Unraveling the time-course of perceptual categorization: Does fastest mean first?

Michael L. Mack (michael.mack@vanderbilt.edu)
Vanderbilt University, Psychology Department
301 Wilson Hall, Nashville, TN 37203 USA

Alan C.-N. Wong (alan.wong@vanderbilt.edu)
Vanderbilt University, Psychology Department
301 Wilson Hall, Nashville, TN 37203 USA

Isabel Gauthier (isabel.gauthier@vanderbilt.edu)
Vanderbilt University, Psychology Department
301 Wilson Hall, Nashville, TN 37203 USA

James W. Tanaka (jtanaka@uvic.ca)
University of Victoria, Department of Psychology
A189 Cornett Building, Victoria, BC Canada

Thomas J. Palmeri (thomas.j.palmeri@vanderbilt.edu)
Vanderbilt University, Psychology Department
301 Wilson Hall, Nashville, TN 37203 USA

Abstract

Perceptual categorization at the basic level is faster than categorization at more superordinate or subordinate levels (Rosch et al, 1976). For categories of perceptual expertise, this basic-level advantage is attenuated such that subordinate levels are categorized as fast as the basic level (Jolicoeur et al, 1984). But, what does it mean to be fastest? One explanation is that levels of abstraction that are categorized faster are processed first. We tested this "fastest means first" hypothesis by contrasting the time-course of basic- and subordinate-level categorization for novice and expert categories with a signal-to-respond task. Results indicated no qualitative differences in the time courses of perceptual decisions for novice and expert categories, nor was there evidence for a basic-level stage preceding a subordinate-level stage. Simulations with an extant object categorization model investigated how seemingly qualitative differences between novice and expert categorization can be accounted for with quantitative changes to model parameters over learning. Together, the behavioral data and simulation results suggest that fastest does not necessarily mean first in perceptual categorization.

Keywords: categorization; basic level; time-course; cognitive modeling

Introduction

The human visual system allows us to rapidly and accurately recognize objects in the world around us. At a glance, we can detect that an object is there, categorize what kind of object it is, and identify the object with a unique name. An important and long-standing question about object processing concerns when these different levels of abstraction become available to the perceiver. Given an image of a dog, how much time is required before we know that the image contains an object, or that this object is an animal, a dog, or a golden retriever? Some of these perceptual decisions are made more quickly than others (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). But does fastest mean first? Are certain perceptual decisions made prior to others during visual object recognition (Palmeri, Wong, & Gauthier, 2004; see also Grill-Spector & Kanwisher, 2005)?

Rosch et al. (1976) found that participants were faster at verifying objects at a basic level (e.g., *dog*) than more superordinate (e.g., *animal*) or subordinate (e.g., *Poodle*) levels of abstraction. While this *basic-level advantage* is robustly found across many object categories, situations exist where this effect is significantly attenuated. For example, with objects of perceptual expertise (Tanaka & Taylor, 1991), subordinate-level categorization occurs as quickly and as accurately as basic-level categorization. This finding has sometimes been characterized as an *entry-level shift* (Jolicoeur, Gluck, & Kosslyn, 1984); for novice categories, the basic level is the entry level, but for expert categories, more subordinate levels become the entry level.

But what does it mean for a particular level of abstraction to be the “entry” level? One simple and straightforward possibility is illustrated in Figure 1. In this box-and-arrow model, objects from novice categories, after low-level visual processing, are categorized first at the basic level (the entry level) before being categorized at more subordinate levels. Basic-level categorization is faster than subordinate-level categorization because basic-level categorization occurs *before* subordinate-level categorization – fastest means first. But for objects from expert categories, there is an entry level shift, whereby objects are categorized at subordinate levels of abstraction without first being categorized at the basic level.

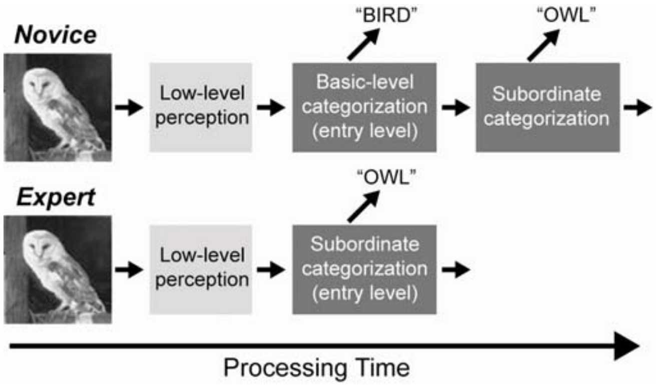

Figure 1: Descriptive model of basic-level advantage (top) and entry-level shift (bottom).

The box-and-arrow model is intuitively plausible and it accounts for the basic-level advantage as well as the entry-level shift with expertise. However, most extant computational models of object recognition and perceptual categorization propose no such preliminary basic-level stage of processing. Instead, decisions about the basic-level category or subordinate-level identity of an object are made after perceptual processing and access to knowledge representations. For example, the EBRW model of perceptual categorization (Nosofsky & Palmeri, 1997; Palmeri, 1997) assumes that the perceptual representation of an object is used to probe memories for past experiences with that object. These probes generate evidence for a particular categorization of an object, whether at a basic level or subordinate level, and evidence drives a random walk decision process. The time to make a decision is determined by how long it takes this stochastic accumulation of evidence to reach a decision threshold, which can be influenced by factors such as between- and within-category similarity. Similarly, a model of object recognition (Joyce & Cottrell, 2004) assumes an object goes through stages of Gabor filtering, principal component analysis, and a neural network mapping perceptual representations onto category labels. Decisions at different levels of abstraction are all driven by trained weights leading to output units at the same output layer of the neural network. Another neural network model of object recognition (Reisenhuber & Poggio, 2000) assumes a hierarchy of information processing that begins with low-level features, moves to view-based representations, object representations, and ultimately category labels. Like the other two models, perceptual decisions at different levels of abstractions are all instantiated at the same output layer of the network. None of these models have an explicit basic-level stage of processing. If there is truly a basic-level stage of processing – as suggested by some interpretations of entry level phenomena – then this would challenge many current computational models of perceptual categorization and object recognition.

The current work attempts to unravel the time-course of basic-level categorization and subordinate-level identification. Does fastest mean first? Is basic-level categorization performed before subordinate-level identification, and does that change with perceptual expertise? We compared categorization at basic and subordinate levels for novice categories (birds and dogs) and an expert category (faces). Normally-functioning adults can be considered face experts (Carey, 1992; Carey & Diamond, 1994; Gauthier & Tarr, 1997). Whether face expertise is qualitatively different from other forms of expertise is hotly debated (Kanwisher, 2000; Tarr & Gauthier, 2000). For the purposes of the present work, we begin with the finding that faