. Our focus here is on behavioral and neural changes in visual cognition that underlie *perceptual* expertise. Here we provide a brief summary of some of the phenomena that any comprehensive computational theory of the development of perceptual expertise must ultimately account for:

- Novices often rely on explicitly verbalized category knowledge in the form of rules or ideal cases that are acquired from reference manuals or explicit instruction (e.g., Allen & Brooks, 1991) or that are created through induction (e.g., Johansen & Palmeri, 2002). By contrast, although experts have more verbal knowledge about a domain, expert categorization often seems removed from explicit and conscious deliberation (e.g., Brooks, Norman, & Allen, 1991; Sloman, 1996). What accounts for this shift from conscious deliberation to more automatic decisions?

- Novices are slow and deliberate in their decisions, perhaps reflecting their use of explicit rules and strategies. The development of expertise is accompanied by a marked speedup in processing, originally characterized by the power law of practice (Newell & Rosenbloom, 1981; but see Heathcote, Brown & Mewhort, 2000; Rickard, 1997; Palmeri, 1999). What causes this increase in the speed of decisions with perceptual expertise?
- One important aspect of this speedup is the so-called "entry level shift" (Jolicoeur, Gluck, & Kosslyn, 1984; Tanaka & Taylor, 1991). For novices, categorizations at the basic level ("dog" or "bird") are faster than categorizations at either a superordinate ("animal" or "plant") or a subordinate level ("robin" or "terrier"). The fastest level of categorization is often described as the entry-level into conceptual knowledge. For experts, there is an entry-level shift whereby subordinate-level categorizations are made as quickly as basic-level categorizations (Johnson & Mervis, 1997; Tanaka & Taylor, 1991). Does this shift reflect a qualitative change in how expert categories are processed, or is it a manifestation of a more continuous quantitative change in the efficiency of processing over learning (Joyce & Cottrell, 2004; Mack, Wong, Gauthier, Tanaka, & Palmeri, 2007; Tong et al., 2008)?
- Novices and experts show different patterns of interference. Novices are easily distracted whereas experts may be able to simultaneously engage in other tasks while making expert decisions. Part of this apparent lack of interference may be because experts no longer use explicit verbalizable routines, so concurrent verbal activity does not interfere with performance. But when experts engage in tasks that tap the same representational resources used for other domains of expertise, they suffer interference in ways unseen in novices (Gauthier & Curby, 2005; Gauthier, Curran, Curby, & Collins, 2003; Rossion et al., 2004; see also Curby & Rossion, this volume). What accounts for these different patterns of interference in experts and novices?
- Novices can attend to part of a complex object while ignoring irrelevant parts. By contrast, experts show interference from irrelevant variation in an unattended part. For example, in a part-matching task—adapted from work in the face recognition literature (Young, Hellawell, & Hay, 1987)—subjects are asked to attend to the top part of a whole object. After a brief delay, a second object is shown with the irrelevant bottom either matching or mismatching the bottom of the first object. When judging whether the top is the

same or different, novices are unaffected by the irrelevant bottom, whereas experts show facilitation when the irrelevant bottom would lead to the same decision, and interference when the irrelevant bottom would lead to a different decision (Cheung, Richler, Palmeri, & Gauthier, 2008; Gauthier et al., 2003; Richler et al., 2008). However, the direction of this interference depends upon the objects of expertise—for example, Chinese readers do not suffer this interference when viewing Chinese characters, while novices do (Hsiao & Cottrell, 2008). What causes this nominal processing cost associated with expertise, and what explains when the expert will show this cost?

- Experts generalize their knowledge. Experts can learn to categorize and identify new objects more quickly than novices, and can discriminate novel objects better than novices, at least so long as the new objects are similar to other objects in their domain of expertise (i.e., they vary systematically in the same way as other learned objects; Gauthier & Tarr, 1997, 2002; Tanaka, Curran, & Sheinberg, 2005).
- The ability of experts to generalize is also limited in specific ways (Palmeri, 1997). Experience is often limited to particular viewpoints. In much the same way that face recognition is impaired by inversion, expert object recognition is impaired by inversion as well (Diamond & Carey, 1986). For example, experts are highly sensitive to changes in the configuration of features, but only when objects are presented in a familiar orientation (Maurer, LeGrand, & Mondloch, 2002; Mondloch, LeGrand, & Maurer, 2002; Gauthier & Tarr, 1997). What does this limited generalization and sensitivity to orientation or viewpoint imply about how experts represent their perceptual knowledge?
- Finally, experts show different patterns of brain activity than novices. For example, with fMRI it has been shown that the fusiform face area (FFA) is not just involved in face recognition but is activated by objects of expertise in real-world experts such as birders (Gauthier, Skudlarski, Gore, & Anderson, 2000; Xu, 2005; but see Grill-Spector, Knouf, Kanwisher, 2004) and by objects of expertise created in the lab (Gauthier & Tarr, 1997, 2002). Similarly, event-related potential (ERP) markers for face recognition, such as the N170, which shows highest amplitude for faces, also show higher amplitude when observing objects of visual expertise over objects that are not (Tanaka & Curran, 2001; but see Scott, Tanaka, Sheinberg, & Curran, 2006). Why are brain areas that are devoted to one domain of expertise, in this case faces, recruited for another domain of expertise? What is different about an expert domain such as letter perception, which recruits different brain areas entirely (Gauthier, Tarr, et al., 2000; Wong & Gauthier, 2006)?

No single computational model can, at present, account for all of the various behavioral aspects of the development of perceptual expertise. At the same time, some models do speak to certain aspects of expertise, and a comprehensive computational theory may be possible by combining complementary models (or lessons learned from those models). This chapter provides an overview of models from visual object processing, perceptual categorization, and face recognition that we believe provide insights into the mechanisms underlying the development of perceptual expertise. At the end of the chapter, we sketch some possible avenues for theoretical integration toward a comprehensive model of perceptual expertise. Ultimately, we need a model that captures the long-term dynamics of learning throughout the development of expertise, the short-term dynamics of novice and expert decisions, and the dynamic interplay of the various brain structures that are recruited at various stages of expertise and how those brain structures are molded by experience.

SOME WHAT'S, WHY'S, AND HOW'S OF MODELING

First, what is a model? Definitions vary widely. For our purposes, we define a model as a theory that has been formalized in terms of mathematical equations or computer simulations. An advantage of formalization beyond mere verbal description is that it forces the theorist to be explicit about all components of the theory, allowing those theories to be clearly articulated, rigorously evaluated, and potentially falsified (e.g., Hintzman, 1990). Intuitions about how components of a theory interact are often overly simplistic or downright wrong. Thus, models often generate new and novel predictions regarding empirical work that would otherwise be unavailable.

There are many varieties of models (e.g., Luce, 1995). There are *statistical models* of data, such as structural equation modeling, multidimensional scaling, factor analysis, principal components analysis, or nonlinear regression. Statistical models can be applied to any data from any domain, at least so long as those data abide by the assumptions underlying the valid use of those models. As models of data, statistical models do not explain why the data was observed. They analyze what was observed. There are *normative models*, such as optimal control theory, Bayesian decision theory, and expected utility theory. Normative models attempt to explain what should be done in a particular situation based on various optimality considerations. To the extent that individuals deviate from optimality, these models fail to explain what people actually do, although often the real issue here is determining what the individual's utilities are. There are *artificial intelligence* and *machine learning models* for computer vision, face recognition, expert reasoning and problem solving, and spoken language recognition. These models attempt to mirror the complex behavior of humans and may even make use of what is known about the processes underlying human perception and cognition, but ultimately they aim to see, hear, or reason as well as people, or even better than people, irrespective of whether the underlying mechanisms bear any resemblance to human mental and neural processes.

Our focus is on *process models* of cognition and perception. They attempt to explain how and why people think, remember, and perceive the way they do. They formally instantiate hypotheses about the mechanisms that lead to observed behavior and are often grounded in neurally inspired computational mechanisms. The somewhat counterintuitive goal of these models is to make the same errors people make, to be slow when people are slow and fast when people are fast, and to be able to mimic the effects of brain damage and mental illness. While process models have different goals than statistical models, normative models, and artificial intelligence models, the initial development of a process model may be closely related to those models. For example, it could be possible to develop a process model that mechanistically instantiates Bayesian decision theory and then see if this model accounts for human behavior, or an existing process model that accounts well for human behavior may end up being related mathematically to Bayesian decision theory years after the model was first developed (e.g., Myung, 1994). Sometimes whole classes of models may be unknowingly related; it took many years for the field to fully realize the intimate relationship between neural networks and statistical models (e.g., Bishop, 1995).

So why model?

Models Rush in Where Theories Fear to Tread

Theories are relatively high-level verbal descriptions of the processes underlying behavior. As such, they are often vague about specific mechanisms, which can make it difficult to make a priori predictions based upon them. However, using machine learning or statistical techniques, one can often create a *working model* of the process involved, even in cases where there is no theory. These can then be examined in order to gain insights into how the process might work. For example, an early class of face processing models was built from "eigenfaces" extracted using principal components analysis[1], a well-known statistical analysis technique (O'Toole, Abdi, Deffenbacher, & Valentin, 1993; Turk & Pentland, 1991). Figure 7.8 shows some eigenfaces, which served as the features for the model. These whole-face templates are

[1] Eigenfaces are whole-face templates that arise from finding the directions of maximum variance in a data set of faces. In particular, they encode the strongest *covariances* between the pixels in a set of faces. More formally, they are the principal eigenvectors of the covariance matrix of the data. For example, in a data set of male and female faces, the first principal component, or the first eigenface, will often correspond to the distinction between male and female. In a data set of Asian and Caucasian faces, the first principal component may correspond to the difference between Asians and Caucasians. It is interesting to note here that neural network models developed around the same time are formally equivalent to these eigenface models (Cottrell & Metcalfe, 1991; Fleming & Cottrell, 1990; Furl, Phillips, & O'Toole, 2002; Golomb, Lawrence, & Sejnowski, 1991).

clearly holistic features. Each face is represented by its correlation with the eigenfaces, giving a vector of numbers that can be matched with a set of stored representations with labels (e.g., identity or emotion). The label of the closest stored face is the label chosen by the model. The original face can be reconstructed as a weighted sum of these eigenfaces. In other words, faces are a point in this high-dimensional eigenspace (the number of dimensions corresponds to the number of eigenfaces used). In this sense, they are one of the first computational instantiations of Valentine's "face space" account of face processing (Valentine, 1991), and provide insights into how a face space might arise, and how holistic effects can be accounted for (more on this later).

Models Can Make Counterintuitive Predictions

Our opening quote by Hintzman (1990) makes this point well. Hintzman's own work showed that behavior that seems to clearly indicate some form of abstraction from specific experiences could actually emerge from a simple learning mechanism with no abstractions in the model whatsoever, just memory for specific experiences (Hintzman, 1986; see also Medin & Schaffer, 1978; McClelland & Rumelhart, 1985; Nosofsky, 1984) (more on this later, too). The behavioral abstraction of the model arose from the mechanism that accessed and used specific memories. Thus, models can be intuition pumps for alternative conceptualizations of hypothetical mechanisms and how they might work.

Models Can Be Manipulated in Ways People (and Animals) Cannot

A computational model allows the modeler to explore "what if" questions that cannot be easily explored with humans and that would be difficult or impossible to explore, even with animal models. By performing these experiments that may go beyond the parameters that are reasonable for the human brain or beyond real-world experience, one can begin to see why things are the way they are. For example, with models we can explore systematically the effects of manipulations such as variations in cortical architecture (e.g., hemispheric vs. "monolithic" brain models, Shillcock & Monaghan, 2001), variations in processing resources (e.g., variations in number of hidden, Plaut et al., 1996), variations in the environment (e.g., What if our parents were cans, cups, or books instead of humans? i.e., Is there something special about face expertise versus visual expertise in general? Joyce & Cottrell, 2004; Sugimoto & Cottrell, 2001; Tong et al., 2008), or variations in brain damage within the very same "brain" (e.g., Hinton & Shallice, 1991; Plaut et al., 1996; Plunkett & Juola, 1999).

Models allow us to explore the effects of a far denser space of potential brain lesions and brain damage than possible with our limited neuropsychological samples. Extremely rare disorders seen in single cases may be attributable to the tails of a distribution of cases seen in an extremely large sample of possible cases (e.g., Thomas & de Wet, 1998). Or extreme cases may correspond to behaviorally distinct regimes. For example, Plunkett and Juola (1999) showed by lesioning their model of past tense formation that one could get behavior that appeared as if "the rule system" was broken, and other behavior that appeared as if "the rote memory system" was broken, even though their model had neither of these components. Such behavioral dissociations and double dissociations are often interpreted as evidence for modules that map directly onto the particular behaviors that are preserved or damaged. However, while modular accounts certainly do predict dissociations and double dissociations, models (combined with simple logic) show that the arrow of implication only goes in one direction: observing dissociations and double dissociations does *not* imply a modular organization (Palmeri & Flanery, 2002; Palmeri & Gauthier, 2004; Plaut, 1995; Plunkett & Juola, 1999; Thomas & de Wet, 1999). Only by having an explicit computational model is it possible to explore how brain damage might affect behavior by breaking the model in various ways (e.g., Dailey & Cottrell, 1999), providing a more comprehensive theoretical account of the deleterious effects of brain damage and mental illness on behavior (e.g., Treat et al., 2007).

Models Can Be Analyzed in Ways People (and Animals) Cannot

For example, in neurocomputational models, one can perform single-cell recordings from "birth" to "death" and fully map out the receptive and projective fields of every unit in the network. We can selectively ablate and restore parts of the network, even down to the single-unit level, to assess their contribution to processing. We can measure responses at different layers of processing to find the best match to human data (e.g., which level accounts for a particular judgment: perceptual processing, object representation, or categorization, Dailey et al., 2002). In general, with any model, we can analyze the interactions between components at a level of detail unavailable in biological preparations.

In this way, models can also generate new theories, in that they may allow the theorist to see formal relations between different aspects of behavior that might not be obvious at first blush (e.g., Logan, 2004). We also remark here that models that are formalized to account for data from one domain can often be extended to apply to data from other domains. Indeed, one might argue that such generalization is the hallmark of a "good" model. As we will see in this chapter, some models of face and object processing can be extended to perceptual expertise more generally (e.g., Joyce & Cottrell, 2004). Models of categorization, identification, and automaticity have also been combined into a more general model of perceptual categorization and perceptual expertise (see Palmeri et al., 2004).

Of course, modeling is not all wine and roses. One common criticism of models goes something like "with enough parameters to tweak, a model can predict anything." This is a valid criticism when a model has too many free parameters relative to the number of degrees of freedom in the data. And too many free parameters can also risk overfitting, where the model accounts not only for stable quantitative and qualitative effects, but also the random variability—the "noise"—in the data as well. In addition, it's true that just as underspecified verbal theories are unfalsifiable, underspecified or weakly constrained computational models risk being unfalsifiable as well. The source of this common criticism may partly stem from the way modeling results are commonly portrayed in the literature. Often the focus is on a particular parameterization of a particular model that fits a particular pattern of observed data. In that sense, these are often more like cases of "postdiction" rather than prediction (e.g., Roberts & Pashler, 2000), or at least they can be (perhaps falsely) interpreted that way. On the one hand, demonstrations of the sufficiency of a particular formal model can be important in showing that a hypothetical mechanism *can* work, even if that mechanism seems counter-intuitive at first blush. On the other hand, demonstrating sufficiency is only the first step.

The fact that most models have free parameters doesn't mean that those models are unconstrained. Models may produce similar qualitative predictions for a wide range of (plausible) parameter values, demonstrating that the pattern of behavior is inherent in the model structure rather than in a specific choice of parameter values (e.g., Johansen & Palmeri, 2002; Pitt, Kim, Navarro, & Myung, 2006). Sometimes model parameters can be chosen a priori to have values that correspond to a range of values that have been measured, either through neurophysiology, psychophysics, or scaling techniques (e.g., Boucher et al., 2007; Nosofsky, 1992b). Specific parameter values may also be chosen because they can be shown to be optimal in some way (Nosofsky, 1998). Parameter-free predictions can also be generated by first fitting the model to one part of the data, such as training data in the case of learning models, and then model predictions with fixed parameters can be compared to observed data that are outside the range of the training data (e.g., Busemeyer & Wang, 2000; Dailey et al., 2002). It's up to the modeler to go through these extra steps.

Perhaps more vexing is that multiple models can produce the same behavior; indeed, there are an infinite number of possible models that can produce the same input–output behavior (Moore, 1956). This is where model selection, and even model competition, may come in. One criterion is to select models based upon an application of Ockham's razor—simpler models that account for the data are to be preferred to more complex ones. Complexity can be measured in, for example, the number of parameters of the model, or models may be nested within one another. Various statistical tests can weigh whether additional complexity leads to a significantly better account of the data. While as a first-order approximation, the more constraints there are on a model's parameters, or the fewer free parameters a model has, the simpler the model is, in actuality, quantifying model flexibility and model falsifiability can get quite complicated (e.g., see Pitt et al., 2002). If competing models are of comparable complexity, and they account for the same set of data, then the models must be weighed on the basis of their predictions. Some of the best modeling work adopts a strong inference approach (Platt, 1964) by designing critical experiments that contrast the predictions of competing models (e.g., Ashby & Waldron, 1999; Nosofsky & Palmeri, 1997).

MODELS OF OBJECT PROCESSING

One approach to developing a model of perceptual expertise is to first turn to models of generic object processing and ask how those models might account for the development of perceptual expertise after the right kind and the right amount of learning has taken place. Of course, this assumes that perceptual expertise can be seen as the end point of the trajectory of normal learning, which need not be the case. However, if we can do so without making any additional assumptions, then by Ockham's razor, we should. In this section, we begin by briefly describing the problems that models of object processing try to solve. We then turn to a number of extant models of object processing and discuss how those models might account for the development of perceptual expertise.

How do we know that an object is the same object we have seen before? Or at least how do we know that it is the same kind of an object that we have seen before? At first glance, what could be simpler? We just open our eyes and we know what things are and whether we recognize them or not. Of course, these naïve intuitions belie the tremendous computational challenges facing our visual system with every glance at the world around us. 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. Somehow, our visual system overcomes this tremendous variation in visual information to create a stable perception of the world. Three-dimensional objects seem stable as we move around, as objects move around, and as the lighting changes. But how does the visual system allow us to perceive this stability when the two-dimensional images falling onto our retinae are changing so dramatically? This is known as the *invariance problem.*

Almost all solutions to the problem of visual object recognition 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 (semi) independently encode a different local aspect of the visual scene. The visual system transforms this high-dimensional stimulus representation into the activation of a million nerve fibers. While this is hardly low dimensional, it is relative to the dimensionality of the retinal stimulation. However, once the cortex is reached, the dimensionality increases again, around 100-fold. This high-dimensional representation allows the cortex to extract many independent features from the input that are relatively sparse (meaning, a small fraction of the neurons fire) and distributed. These independent features are then used to recognize objects. Different theories propose varying solutions to the

9780195309607_0197-0244_Gauthier_GAUT_Ch07 10/7/2009 16:59 Page:206

problem of creating a low-dimensional object representation, differing markedly in the form of that representation, and in how great a dimensionality reduction is assumed. Most stop there and do not consider the dimensionality expansion that occurs (although ongoing work on "overcomplete" independent components analysis is suggesting ways this might be done).

Some of the earliest models of object processing assumed that the fundamental goal of vision was to create a faithful description of the objects in the world, in a sense reconstructing the three-dimensional structure of objects and their spatial relations within visual representations. One of the most intuitive proposals for constructing such representations, originally put forth by Marr and Nishihara (1978; Marr, 1982), assumes that every given object can be described in terms of generic three-dimensional primitives and their spatial relations. This idea was adopted by Biederman (1987) and implemented as a neural network simulation by Hummel and Biederman (1992), in the "recognition-by-components" (RBC) theory. RBC assumes a small vocabulary of three-dimensional primitives called "geons" and specifies the rules, based upon "viewpoint invariant properties" for extracting geons from images (see Figure 7.1). The key idea is that the geons have properties that are invariant to some distortion and viewing angle, thereby directly addressing two of the twin challenges facing vision. This represents an extreme dimensionality reduction: different views of an object and different exemplars within an object class all map onto the same configuration of three-dimensional geon primitives.

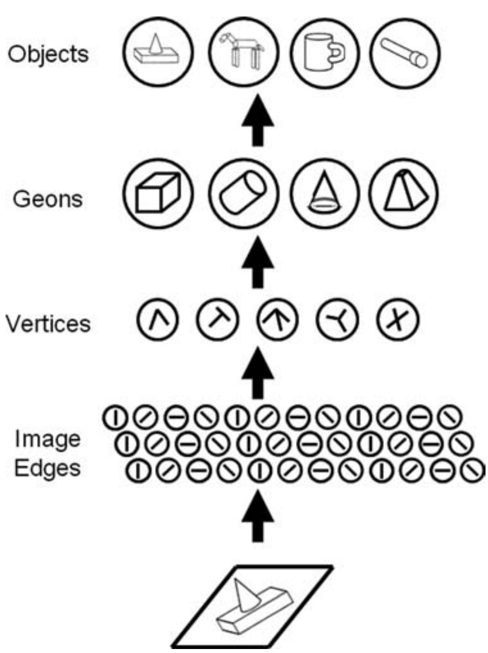


Figure 7.1 Recognition-by-components (Biederman, 1987; Hummel & Biederman, 1992) assumes that a retinal image is initially described in terms of its edges. A variety of nonaccidental primitive features are extracted from this edge description, such as "L" junctions, "Y" junctions, and other properties. Combinations of various viewpoint-invariant primitives signal the presence of one of the small number of geons. Viewpoint-invariant object recognition involves recognizing the particular combination and relative configuration of the viewpoint invariant geon representations extracted from a complex object.

In the context of RBC, any basic-level visually defined category may be uniquely represented by a small subset of geons in particular spatial configurations. For example, a wide variety of birds are made up of roughly the same parts—head, body, wings, beak—perhaps with the exception of atypical birds like penguins. The problem of basic-level categorization is solved because all birds map onto the same structural description. Of course, if these structural descriptions lack metric information about the relative sizes of geons and the quantitative location of geons with respect to one another, which early versions of RBC did, then within-category discrimination is nearly impossible. That is, if one believes that visual expertise is just an end point of normal object recognition, it is unclear how this theory could be generalized to that situation. More recent versions have added a separate pathway for metric information in order to solve the problem of face recognition, but the evidence for two distinct recognition pathways in the brain is weak or nonexistent. In any case, it is unclear how such a model would account for *any* of the phenomena of perceptual expertise.

These structural description models have been challenged based on other objections (e.g., Edelman, 1997, 1999). Specifically, a variety of laboratories have shown that object recognition depends on experience with particular views of an object (e.g., Bülthoff & Edelman, 1992; Tarr, 1995; Tarr & Pinker, 1989). Viewpoint-invariant recognition derives from experience with multiple views (Tarr, Kersten, & Bülthoff, 1998), not because object representations are inherently viewpoint independent, as suggested by RBC and its variants. These and other results led to an alternative class of object processing models based on stored representations of previously experienced views of objects (see Figure 7.2). These theories often begin with a very different assumption of the goals of vision. Rather than assuming that we need to reconstruct the three-dimensional world inside our heads, view-based approaches typically stress the importance of generalization from past to present experiences (Edelman, 1999; Shepard, 1994). Considering that we are vanishingly unlikely to ever experience the same situation twice, and that similar objects often give rise to similar consequences, survival demands that we recognize the similarities (Shepard, 1987). So one solution is to create representations that preserve the similarity structure between objects even if those representations do not encode three-dimensional structure explicitly (Edelman, 1999).

View-based models solve the problem of viewpoint invariance by generalizing according to the similarity between the current representation of an object and the stored representations of objects in memory, without any need for explicit image transformations (e.g., Poggio & Edelman, 1990; Reisenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007). Models of this sort account well for experimental patterns of interpolation between learned views and limited extrapolation beyond learned views (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992). The most recent instantiations of view-based models (e.g., Jiang et al. 2006; Reisenhuber & Poggio, 1999, 2000) incorporate a number of neurobiological constraints in terms of

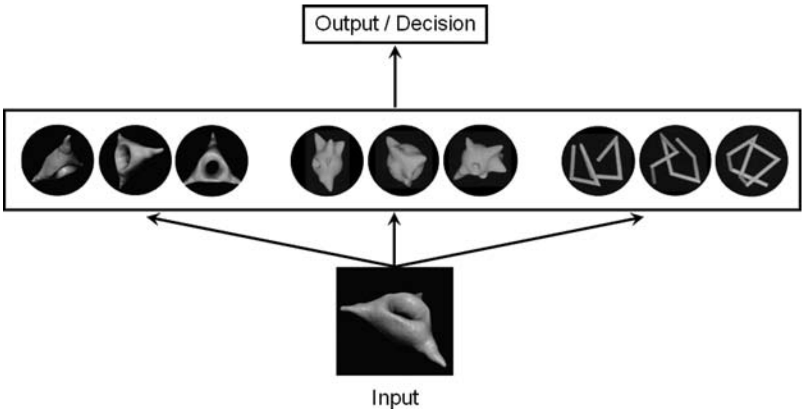


Figure 7.2 View-based models assume that objects are represented in terms of stored views. Interpolation to new views differing from experienced views in terms lighting, viewpoint, and other factors is based on similarity to stored views. In the model illustrated in the figure, an input image is matched against the stored views depicted for three different objects. The output depends on what the model is trained to reproduce. It can be trained to generate a canonical view of an object plus information about the pose of the viewed image (e.g., Poggio & Edelman, 1990). Or it can be trained to categorize, identify, or recognize the viewed image.

computations that are performed and the hierarchy of transformations performed by earlier stages of visual processing (illustrated later in Figure 7.9). One of the appealing aspects of using similarity as a means to invariance is that the same kind of mechanism can account for how we generalize across both viewing and category variation (more on the latter later). While invariance over view can be achieved by encoding multiple views of individual objects, invariance over kind can be achieved by encoding multiple views of multiple objects of that kind. Given sufficient views of objects from multiple classes, both view and class invariance can be achieved.

Unlike the RBC model, view-based models can be naturally extended to account for aspects of perceptual expertise. For example, if one has a lot of experience with certain classes of objects, for example, faces, the model would store many representations of views of the same face. This dense representation would clearly lead to the ability to finely discriminate between faces, while discrimination would be poorer for the more sparse representations of other objects (Palmeri et al., 2004). Thus, view-based accounts provide a language for thinking about the kind of perceptual learning that takes place with expertise. Perceptual experts are better able to visually discriminate between objects within their domain of expertise, even if those objects are unfamiliar (Tanaka, Curran, & Sheinberg, 2005). To the extent that experts have learned a wide range of views from a wider range of objects, they have a larger "vocabulary" of stored images from which to encode a new shape and represent its similarity to other objects. But the extent of this increase in perceptual discriminability is limited to the range of objects experts have had experience with (Gauthier & Tarr, 1997), and view-based models naturally account for the limited extrapolation beyond the experienced set (see also Palmeri, 1997).

Models like this can account for perceptual speedups by assuming an inverse relationship between response time and categorization certainty, with certainty proportional to the density of the (correctly labeled) representations[2]. Furthermore, if the label stored with the representations corresponds to the subordinate level, this would also account for the entry level shift with expertise, as the density of subordinate-level representations would be greater than category-level ones. Also, it is clear that if most representations are in a canonical orientation, then inversion effects should fall out of the representation. It is less clear how they would account for the difference in interference patterns between novices and experts, the shift from rule-based behavior to similiarity-based, or fMRI results showing the use of similar regions of cortex for different areas of expertise. Finally, if a view-based model uses full-image representations, then it would seem that all objects should show holistic responses, not just objects of expertise.

Ullman, Vidal-Naquet, and Sali (2002; Ullman, 2007) suggested an alternative model of object processing that in some sense combines elements of structural description models and view-based models. They showed that view-based features of "intermediate complexity" best account for basic-level classification, where the particular features and their size were determined by the mutual information between the patch and the category label (see Figure 7.3). For faces, these features might include what we would generally call the "parts" of a face such as the eyes, nose, or mouth, and for a car these might include "parts" like the wheel or the driver's side window. It is important to emphasize that these are not parts in any way like geons are parts. These are viewpoint-dependent view-based fragments. They are generally not full images, although full-image representations can be part of the suite of image features. Moreover, spatial relationships between these parts are not explicitly encoded, but if the local context is preserved and local features overlap, there is an implicit representation of configural information. So generic object recognition at the basic level may be view based, and it need not depend on full images of objects. Moreover, it is tempting to speculate about the relationship between these view-based fragments and the kinds of ad hoc feature sensitivities seen in neurons in TEO (Tanaka, 1996, 2002; see also Serre et al., 2007). Neither correspond directly to what we might typically think of as a distinct object part, or to anything like a geon.

[2] Ultimately, of course, any complete account of the time-course of perceptual expertise demands models that incorporate true temporal dynamics, not simply correlating time with some other nontemporal measure.

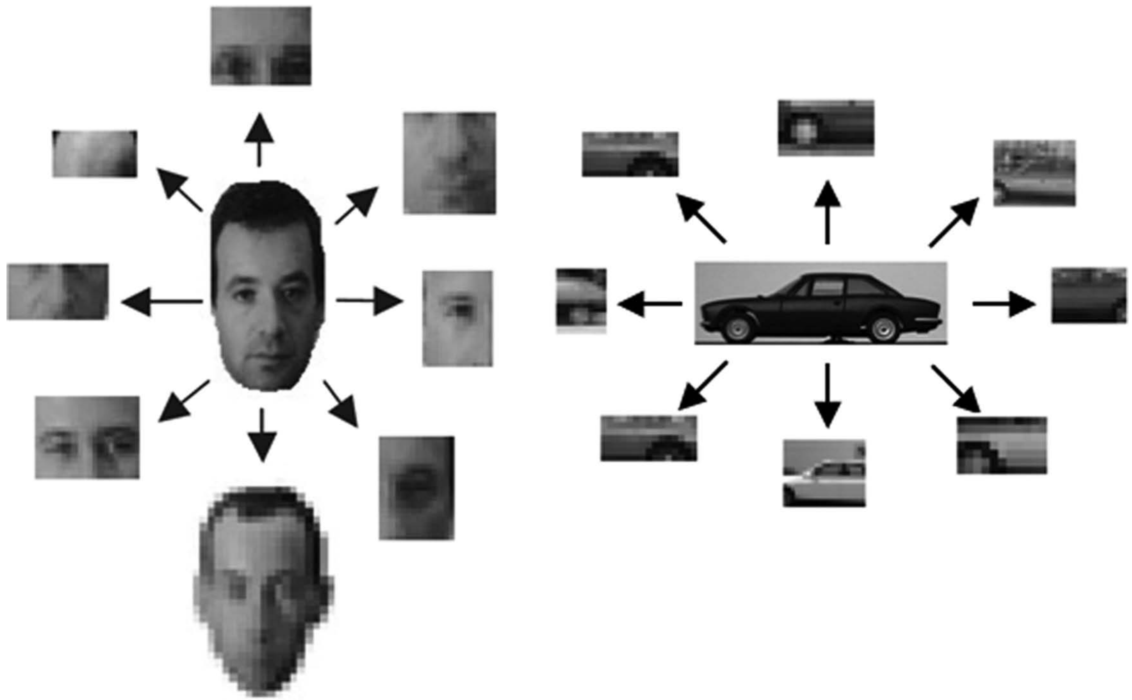


Figure 7.3 Illustration of view-based visual features of intermediate complexity maximize delivered information with respect to a basic-level category of objects. The figure shows examples of face fragments and car fragments (adapted with permission from S. Ullman, M. Vidal-Naquet, & E. Sali, 2002, "Visual features of intermediate complexity and their use in classification," *Nature Neuroscience*, *5*(7) © 2002 Nature Publishing Group). Zhang and Cottrell (2005) found somewhat larger and more complex view-based visual features for subordinate identification.

9780195309607_0197-0244_Gauthier_GAUT_Ch07 10/7/2009 16:59 Page:211

The approach proposed by Ullman et al. (2002) was intended to be a solution to basic-level classification—classifying an object as a face or a car—not more subordinate-level classifications—classifying an object as Barack Obama or a Toyota Prius. But recently, Zhang and Cottrell (2005) extended this approach to discover the image features that have maximal informativeness for subordinate-level classification. What they found was that these image features were larger and more complex than the features Ullman et al. reported for basic-level classifications. 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 the emergence of configural and holistic effects with expertise requires assembling hierarchies of features, not simply relating them in a single level of spatial relations (e.g., Gauthier & Tarr, 2002; Maurer, Le Grand, & Mondloch, 2002). Holistic effects emerge with perceptual expertise as larger and more complex view-based fragments, or even entire images, are learned.

MODELS OF PERCEPTUAL CATEGORIZATION

To recognize an object is to decide that its perceptual representation is similar to an object representation created and stored during some previous experience with that object. But to identify or categorize an object, its perceptual representation must be compared with a knowledge representation that summarizes what is known about the identity or category of that object. Ultimately, models of object processing and models of perceptual categorization both aim to explain how people recognize, identify, and categorize objects. But whereas models of object processing typically emphasize the nature of the perceptual representations created by high-level vision, models of perceptual categorization have focused more on the nature of the knowledge representations and decision processes underlying recognition, identification, and categorization (Palmeri & Gauthier, 2004). Perceptual categorization models often begin with relatively simplified assumptions about how objects are represented, commonly assuming that objects are represented in a multidimensional psychological space (Ashby, 1992) with visually similar objects close together in that space and visually dissimilar objects far apart in that space (similar multidimensional representations were adopted by "face space" theories discussed in the next section, e.g., Valentine, 1991). These multidimensional object representations are not chosen arbitrarily in a post hoc manner; they are typically derived a priori from known psychophysical mappings between physical and psychological dimensions or using various psychological scaling techniques (e.g., Nosofsky, 1992b). So while object processing models differ in how objects are perceptually represented, categorization models often assume the same multidimensional perceptual representations of objects. Categorization models differ in how knowledge about an object's identity and category are represented.

One key dimension on which categorization models differ is the abstraction of the category representations. A hallmark of visual cognition is generalization; even young children know when two visually similar but distinct 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. An early solution was to assume that 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. In a related vein, early theories assumed that people learn new categories by forming abstract logical rules, and research focused on what kinds of rules people found more or less difficult to learn (e.g., Bruner, Goodnow, & Austin, 1956; Hunt, Marin, & Stone, 1966; see also Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Nosofsky & Palmeri, 1998). Subsequent research instead assumed that people learned abstract category representations based on prototypes—statistical central tendencies of experienced category exemplars—rather than rules (e.g., Homa, 1978; Minda & Smith, 2000; Posner & Keele, 1968; Reed, 1972). Both rule-based and prototype-based theories assume that because category knowledge can be applied abstractly, the underlying category knowledge representations must themselves be abstract (Figure 7.4).

This solution is similar to the solution to the invariance problem in object recognition proposed by Biederman (1987). Objects differ in viewpoint. RBC achieves viewpoint invariance by constructing abstract perceptual representations that are invariant over object view. Objects from the same category look different. Prototype and rule models achieve class invariance by constructing abstract category representations that are invariant over category instance. But in the same way that view-based models of object processing can achieve viewpoint invariance using viewpoint-dependent representations (Bülthoff & Edelman, 1992; Poggio & Edelman, 1990; Tarr & Pinker, 1989), so-called "exemplar-based" models of categorization can achieve class invariance using instance-specific representations (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1984); both "views" and "exemplars" are representations tied to specific object experience. Computationally, there is a common solution to recognizing an object from a novel viewpoint and categorizing a novel instance of a category using experience-specific representations (Edelman, 1999).

As the name implies, exemplar models of categorization assume that categories are represented in terms of the specific exemplars that have been experienced. The perceptual representation of an object to be classified activates these stored exemplars depending on its similarity to those exemplars, with similarity a decreasing function of distance in multidimensional psychological space. The probability of classifying the object into a particular category depends on how similar it is to exemplars of that category relative to its similarity to exemplars of other categories (for details see Kruschke, 1992; Lamberts, 2000; Medin & Schaffer, 1978; Nosofsky, 1984; Nosofsky, 1992a). Exemplar models naturally account for many phenomena thought to

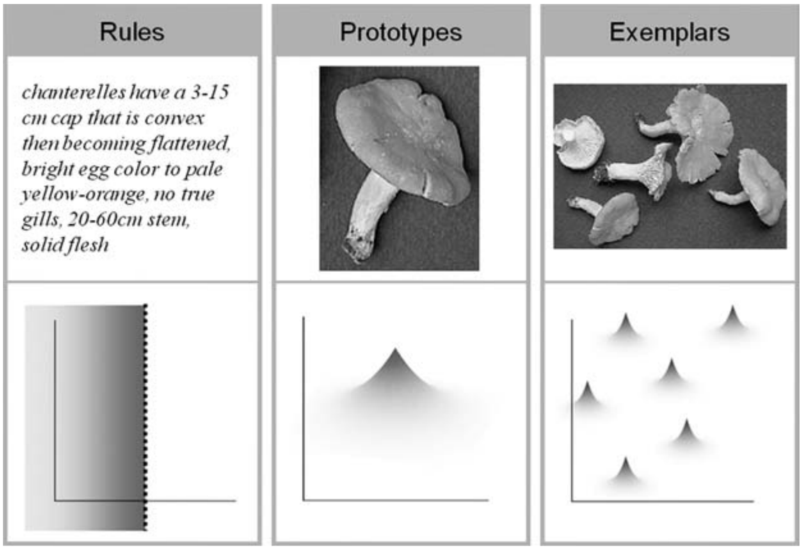


Figure 7.4 Illustration of rule-based, prototype-based, and exemplar-based category representations. Verbal rules for categorizing edible chanterelle from similar nonedible or even poisonous mushrooms can be quite complex (top row). As an illustration of a simple one-dimensional rule, the space of objects is carved into those defined by the rule in gray shading and everything else (bottom row). The most prototypical chanterelle (top row) would be an average of experienced exemplars. The prototype lies in the center of the space of category examples, with the generalization gradient around the prototype defining the typicality (bottom row). Knowledge of chanterelles can also be represented by the range of examples that have been experienced (top row). Categorization is determined by the similarity to stored exemplars in the space of possible objects (bottom row).

demonstrate the formation of abstract rules or prototypes (e.g., Busemeyer, Dewey, & Medin, 1984; Hintzman, 1986; Nosofsky, 1986; Shin & Nosofsky, 1992); for example, category prototypes are well classified because they are similar to many stored exemplars, without any need to additionally store an abstracted prototype explicitly (Palmeri & Nosofsky, 2001). A large body of research demonstrated the theoretical success of exemplar-based models in accounting for a range of categorization and related phenomena (e.g., Estes, 1994; Kruschke, 1992; Lamberts, 1995, 2001; Nosofsky, 1988; Nosofsky & Palmeri, 1997; Figure 7.5 illustrates the formal relationships between various models). Computationally, the exemplar representations in many exemplar models of categorization (e.g., Kruschke, 1992) are quite similar to the view representations in view-based models of object processing (e.g., Poggio & Edelman, 1990; Riesenhuber & Poggio, 1999). [3] As such, exemplar models of

[3] Exemplar models have also recently been shown to be computationally similar to certain popular machine learning algorithms (Jäkel, Schölkopf, Wichmann, 2007, 2008).

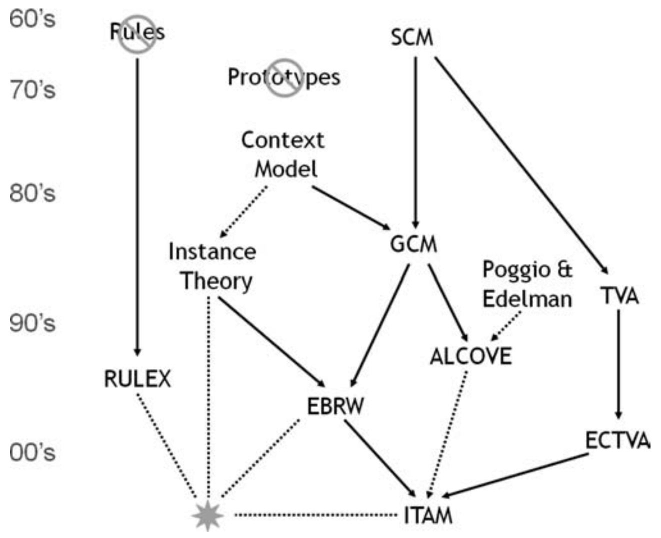


Figure 7.5 A family tree expressing the relations between a class of computational models over time. Models connected by solid lines have formal mathematical relationships. Models connect by dotted lines share computational principles. Important aspects of the development of perceptual expertise can be explained by conjoining aspects of rule-based categorization models like RULEX, theories of automaticity like instance theory, and exemplar-based categorization models like EBRW. These models also bear important formal relations to models of visual attention like ITAM. SCM = similarity choice model (Luce, 1959; Shepard, 1957); context model (Medin & Schaffer, 1978); GCM = generalized context model (Nosofsky, 1984, 1986); instance theory (Logan, 1988; Poggio and Edelman, 1990); TVA = theory of visual attention (Bundesen, 1990); ALCOVE = attention learning COVEring theory (Kruschke, 1992); RULEX = RULe-plus-EXception model (Nosofsky, Palmeri, & McKinley, 1994); EBRW = exemplar-based random walk model (Nosofsky & Palmeri, 1997; Palmeri, 1997); ECTVA = executive control of TVA (Logan & Gordon, 2001); ITAM = instance theory of attention and memory (Logan, 2002).

categorization share many of the same qualities—and shortcomings—of view-based models of object processing in terms of accounting for core phenomena of perceptual expertise as described in the last section.

Some defining characteristics of perceptual expertise entail comparisons between categorization and subordinate-level identification. On the one hand, categorization and identification seem diametrically opposed, with identification highlighting discrimination between stimuli and categorization rendering discriminable stimuli equivalent. Indeed, early work by Shepard and colleagues (Shepard & Chang, 1963; Shepard, Hovland, & Jenkins, 1961) suggested that the same exemplar generalization mechanism could *not* account jointly for categorization and identification performance. But Nosofsky (1984, 1986, 1987) showed how exemplar models could naturally account for both kinds of decisions using the same exemplar representations. The key insight was that exemplar generalization could be task dependent, unlike the task-independent exemplar generalization assumed by Shepard. Specifically, Nosofsky assumed that some psychological dimensions could be weighted more heavily than others depending on their diagnosticity for categorization (see also Kruschke, 1992; Lamberts, 2000; Nosofsky & Kruschke, 1992). Whereas all dimensions may be important for discriminating objects for purposes of identification, certain dimensions may be more (or less) relevant than others for categorizing objects; this makes exemplars that differ along nondiagnostic dimensions more similar than exemplars that differ along diagnostic dimensions. In addition to accounting for relations between identification and categorization, this dimensional weighting (called "dimensional selective attention" in these models) is necessary for exemplar models to account for the time course of learning categories; models without dimensional weighting have difficulty accounting for category learning (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Nosofsky & Palmeri, 1996). Many models of object processing (e.g., Riesenhuber & Poggio, 1999) also lack this dimensional weighting and assume task-independent generalization (Jiang et al., 2007), so it is quite possible that they will be unable to account for the full gamut of object categorization data either.

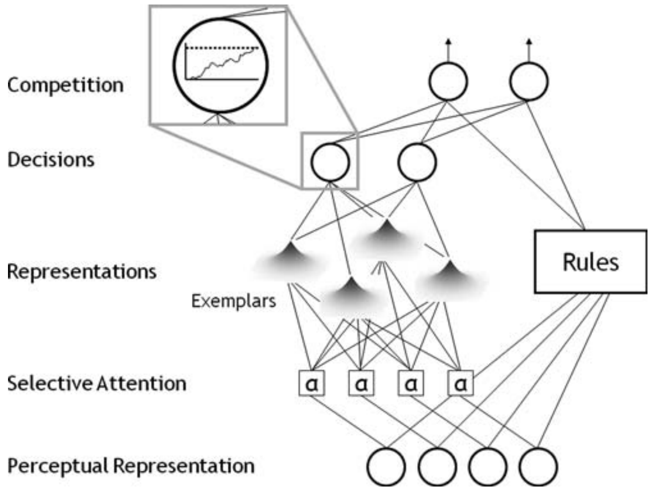
The flexibility imbued by dimensional selective attention in exemplar models seems to fly in the face of the limits on selective attention seen with perceptual expertise. While novices can attend selectively to part of a complex object, experts show interference from variation in an irrelevant part. Experts represent objects holistically. Perhaps what's key to resolving this paradox is that most experiments demonstrating holistic processing have made irrelevant (temporarily) a part of an object that has always in the past been diagnostic for identification or categorization. From a subject's perspective, for decades both the top and bottom parts of a face have been important for face recognition; for car experts both the top and the bottom of a car have always been relevant for telling apart car models. Now in the experiment, they are told that the bottom is no longer relevant for some decision they are asked to do (for the next few minutes). Like novices they are able to do the task, and attend to the top while ignoring the bottom. However, selective attention is never perfect, especially for spatially contiguous parts. Decades of experience have caused long-term exemplar representations of faces or cars to include both the top and the bottom because both parts are critical to successful identification or categorization of faces or cars. So even a small failure to ignore the irrelevant part can end up having a large effect on observed behavior, manifested by an interference by the irrelevant part for experts. It's likely that the flexibility in selective attention cannot override extensive past experience that has created more permanent representations (see Gauthier & Palmeri, 2002; Palmeri et al., 2004), but this dynamic in exemplar models between extensive past experience and current task demands has not yet been fully explored.

Understanding dynamics has to play a key role in fully understanding perceptual expertise. Experts are fast. Several core features of perceptual expertise involve significant speedups in processing. Unlike many other models of categorization and object processing, exemplar-based models of categorization have taken time seriously (e.g., Cohen & Nosofsky, 2003; Lamberts, 2000; Logan, 1988; Nosofsky & Palmeri, 1997; Palmeri, 1997; see also Ashby et al., 2007). These models make specific assumptions about the time for perceptual processing, the time it takes to match perceptual representations with stored representations in memory and how those times change with experience, and the time to accumulate evidence from these matches in order to make rapid perceptual decisions.

For example, the exemplar-based random walk (EBRW) model (Nosofsky & Palmeri, 1997; Palmeri, 1997) assumes that when an object is presented, its perceptual representation is used as a probe of stored exemplars. These stored exemplars race to be retrieved with rates proportional to their similarity to the presented object. The winning exemplar provides incremental evidence for a particular categorization decision. Retrieval is noisy, so multiple exemplar retrievals are needed to obtain reasonably accurate decisions. The results of these multiple retrievals are accumulated over time, with each potential decision associated with a different accumulator. Whichever accumulator reaches its threshold first determines which decision is made and when it is made. EBRW is a member of a family of stochastic (noisy) accumulator models (random walk models and diffusion models) that provide excellent accounts of things like speed–accuracy tradeoffs and shapes of response time distributions (e.g., Ratcliff, 1978; Ratcliff & Rouder, 1998), and these models appear to have some grounding as the neural basis of perceptual decisions (Boucher et al., 2007; Schall, 2004; Smith & Ratcliff, 2004). What distinguishes EBRW from more general diffusion-type models is that it provides a specific theory of the evidence that drives the stochastic accumulation of evidence to a threshold.

Objects that are hard to categorize, because they are similar to objects in other categories, are categorized slowly. According to EBRW, confusable objects will tend to retrieve objects from multiple competing categories, causing a stochastic accumulation that vacillates between competing alternatives, causing longer response times. But even difficult-to-categorize objects will be categorized more quickly and more accurately as people develop perceptual expertise. EBRW assumes that with more and more experience with exemplars, more and more exemplar information is stored in memory (Logan, 1988). As more exemplar information is stored in memory, the right exemplars are retrieved (Lamberts, 2000; Nosofsky & Alfonso-Reese, 1999). Exemplar retrieval also takes place ever more rapidly. More rapid retrieval causes more rapid accumulation of evidence to a threshold and faster decisions. EBRW naturally accounts for the ubiquitous power law of learning observed throughout the skill learning and expertise literatures (Logan, 1988, 1992; Newell & Rosenbloom, 1981; Palmeri, 1997; but see Heathcote et al., 2000; Palmeri, 1999; Rickard, 1997). With the sharpening of exemplar representations that comes with experience (Nosofsky, 1987), EBRW naturally accounts for the relative speedups in basic-level categorization versus subordinate-level identification seen over the development of perceptual expertise. The entry-level shift with expertise emerges directly from quantitative changes in exemplar representations rather than any qualitative shift in processing strategies (Mack et al., 2007; Palmeri et al., 2004).

While exemplar models have provided compelling accounts of a range of phenomena, there has been growing interest in reexamining the potential role of more abstract forms of category representation, such as rules or prototypes (but see Nosofsky & Johansen, 2000). Various hybrid theories (see Figure 7.6) have been proposed that involve mixtures of rules and exemplars (e.g., Anderson & Betz, 2001; Erickson & Kruschke, 1998; Johansen & Palmeri, 2002; Noelle & Cottrell, 1996; Nosofsky, Palmeri, & McKinley, 1994; Palmeri, 1997; Smith, Patalano, & Jonides, 1998; Thomas, 1998), prototypes and exemplars (e.g., Anderson, 1990; Love, Medin, &



AQ1

Figure 7.6 An illustration of a broad class of categorization models (e.g., Ashby et al., 1998; Erickson & Kruschke, 1992; Johansen & Palmeri, 2002; Kruschke, 1992; Nosofsky & Palmeri, 1997; Palmeri, 1997). Objects are represented along multiple perceptual dimensions and features (Perceptual Representation). Categories can be represented using exemplars or rules. Along the exemplar route, dimensions can be selectively attended according to their diagnosticity (e.g., Gauthier & Palmeri, 2002; Kruschke, 1992). As such, exemplars are activated according to their similarity to the presented object, but with diagnostic dimensions carrying more weight than nondiagnostic dimensions (Nosofsky, 1984). Depending on the model, exemplars are associated with learned categories by weights tuned using Hebbian or error-driven learning mechanisms. Categorization decisions can be driven by a variety of accumulation of evidence models (e.g., Nosofsky & Palmeri, 1997; Smith & Ratcliff, 2004). A variety of processes can be used to resolve the competition between rule-based and exemplar-based categorization decisions.

Gureckis, 2004; Smith & Minda, 1998), and various kinds of linear and nonlinear decision boundaries (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; see also Ashby, Ennis, & Spiering, 2007). Let us consider the possible interplay between abstract knowledge and more specific knowledge as it might play out in a domain of perceptual expertise. We can imagine that a novice searching the woods for prized Chanterelle mushrooms must refer to a set of fairly complex rules for telling them apart from many visually similar, yet quite poisonous species, such as the Jack O'Lantern mushroom (Phillips, 1991). Although these rules may become internalized, categorizing mushrooms as edible versus poisonous (without requiring reference to a field guide) may still appear to involve deliberate use of explicit rule-based knowledge. With experience, however, a mushroom gatherer eventually seems to shift from this potentially slow, deliberate, attention-demanding mode of categorizing to a far more rapid and automatic mode of processing that seems to characterize more expert-like performance (it's likely that after finding the mushroom at a glance, they may still check whether they are correct by using the well-known rules since the consequences of misclassification could be dire). Understanding the kinds of representational changes that allow someone to become a skilled mushroom gatherer who can recognize the prized Chanterelle so quickly and effortlessly without needing to make recourse to explicit rules is a key question of perceptual expertise. These hybrid categorization models have attempted to explicitly understand these changes in category knowledge.

One way of thinking about these shifts is to view categorization as just another domain in which people develop cognitive skills with experience. According to Logan's (1988) instance theory, automaticity in a range of skills is attributed to a shift from strategic and algorithmic processes, such as the use of explicit rules, to the retrieval of exemplars from memory. Automaticity is a memory phenomenon. Exemplars are memories. Could such shifts characterize the development of expertise in perceptual categorization whereby people initially use simple rules to categorize objects but eventually come to rely on similarity-based retrieval of exemplars? Palmeri (1997, 1999; see also Palmeri et al., 2004) found evidence for shifts from rules to exemplars in a paradigm in which subjects were supplied an explicit rule for initially classifying objects into different categories. In a different paradigm, Brooks and colleagues (Allen & Brooks, 1991; Regehr & Brooks, 1993) found evidence for intrusions of similarity-based retrieval even when subjects were supplied an explicit categorization rule. But in many experimental paradigms and in many real-world situations, people are not supplied categorization rules prior to learning about categories of objects. Johansen and Palmeri (2002) found that in situations where no explicit rule is provided, people still seem to adopt an analytic strategy of developing simple rules at the outset of category learning. However, over the course of learning, these rules eventually give way to processes more akin to similarity-based exemplar retrieval. According to this view, because much of expert performance is based on similarity, generalization to new objects can be rather limited. Experts may sometimes be able to turn to rules when exposed to truly novel objects in novel situations, but then performance may not have the automaticity and fluency that often distinguishes experts from novices.

MODELS OF FACE RECOGNITION

Much of the research described throughout this volume begins with the hypothesis that what makes faces special is not that they are faces per se, but that faces are a domain of perceptual expertise; needless to say, this is seen as a controversial hypothesis by some researchers (Farah et al., 1998; McKone, Kanwisher, & Duchaine, 2007; Robbins & McKone, 2007), as addressed in some detail in other chapters. Starting with the hypothesis that expert face processing shares important computational principles with other domains of perceptual expertise, models of face processing and face recognition can be a fruitful starting point for a computational understanding of other domains of perceptual expertise. Face recognition differs from common object recognition in the type of problem: object recognition typically requires *ignoring* within-class variability, in order to recognize all of the variants of a class, while face recognition requires paying a great deal of attention to within-class variability, because that is the signal that separates the individual members of the class. One could therefore imagine treating face recognition as another type of object recognition, just one level down the hierarchy of objects, but in this case, all of the objects share overall shape and parts. This is what makes the classic RBC theory far from able to account for expertise.

We will begin this section by describing some classic and more recent models of face recognition and then turn to work that makes direct links between face recognition and other kinds of expert recognition within the same general processing architecture.

There are at least three kinds of face recognition models: psychological models, computer vision models, and models that try to combine these approaches to generate computational cognitive models of face recognition. Within the latter class, there are models that emphasize the relationship to the brain, and attempt neural plausibility, while others abstract away from the neural architecture.

We will start with the best-known psychological model of face recognition, the classic Bruce and Young (1986) model (Figure 7.7). While it is not a computational process model, it serves as a useful starting point and gave rise to a later implementation. Bruce and Young's model, which they termed a "functional"[4] model, was designed to account for a wide range of behavioral and neuropsychological data. They began by distinguishing seven different

[4] Here, "functional" means that the model tries to account for all of the functions required, but the model is not computational in the sense that it is not implemented on a computer.

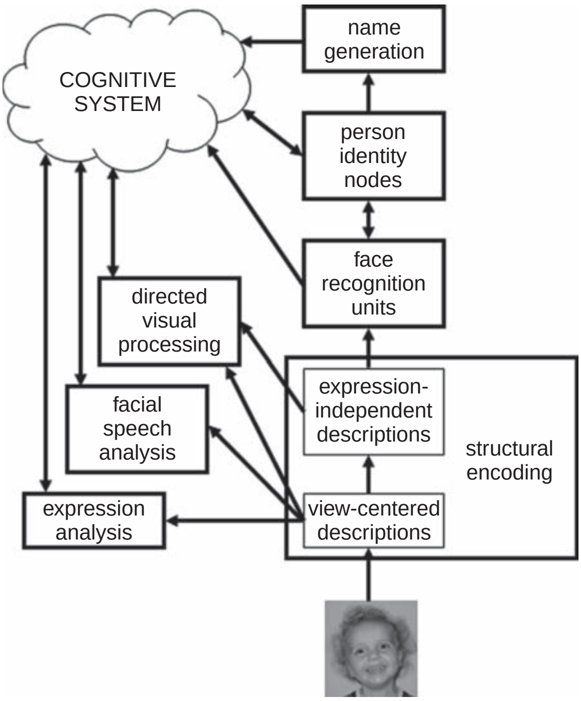


Figure 7.7 Illustration of the Bruce and Young model (reproduced with permission from Bruce & Young, 1986, © The British Psychological Society).

"codes" or representations that were necessary in order for their model to account for this data. In the model, face processing starts with a pictorial code, corresponding to Marr's (1982) "primal sketch." This is followed by a view-based code, analogous to Marr's "2.5D sketch", useful for face recognition memory of unfamiliar faces, recognition of facial expressions, and visual speech coding, all of which branch off at this point. For identifying known faces, the view-centered description is further transformed into an expression-independent representation, corresponding to some extent to Marr's "3D object representation" level, allowing the recognition of familiar faces in new situations (e.g., in novel lightings, orientations, etc.). However, the structural codes were considered to be very different than those posited by Marr, who was interested in visual object recognition; in face recognition, within-class discrimination is the primary problem. Hence, the structural encoding must have fine-grained configural information encoded. The structural encoding stage then activates face recognition units, which allow for recognition of a face as known, without necessarily recognizing who it is. This is the job of the person identity units, which activate semantic information about the person and the name unit for the person.

These various levels were deemed necessary to account for a great deal of data. For example, prosopagnosics would have damage somewhere along the stream that activated the face recognition units, without damage to the person identity nodes, since they can recognize people from voice or other means. The name units were differentiated from other semantic units because information about a person can be accessed without accessing their name. This model and its later implementations (Burton & Bruce, 1993; Burton et al. 1990; Burton et al., 1999) have continued to have a major influence on theories of and experiments in face recognition. However, no learning mechanism was included in the model, so it is hard to say how it would account for the acquisition of perceptual expertise.

Another influential psychological model is Tim Valentine's face space model (Valentine, 1991). His idea was that faces are represented as points in a high-dimensional space (which could be the equivalent of the structural encoding stage of the Bruce and Young model). In terms of a computational model, these dimensions could each correspond to a feature, and the actual point then specifies values for each feature. These kinds of representations are akin to the multidimensional representations used in many of the perceptual categorization models described in the previous section (e.g., Ashby, 1992; Lamberts, 1995, 2000; Nosofsky & Palmeri, 1997). Thinking of faces in this way leads to consideration of clusters in that space, for example, for unfamiliar-race faces versus familiar-race faces. Valentine posited that unfamiliar-race faces were clustered in a less differentiated ball in the face space, while familiar-race faces were more differentiated, explaining the so-called "other-race effect" (ORE) in face recognition. This model is highly compatible with current computational models that use patterns of activation over a set of units to represent faces. Indeed, such models have been used to explain the ORE in much the same way as Valentine envisioned it (e.g., Haque & Cottrell, 2005; O'Toole, Deffenbacher, Valentin, & Abdi, 1994).

Another influential psychological model is that of Farah, Wilson, Drain, and Tanaka (1998), in which they posited holistic representations for faces. The idea of a holistic representation is that the features of a face are connected in some way with one another, achieving a more important status than the representations of the individual parts. While they were somewhat agnostic concerning the actual form of this representation—it could correspond to whole-face templates or it could correspond to strong connectivity between the parts of a single face—the holistic representation can be used to explain the whole-face superiority effect. This effect, similar to the word superiority effect, corresponds to the ability of subjects to better discriminate a nose in the context of a face than a nose in isolation.

Computational models of face identification (in contrast with the face categorization models noted above) began with Takeo Kanade's doctoral thesis in 1973. His model was the first to actually use real images as inputs. It recognized faces by measuring distances between face features. Unlike many modern models, it actually had a top-down component that reanalyzed the face if there was not a good enough match. The next generation of models began with Kohonen's neural network model (Kohonen et al., 1977), which essentially used singular value decomposition to learn a map from pixels to names; another way of thinking about this model is that it was a linear neural network that learned to map from faces to names. This results in an appearance-based, holistic model. He showed that he could recognize faces in novel orientations by training the system on multiple orientations. This model thus was one of the first to show how an appearance-based model could generalize by interpolating between learned views.

The next set of models all began around 1990 when there was a sudden burst of interest in the use of principal components analysis (PCA) to represent faces. Turk and Pentland's (1991) "eigenface" model used principal components of gray scale images to learn a representation of faces (see Figure 7.8). These

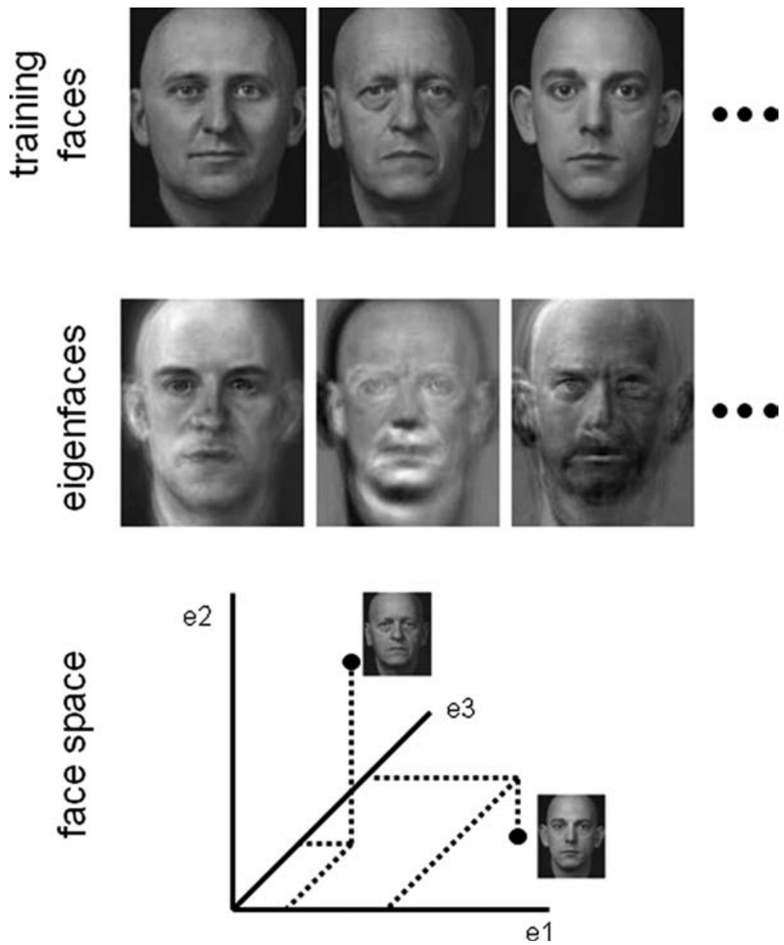


Figure 7.8 Examples of training faces (top row) and the first three eigenfaces (middle row) found using PCA (Adapted with permission from M. N. Dailey, G. W. Cottrell, and T. A. Busey, 1998, "Eigenfaces for familiarity," In *Proceedings of the Twentieth Annual Cognitive Science Conference*, © 1998 Lawrence Erlbaum Associates.). Faces are represented in terms of a linear combination of eigenfaces, generating a multidimensional face space representation (bottom row).

models are holistic in the sense that a face is represented by a weighted combination of whole-face features, that is, eigenfaces. Once the eigenfaces were extracted from a set of images, novel images could be categorized by projecting them onto the eigenfaces, and then labeling them with the identity of the closest projection from the training set (Figure 7.8). A version of this model won portions of Defense Advanced Research Projects Agency's (DARPA's) face recognition bake-off (Face Recognition Technology, FERET) in 1994 and 1995. A number of similar models based on neural networks were developed at this time, but they were considered cognitive models, and we discuss them below. In that section, we also discuss some of the interesting cognitive modeling properties of eigenfaces.

The model that performed the best overall in the FERET 1994 and 1995 tests was von der Malsburg's system (Okada et al., 1998). It used a deformable template that was fitted over the face. At each node of the template was the response of a set of Gabor filters (wavelets), which can be thought of as a kind of "zip code" for that portion of the face. There was one such template for each person in the training set. The links between the nodes can be thought of as "springs" that were stretched when matching a new face, which gave the model the ability to match faces that were displayed in quarter and full profile. It was this ability to match rotated faces that gave this system the edge over eigenfaces, which had no method for deforming them in order to match such rotations. This fitting process must be repeated for each template in the database, which is computationally more expensive than the nearest-neighbor technique used in the eigenface system. Given that the training faces were frontal, it is unclear how the eigenface method could be adapted to this test, without learning in advance how faces are transformed by rotations out of the image plane.

Around this same time, cognitive models based on PCA or their neural network equivalent, autoencoder networks, were being developed (Cottrell & Metcalfe, 1991; Fleming & Cottrell, 1990; Hancock et al., 1996; O'Toole et al., 1993). What was demonstrated in these early models is that there is a neurally plausible architecture for extracting principal components (eigenfaces when applied to faces) and that these representations are sufficient for a number of face processing tasks when used with a discriminative classifier, including identity, gender, and emotion classification (Padgett & Cottrell, 1997). These models made contact with the psychological literature in a variety of ways. For example, the other-race effect (ORE) was explained in a manner very close to that envisioned by Valentine by encoding a greater proportion of one race versus the other in an autoassociative matrix (O'Toole et al., 1994). The matrix then reproduced the less-represented-race faces with less fidelity than the faces of the more frequently encoded race. Padgett and Cottrell (1998) compared their model's categorization of morph stimuli to human responses, and by comparing internal representations, showed that the model could discriminate facial expressions better near a category boundary, as people do (Young et al., 1997). The PCA approach has also been combined with an interactive activation and competition version of the Bruce and

Young (1986) model to create a model that can explain face priming effects, such as Stan Laurel's face priming Oliver Hardy's, as well as face repetition effects (Burton et al., 1999). O'Toole and her colleagues have extended the idea of statistical analysis of 2D faces, as represented by PCA, to 3D face representations, and demonstrated the psychological validity of the representation through adaptation effects (Blanz et al., 2000).

More recently, neurophysiologically realistic models of general object processing (Riesenhuber & Poggio, 1999) have been applied to face recognition (Jiang et al., 2006). By so doing, these models make explicit the hypothesis that computations performed in the context of recognizing faces are qualitatively the same as those for recognizing common objects. As shown in Figure 7.9, this model starts with model cells similar to simple cells and has layers that alternate between layers of cells that encode complex cell responses by combining the responses of simple cells using a "max" operation, and layers of cells that combine the max cell responses into shape representations with linear rules. At the uppermost layer, cells with a Gaussian response function are trained to respond to particular individuals. While the model is illustrated for a face recognition problem, the basic structure of the model is identical for nonface object recognition as well. The training is via a brute force search of thousands of parameter settings until some are found that give response profiles that are not significantly different from the desired responses. After this training, the model has view-tuned units that have responses in quantitative agreement with human response profiles on the same data. This kind of modeling gives a proof of concept that such models can show both configural and featural effects without an explicit encoding of either. However, this kind of result can also be obtained by PCA models of face processing (Zhang & Cottrell, 2006), which use a more realistic training mechanism of learning the statistical structure of faces from the responses of Gabor filters used to model V1. However, the Jiang et al. model then makes predictions concerning the tuning and connectivity properties of FFA neurons that can be checked experimentally. Indeed, Jiang et al. (2007) use an fMRI-RA (rapid adaptation) paradigm to test the model's prediction that there should be an asymptote to discrimination performance that occurs when the population of neurons in the FFA responding to the two faces becomes disjoint. Thus, using a combination of computational modeling, fMRI, and behavioral techniques, they found that both human face discrimination performance and FFA activation can be quantitatively explained by a simple shape-based model in which human face discrimination is based on a sparse code of tightly tuned face neurons.

We now turn to a series of models developed originally by Fleming and Cottrell (1990) and Cottrell and Metcalfe (1991). The models reached their mature form in Dailey and Cottrell (1999) and then were further elaborated upon over the next 10 years by Cottrell and colleagues, although we will simply refer to these variants as "the model." Whereas Jiang et al. (2006) originated from a model of object processing (Riesenhuber & Poggio, 1999) that was later extended to face recognition, this model originated as a model of face

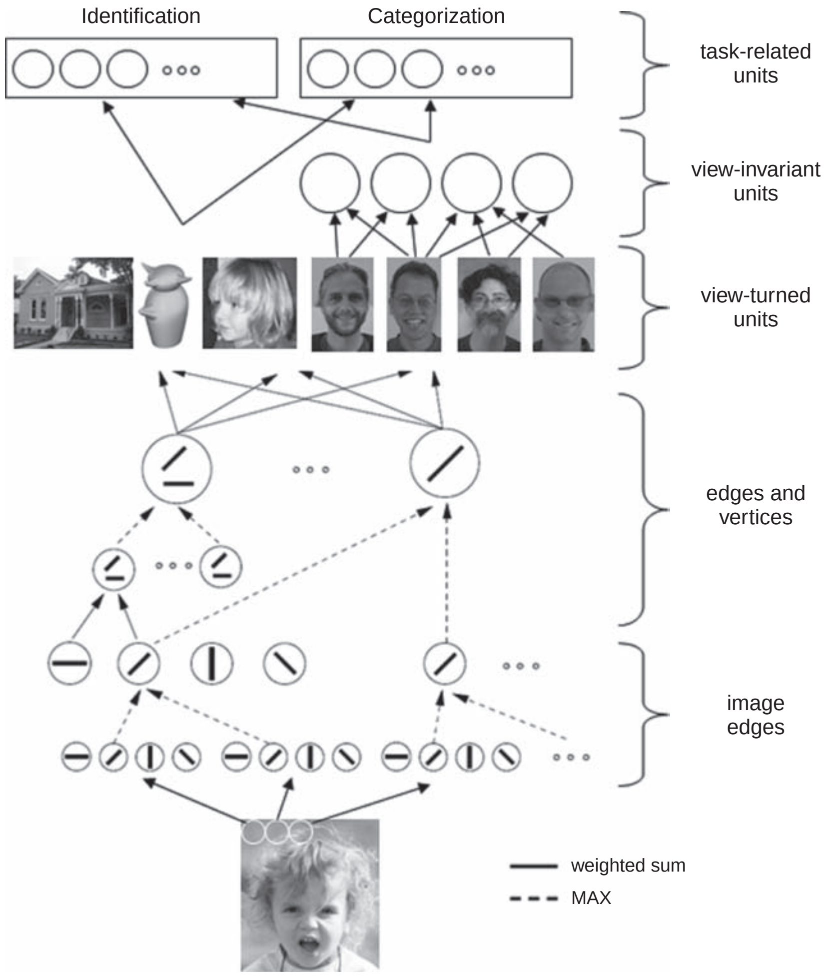


Figure 7.9 Illustration of the network architecture of the object and face recognition model developed by Riesenhuber and Poggio (1999, 2000; reproduced with permission from Jiang et al., 2006). An image is initially processed by a hierarchy of levels that extract edges and vertices, representing visual areas V1–V4 in cortex. Scale and translation invariance is performed by layers that calculated a weighted sum of inputs (solid lines) and layers that compute the unit with maximal activity (dotted lines). These feed into high-level view-tuned and view-invariant representations in IT. Task-specific units for making categorization and identification decisions are assumed to be in prefrontal cortex. Reprinted from X. Jiang, E. Rosen, T. Zeffiro, J. VanMeter, V. Blanz and M. Riesenhuber, 2006, "Evaluation of a Shape-Based Model of Human Face Discrimination Using fMRI and Behavioral Techniques," *Neuron, 50*(1), pp. 159–172. Copyright 2006. Used with permission from Elsevier.

recognition that was later extended to object recognition and perceptual expertise. The common theoretical insight is that qualitatively the same computational principles account for both object recognition and face recognition.

The basic structure is that of a neural network with four processing layers (Figure 7.10). The first layer represents the processing by V1, which is modeled as Gabor filters of five scales and eight orientations. The second

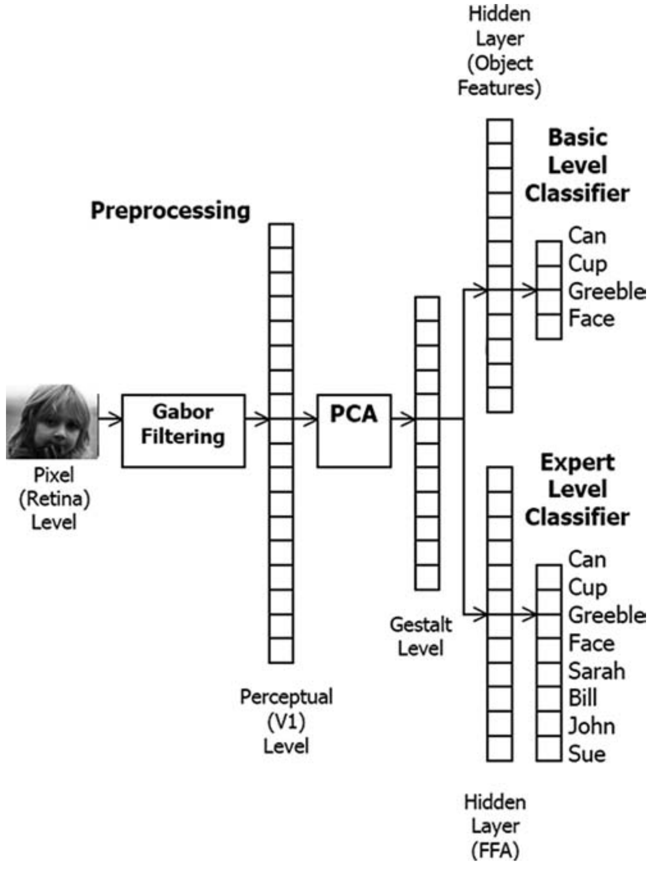


Figure 7.10 Illustration of the network architecture of a perceptual expertise model developed by Cottrell and colleagues (adapted with permission from Tong, Joyce, & Cottrell, 2005). The retinal image is preprocessed by a bank of Gabor wavelet filters at a number of different spatial scales and orientations to approximate the processing that occurs in early visual areas. The result is submitted to principal components analysis to reduce the dimensionality to a PCA projection (note that the PCA layer is "trained" on a different set of images from those used to test other predictions of the model). The resulting PCA representation is labeled as the Gestalt level in the figure. This representation is associated with basic-level and subordinate-level categories via a hidden layer representation, with weights trained using standard back-propagation. In this particular instantiation of the model, there are separate basic- and expert-level subnetworks, with the expert-level subnetwork hypothesized as a model FFA.

layer uses principal components to extract covariances from the Gabor filters and reduce the dimensionality of the representation. This can be thought of as the structural description layer from the Bruce & Young model and may correspond to later occipital layers. The next layer is an adaptive hidden layer trained by back-propagation to learn features that are appropriate for the task, and corresponds to either FFA when the model is trained at an expert level, or object layers in the ventral stream when trained to categorize. While back-propagation itself is not thought to be neurally plausible, more biologically realistic training methods exist that give rise to similar representations (Plaut & Shallice, 1993). This adaptive layer is often divided into several separate areas that are considered to be separate cortical areas competing to solve the tasks given to them, an assumption that was explicitly modeled by Dailey & Cottrell (1999). The fourth layer represents the output of the model, providing category or individual labels to the inputs, and likely corresponds to frontal areas (Palmeri & Gauthier, 2004; Riesenhuber & Poggio, 1999). The level of differentiation at this layer determines the fineness of the hidden layer features needed to discriminate the categories. Structurally, both the Cottrell model and a recent extension of the Riesenhuber and Poggio model (Serre et al., 2007) propose an initial stage of image filtering, followed by unsupervised learning of visual representations, followed by supervised learning of object categorizations.

One of the first issues addressed by this model was how the FFA develops in the first place (Dailey and Cottrell, 1999; see Figure 7.10). In their model, the banks of Gabor filters of different scales were treated as separate input channels. The representation of each scale was processed by a channel-specific PCA, so there was a low–spatial frequency input channel, up through a high–spatial frequency input channel, and PCA captured the covariances within each channel. It is well known that babies have poor contrast sensitivity in higher spatial frequencies, so using this scheme they could model the developmentally appropriate lower spatial frequency input to the cortex by attenuating the input from the higher spatial frequencies. The model then consisted of competing representations, which can be conceptualized as modeling the left and right hemispheres, one receiving relatively low spatial frequencies and one receiving relatively high spatial frequencies (Ivry & Robertson, 1998). The model was then trained to either categorize four classes of 12 objects each (five images per object) at a basic level, or to individuate one of the classes into its 12 identities, while continuing to simply categorize the other three classes of stimuli. To explain the differentiation of cortical areas for different tasks, the model assumed that different cortical areas competed for task through a gating network that fed more error back to the classifier with the lowest error—a rich-get-richer approach (this is also known as a "mixture of experts" model). Experiments using these differing levels of classification, as well as identical versus differential spatial frequency inputs, demonstrated that only in the case of different spatial frequency inputs to the two hidden layers, and the task of face identification, was there consistent specialization of one of the networks for face processing.

Consistent with behavioral data, the network receiving the lower spatial frequencies was always superior in the face identification task. Differential damage to the model produced prosopagnosia and object agnosia. Further experiments demonstrated that networks using lower spatial frequencies generalized to new images of the trained individuals much better than networks using high spatial frequencies, explaining why the specialization occurred, as better generalization also implies faster learning. They concluded that the model supported the hypothesis that something resembling a face processing "module" could arise as a natural consequence of the infant's developmental environment—poor visual acuity coupled with the goal of individuating people's faces—without being innately specified.

In terms of how this model accounts for the core features of perceptual expertise, we start with the last one, that experts tend to show activity in the fusiform face area (FFA) for objects of expertise, even after only 10 hours of training in the lab (Gauthier et al., 2000). If this is true, then the question arises as to why the FFA, an area that must start out as a face area in typically developing children, becomes recruited for other expertise tasks. One suggestion is that the FFA is the location of a *process*—fine-level discrimination of homogeneous categories. However, this is an analgesic answer—it makes us feel better, but it just covers up the problem by giving it a name. A modeler wants to know *how* this happens—what is the mechanism? In a series of papers, Cottrell and colleagues have addressed this as a question that can be solved by modeling (Joyce & Cottrell, 2004; Sugimoto & Cottrell, 2001; Tong et al., 2005; Tong et al., 2008). Following Dailey and Cottrell (1999), the model assumes that there are two networks (corresponding to two cortical areas) that have competed such that one becomes specialized for categorizing objects at the basic level (modeling LOC or other visual processing areas), and the other is specialized for the subordinate/expert level (i.e., it corresponds to the FFA). Aside from their tasks, the two networks have identical processing resources in their hidden layers. Despite this, the model displays the entry-level shift in response times—RT's (modeled as the uncertainty of the output) are just as fast for expert-level responses as for basic-level ones as in humans (Gauthier & Tarr, 1997). These two networks are then placed in a competition to learn a new category of objects at the expert level. The consistent result is that the expert network always learns the new expertise category first. This occurs even if the first category of expertise is not faces; hence, there is nothing special about faces per se. Indeed, a model trained to individuate donuts learns to individuate swords faster than a basic-level categorizer. Thus, the rather fanciful conclusion is that if our parents were donuts, the fusiform donut area would win the competition for new categories of expertise.

The model can then be analyzed to determine why the expert network won. The analysis reveals that the expert network spreads the stimuli into broader regions of representational space. That is, as in the Valentine model, faces are spread out in representational space, while objects are localized. This makes sense given that we need to ignore within-class

9780195309607_0197-0244_Gauthier_GAUT_Ch07 10/7/2009 16:59 Page:229

variability in order to categorize an object, while we need to amplify to within-class variability in order to individuate members of a category. The mapping from the inputs to the hidden units amplifies variability for expert objects, and this spreading transformation generalizes to new objects. Thus, the expert network has a head start in differentiating new objects from one another. The model also predicts that there will be relatively greater within-class variability in neural responses within the FFA as compared to object recognition areas, a prediction that can be tested neurophysiologically.

In terms of the other core features of perceptual expertise, the model accounts for speedups in processing due to the development of expertise. This phenomenon can be explained in terms of the connection strengths to the correct outputs growing with learning. The correct outputs are then turned on with less uncertainty, which leads to faster response times. The entry-level shift is explained in a similar fashion, combined with the nonlinearity in the outputs. As the connection strengths to the correct output are increased, the activity of the correct output increases, but it reaches a maximum level, so that uncertainty cannot be reduced beyond a floor. Hence, the subordinate-level category outputs reach the same level of activity as the category-level outputs and can go no further, so the corresponding response times are the same. Thus, this model shows that the entry-level shift can be explained as a continuous, quantitative change in the efficiency of processing over learning (Joyce & Cottrell, 2004). We should note that while the model accounts well for speedups with expertise, the model itself has no inherent temporal dynamics; response times are assumed to be inversely related to categorization certainty.

The model also shows the same pattern as experts in being unable to ignore variation in an unattended part (Cottrell et al., 2002). Specifically, the model can be modified to "pay attention" to half of the face by simply attenuating the input from part of the face. In these circumstances, the model shows interference from unattended parts of the face and also shows the same specific interaction between changes in identity and expression that is seen in human subjects (Calder et al., 2000). This form of holistic processing can be explained in terms of the whole-face templates that are developed at the PCA level (while this layer is like the eigenfaces in other models, in this model, it is "eigen-Gabor-faces," but the principle is the same). Suppose the model is presented with George Bush's face in the upper half of the input, and Al Gore's face in the lower half of the input. If there is a template (principal component) that preferably matches Al Gore's face, it will be partially matched by the input, and so will fire at a reduced level and pass this activation on to later layers. However, there is no way for these later layers to "know" what part of the input was matched – this template is voting for *all* of Al Gore's face, including his eyes, and so there is interference in recognizing the top half of the input as George Bush.

The model is also able to account for the way that experts generalize their knowledge, given the way they are trained to categorize the objects

(Nguyen & Cottrell, 2005). Specifically, subjects will fail to improve their discrimination of new examples of a category of objects when they have been trained only to give the same label (i.e., categorize) that set of objects (Tanaka, Curran, & Sheinberg, 2005). However, when they are trained to the individual level (in this case, the species of a type of bird, owls or wading birds), they then improve their discrimination in a graded fashion depending on the similarity of the novel examples to the trained categories. As with the explanation of the recruitment of the FFA for new objects of expertise, the result is explained in terms of within-class versus between-class variance in the representations required for basic-level categorization versus expert-level categorization. The internal representations are more differentiated for objects of expertise, and when new examples of the expert category are presented, they are more or less differentiated, depending upon their similarity to the trained expert class. On the other hand, when a large number of objects are categorized as “the same” by being given the same label (e.g., “owls” or “wading birds”), then the internal representations of those objects are pushed closer together by the learning mechanism and hence are less differentiated.

Finally, the model is impaired by inversion in the same way human subjects are (Ge et al., 2006; Mondloch et al., 2002). This can be explained entirely by the principal components level of the model (McCleery et al., 2008; Zhang & Cottrell, 2006). The principal components are sensitive to the orientation of the training data. If most of the training data is upright, and the objects of expertise have quite different statistics when inverted, the representation of inverted objects is relatively undifferentiated. The model is able to explain the difference in priming effects between inverted Chinese characters and faces, two types of expertise (McCleery et al., 2008), as well as the *development* of the sensitivity to configural, featural, and contour differences between upright and inverted faces (Zhang & Cottrell, 2006).

This model has not been tested against the data on the transition from verbalizable rules to automated processing, but connectionist mechanisms exist for making this transition (Noelle & Cottrell, 1996). Furthermore, the interference patterns seen in n-back tasks between different areas of expertise (Gauthier & Curby, 2005; Gauthier et al., 2003) have not been investigated. Adding residual activation patterns to the units in the model to implement the role of previous processing might accomplish this, but at this point, this is mere speculation. Even with the success of the model so far, it is important to systematically explore what phenomena it can’t account for.

SUMMARY AND CLOSING THOUGHTS

What makes an expert an expert? Experts perceive objects in their area of expertise better than novices. Do they perceive differently? Experts know more than novices; that’s what makes them experts, after all. They have a deeper and more fine-grained understanding about objects in their area of expertise. Is their knowledge fundamentally different from what novices know, or do they just know more and do they know it better? Experts make faster and more accurate decisions than novices. Are experts using information to make decisions in a way that novices simply cannot? Or are experts using that information more effectively?

Although perception, knowledge, and decisions by experts could be qualitatively different from novices—in that the mechanisms underlying expertise might operate under completely different principles from those underlying novice performance—computational models allow us to explore whether qualitative changes in behavior might emerge from quantitative changes in perception and memory over the course of learning. A number of key expertise effects can indeed be explained by using models of normal object recognition and perceptual categorization that learn representations, select representations, strengthen representations, and sharpen representations, without having to invoke qualitative changes. From a neural perspective, it is far easier to envision incremental changes in the brain from those that support novice object recognition to those that support expert object recognition than it is to envision creating a new special-purpose module for an expert domain. That certainly could happen, especially over the course of evolution or maybe over the course of development, and expertise does take a long time to develop (e.g., Ericsson & Lehmann, 1996; Ericsson et al., 2006). But so far we have not needed to invoke this kind of modular restructuring to explain many of the important behavioral or neural changes that take place over the development of perceptual expertise.

One basic question that seems not fully resolved is the specific locus for changes that take place with perceptual expertise. For example, are holistic processing effects that emerge with expertise best explained by changes in how objects are perceived, how decisions are made about those objects, or some combination of the two (Cheung et al., 2008; Richler et al., 2008)? We could also ask whether perceptual expertise is driven by changes in how we perceive objects or what we know and remember about objects, but the boundaries between perception and memory are rather fuzzy (e.g., Palmeri & Tarr, 2008). Category representations are strongly influenced by the particular category exemplars that have been experienced, and object representations are strongly influenced by the particular views of objects that have been experienced—so what you know about a category of objects and how you perceive objects from a well-known category depends on some kind of memory for what you've seen before. Open questions remain regarding issues such as how densely the space of experienced exemplars is represented (e.g., Ashby & Waldron, 1999; Palmeri et al., 2004), whether view-based representations are full templates (e.g., Riesenhuber & Poggio, 1999), view-based parts (e.g., Ullman et al., 2002), or somewhere in between (Zhang & Cottrell, 2005), and the extent to which representations are localist (e.g., Ashby & Waldron, 1999; Kruschke, 1992) versus distributed (e.g., Dailey & Cottrell, 1999).

Models need to take time seriously. As we documented, many of the key phenomena of perceptual expertise involve time. It is not sufficient to simply correlate the output activations of a model with mean observed response times. Trade-offs between speed and accuracy are far more complex and far more interesting (e.g., Lamberts, 2000; Mack et al., 2007, 2008). There are important details in distributions of response times that are not captured by simply examining the mean or the median RT (e.g., Townsend, 1990), so models should produce distributions of RTs (e.g., Lamberts, 2000; Nosofsky & Palmeri, 1997; Palmeri, 1999). Behavior changes with the temporal dynamics of the stimulus (e.g., Grill-Spector & Kanwisher, 2005; Mack et al., 2008), so models need to be dynamic. The precise timing of neural events is the cornerstone of neurophysiological studies (e.g., Anderson & Sheinberg, 2008; Mruczek & Sheinberg, 2008) and ERP studies (e.g., Scott et al., 2006; Tanaka & Curran, 2001), and techniques are emerging to significantly improve the temporal resolution of fMRI (e.g., Dux, Ivanoff, Asplund, & Marois, 2006). Process models must not only account for the timing of behavior, but the precise timing revealed by the neural markers as well (e.g., see Boucher et al., 2007). While efforts have been made to develop models of perceptual categorization and perceptual decision making that take time seriously, many object processing models are static (e.g., Dailey & Cottrell, 1999; Riesenhuber & Poggio, 1999). Given the growing body of cognitive neuroscience data on the timing of object recognition and how it changes with perceptual expertise, models of high-level vision need to incorporate these temporal dynamics.

Despite the plethora of models we have reviewed in this chapter, there is a surprising amount of theoretical convergence toward a comprehensive computational model of perceptual expertise. Over the years, research in object recognition and perceptual categorization has approached questions of visual cognition with somewhat surprising independence yet has arrived at theories that share important properties (Palmeri & Gauthier, 2004). A challenge will be putting the pieces together. The model of object recognition proposed by Riesenhuber and Poggio (1999; Serra et al., 2007) shares important computational principles of exemplar category representations as models of categorization proposed by Nosofsky, Kruschke, and others (e.g., Kruschke, 1992; Nosofsky & Kruschke, 1992). Categorization models have stressed the critical importance of learned selective attention to diagnostic properties of objects relevant to particular categorizations (e.g., Nosofsky, 1984, 1986) and have played close attention to how perceptual decisions are made (e.g., Lamberts, 2000; Nosofsky & Palmeri, 1997), neither of which have been the focus of object processing models. Object processing models have tried to make reasonable assumptions about lower-level visual processing (Dailey & Cottrell, 1999; Riesenhuber & Poggio, 1999, 2001), whereas categorization models typically start with representations suggested by careful psychophysical scaling studies without worrying about how those representations arise, probably at their detriment.

A comprehensive model of perceptual expertise must incorporate all of these various components. It's probably clear from our review that we believe this comprehensive model of perceptual expertise will also be a comprehensive model of normal object recognition and categorization as well, with expertise merely the end point of normal learning. One challenge in this endeavor is balancing the need for a model that spans perception, memory, and decisions, accounts for existing data, and predicts new results with an honest appraisal of the complexity that might encumber such an omnibus model. While having a model that works in the real world and behaves like people do is the goal of machine learning, artificial intelligence, and robotics, it's not the ultimate goal of psychology and neuroscience. Arguably, having a model that really "works" on real images in real time has an advantage over one that is more theoretical. But if such a model is burdened with so many ad hoc processing assumptions and freely tuned parameters in order to make it "work," then it fails as a viable and testable theory that can be falsified. Our challenge then is to balance these competing goals as we move toward more powerful and comprehensive theories of perceptual expertise.

REFERENCES

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

Anderson, J. R. (1976). Language, memory, and thought. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Betz. J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review, 8,* 629–647.

Anderson, B., & Sheinberg, D. L. (2008). Effects of temporal context and temporal expectancy on neural activity in inferior temporal cortex. *Neuropsychologia, 46*(4), 947–957.

Ashby, F. G. (1992). Multidimensional models of perception and cognition. Hillsdale, NJ, Lawrence Erlbaum Associates, Inc.

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

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

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

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

Bishop, C. M. (1995). Neural networks for pattern recognition. London: Oxford University Press.

Blanz, V., O'Toole, A. J., Vetter, T., & Wild, H. A. (2000). On the other side of the mean: The perception of dissimilarity in human faces. *Perception, 29,* 885–891.

Boucher, L., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2007). Interactive race model of countermanding saccades. *Psychological Review, 114,* 376–397.

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

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). A study of thinking. Oxford, England: Wiley.

Bülthoff, H. H., & 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.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

Burton, A. M., & Bruce, V. (1993). Naming faces and naming names: Exploring an interactive activation model of person recognition. *Memory, 1,* 457–480.

Burton, A. M, Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science, 23,* 1–31.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81,* 361–380.

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

Busemeyer, J. R., & Wang, Y. (2000). Model comparisons and model selections based on the generalization criterion methodology. *Journal of Mathematical Psychology, 44,* 171–189.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 527–551.

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1327–1336.

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

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

Cottrell, G. W., Branson, K., & Calder, A. J. (2002). Do expression and identity need separate representations? In *Proceedings of the 24th Annual Cognitive Science Conference, Fairfax, Virginia*. Mahwah, NJ: Lawrence Erlbaum.

Cottrell, G. W., & Metcalfe, J. (1991). EMPATH: Face, gender and emotion recognition using holons. In R. P. Lippman, J. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems, Vol. 3* (pp. 564–571), San Mateo, CA: Morgan Kaufmann.

Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recognition in modular neural network models. *Neural Networks, 12,* 1053–1073.

Dailey, M. N., Cottrell, G. W., & Busey, T. A. (1998). Eigenfaces for familiarity. In *Proceedings of the Twentieth Annual Cognitive Science Conference*. Madison, WI, Mahwah: Lawrence Erlbaum.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). Empath: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience, 14,* 1158–1173.

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

Dux, P. E., Ivanoff, J. G., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron, 52,* 1109–1120.

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

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

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

Ericsson, K. A., Charness, N., Feltovich, P., & Hoffman, R. R. (2006). *Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptations to task constraints. *Annual Review of Psychology, 47,* 273–305.

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

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

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

Fleming, M., & Cottrell, G. W. (1990). Categorization of faces using unsupervised feature extraction. *Proceedings of the International Joint Conference on Neural Networks, 2,* 65–70.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making and reversal. *Psychological Review, 113,* 300–326.

Furl, N., Phillips, P. J., & O'Toole, A. J. (2002). Face recognition algorithms as models of the other race effect. *Cognitive Science, 96,* 1–19.

Gauthier, I., & Curby, K.M., (2005). A perceptual traffic-jam on highway N170: Interference between face and car expertise. *Current Directions in Psychological Science, 4,* 30–33.

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

Gauthier, I., & Palmeri, T. J. (2002). Visual neurons: Categorization-based selectivity. *Current Biology, 12,* R282–R284.

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

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

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

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C. & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12,* 495–504.

Ge, L., Wang, Z., McCleery, J., & Lee, K. (2006). Activation of face expertise and the inversion effect. *Psychological Science, 17*(1), 12–16.

Golomb, B. A., Lawrence, D. T., & Sejnowski, T. J. (1991). Sexnet: A neural network identifies sex from human faces. In R. P. Lippman, J. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems, Vol. 3* (pp. 572–577), San Mateo, CA: Morgan Kaufmann.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32,* 108–154.

Grill-Spector, K., & Kanwisher, N. (2005). 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., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7,* 555–562.

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & Cognition, 24,* 26–40.

Haque, A., & Cottrell, G. W. (2005). Modeling the other race advantage. In *Proceedings of the 27th Annual Cognitive Science Conference, La Stresa, Italy*. Mahwah, NJ: Lawrence Erlbaum.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). Repealing the power law: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7,* 185–207.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.

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

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

Hsiao, J., & Cottrell, G. W. (2008). Not all expertise is holistic, but it may be leftist: The case of Chinese character recognition. *Psychological Science, 10,* 998–1006.

Homa, D. (1978). Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 407–416.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

Hunt, E. B., Marin, J., & Stone, P. J. (1966). *Experiments in induction*. New York: Academic Press.

Ivry, R., & Robertson, L. C. (1998). The two sides of perception. Cambridge, MA: MIT Press.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2007). A tutorial on kernel methods for categorization. *Journal of Mathematical Psychology, 51,* 343–358.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review, 15,* 256–271.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron, 50,* 159–172.

Jiang, X., Bradley, E., Rini, R.A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron, 53,* 891–903.

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

Johnson, K. E., & 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., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16,* 243–275.

Joyce, C., & Cottrell, G. W. (2004). Solving the visual expertise mystery In H. Bowman & Labiouse, C. (Eds.), *Connectionist models of cognition and* *perception II: Proceedings of the eighth neural computation and psychology workshop*. Hackensack, NJ: World Scientific.

Kanade, T. (1973). *Picture processing system by computer complex and recognition of human faces.* Unpublished doctoral thesis, Department of Information Science, Kyoto University.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131,* 451–476.

Kohonen, T., Lehtio, P., Oja, E., Kortekangas, A., & Makisara, K. (1977). Demonstration of pattern processing properties of the optimal associative mappings. In *Proceedings of the International Conference on Cybernetics and Society*, Washington, D.C.

AQ2

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

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

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

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

Logan, G. D. (1992). Shapes of reaction time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 883–914.

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

Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology, 55,* 207–234.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108,* 393–434.

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

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

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Reviews of Psychology, 46,* 1–26.

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review, 15,* 28–35.

Mack, M. L., Richler, J. J., Palmeri, T. J., & Gauthier, I. (in press). Categorization. To appear in G. G. Berntson & J. T. Cacioppo (Eds.), *Handbook of neuroscience for the behavioral sciences.*

Mack, M. L., Wong, A. C.-N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2007). Unraveling the time-course of perceptual categorization: Does fastest mean first? *Proceedings of the Twenty-Ninth Annual Meeting of the Cognitive Science Society.* New York: Erlbaum.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.

Marr, D., & 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.

Maurer, D., LeGrand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260.

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

McCleery, J. P., Zhang, L., Ge, L., Wang, Z., Christiansen, E. M., Lee, K., and Cottrell, G.W. (2008). The roles of visual expertise and visual input in the face inversion effect: Behavioral and neurocomputational evidence. *Vision Research, 48,* 703–715.

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

McKone, E., Kanwisher, N., & Duchaine, B. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11,* 8–15.

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

Minda, J. P., & Smith, J. D. (2000). 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.

Mondloch, C. J., LeGrand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31,* 553–566.

Moore, E. F. (1956). Gedanken-experiments on sequential machines. *Automata Studies, Annals of Mathematical Studies, 34,* 129–153.

Mruczek, R., & Sheinberg, D. L. (2008). Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. *Journal of Neuroscience, 27*(11), 2825–2836.

Murphy, G. L., & Wright, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 144–155.

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

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Nguyen, N., & Cottrell, G. W. (2005). *Owls and wading birds: Generalization gradients in expertise.* Proceedings of the 27th Annual Cognitive Science Conference, La Stresa, Italy. Mahwah, NJ: Lawrence Erlbaum.

Noelle, D. C., & 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,* 39–57.

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

Nosofsky, R. M. (1992a). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of* *perception and cognition.* Scientific psychology series (pp. 363–393). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

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

Nosofsky, R. M. (1998). Optimal performance and exemplar models of classification. In M. Oaksford and N. Chater (Eds.), *Rational models of cognition*, London: Oxford University Press.

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

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

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7,* 375–402.

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

Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review, 3,* 222–226.

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

Nosofsky, R. M. (1998). Optimal performance and exemplar models of classification. In M. Oaksford & N. Chater, (Eds.), *Rational Models of Cognition*. London: Oxford University Press.

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

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

Okada, K., Steffans, J., Maurer, T., Hong, H., Elagin, E., Neven, H., & Malsburg, C. v. d. (1998). The Bochum/USC face recognition system and how it fared in the FERET phase III test. In H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang (Eds.), *Face recognition: From theory to applications* (pp. 186–205), Berlin: Springer-Verlag.

O'Toole, A. J., Abdi, H., Deffenbacher, K., & Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America, 10,* 405–411.

O'Toole, A. J., Deffenbacher, K., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22,* 208–224.

Padgett, C., & Cottrell, G. W. (1997). Representing face images for emotion classification. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems 9* (pp. 894–900). Cambridge, MA: MIT Press.

Padgett, C., & Cottrell, G. W. (1998). A simple neural network models categorical perception of facial expressions. In *Proceedings of the Twentieth Annual Cognitive Science Conference*. Madison, WI, Mahwah, 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. (1999). Theories of automaticity and the power law of practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 543–551.

Palmeri, T. J., & Flanery, M. A. (2002). Memory systems and perceptual categorization. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp. 141–189), New York, NY: Academic Press.

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

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

Palmeri, T. J., & Tarr (2008). Object recognition and long-term visual memory for objects. In S. Luck & A. Hollingsworth (Eds.), *Visual memory*. London: Oxford University Press.

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

Phillips, R (1991). *Mushrooms of North America*. Boston, MA: Little, Brown and Company.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review, 113,* 57–83.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17,* 291–321.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Plunkett, K., & Juola, P. (1999) A connectionist model of English past tense and plural morphology. *Cognitive Science, 23,* 463–490.

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

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

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

Ratcliff, R., & Rouder, J. N. (1998). Modeling response time for decisions between two choices. *Psychological Science, 9,* 347–356.

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

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

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92–114.

Richler, J., Gauthier, I., Wenger, M., & Palmeri, T. J. (2008). Holistic processing of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 328–342.

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

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

Robbins, R. A., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103,* 34–79.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367.

Rossion, B., Kung, C.-C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences, 101,* 14521–14526.

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

Scott, L., Tanaka, J. W., Sheinberg, D., & Curran, T. (2006). A reevaluation of the electrophysiological correlates of expert object processing. *Journal of Cognitive Neuroscience, 18,* 1453–1465.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Science, 104,* 6424–6429.

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

Shepard, R. N. (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., & Chang, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology, 65,* 94–102.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75* (13, Whole No. 517).

Shillcock, R., & Monaghan, P. (2001). The computational exploration of visual word recognition in a split model. *Neural Computation, 13,* 1171–1198.

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

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

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

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative mechanisms of categorization. *Cognition, 65,* 167–196.

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

Sugimoto, M., & Cottrell, G. W. (2001). Visual expertise is a general skill. In *Proceedings of the 23rd Annual Cognitive Science Conference, Edinburgh, Scotland*, pp. 994–999. Mahwah, NJ: Lawrence Erlbaum.

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 M. Fahle & T. Poggio (Eds.), *Perceptual learning* (pp. 67–82). Cambridge, MA: MIT Press.

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

Tanaka, J. W., Curran, T., & Sheinberg, D. (2005). The training and transfer of real-world, perceptual expertise, *Psychological Science, 16,* 141–151.

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

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 & Review, 2,* 55–82.

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

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

Thomas, M. S. C. & de Wet, N. M. (1998). Stochastic double dissociations in distributed models of semantic memory. In G. Humphreys & D. Heinke (Eds.), *Proceedings of the 5th Neural Computation and Psychology Workshop*. London: Springer.

Thomas, R. D. (1998). Learning correlations in categorization tasks using large, ill-defined categories. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 119–143.

Tong, M. H., Joyce, C. A., & Cottrell, G. W. (2005). Are Greebles special? Or, why the Fusiform Fish Area would be recruited for sword expertise (if we had one). In *Proceedings of the 27th Annual Cognitive Science Conference, La Stresa, Italy*. Mahwah, NJ: Lawrence Erlbaum.

Tong, M. H., Joyce, C. A., & Cottrell, G.W. (2008). Why is the fusiform face area recruited for novel categories of expertise? A neurocomputational investigation. *Brain Research*.

Townsend, J. T. (1990). The truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin, 108,* 551–567.

Treat, T. A., McFall, R. M., Viken, R. J., Kruschke, J. K., Nosofsky, R. M., & Wang, S. S. (2007). Clinical-cognitive science: Applying quantitative models of cognitive processing to examination of cognitive aspects of psychopathology. In R. W. J. Neufeld (Ed.), *Advances in clinical-cognitive science: Formal modeling and assessment of processes and symptoms* (pp. 179–205). Washington, DC: APA Books.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience, 3,* 71–86.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11,* 58–64.

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

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A,* 161–204.

Wong, C.-N., & Gauthier, I. (2006). An analysis of letter expertise in a levels-of-categorization framework. *Visual Cognition 15*(7), 854–879.

Xu, Y. (2005). Revisiting the role of the fusiform and occipital face areas in visual expertise. *Cerebral Cortex, 15,* 1234–1242.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747–759.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63*, 271–313.

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

Zhang, L., & Cottrell, C. (2006). Look Ma! No network: PCA of Gabor filters models the development of face discrimination. In *Proceedings of the 28th Annual Cognitive Science Conference, Vancouver, BC, Canada*. Mahwah, NJ: Lawrence Erlbaum.

QUERIES TO BE ANSWERED BY AUTHOR (SEE MANUAL MARKS)

IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. Do NOT mark your corrections on this query sheet.

Chapter 7

Q. No.	Pg No.	Query
AQ1	218	The Erickson & Kruschke (1992) work that is cited in the Figure 7.6 caption doesn't appear in the References (Please add).
AQ2	238	Please provide publisher for reference 'Kohonen et al., 1977'.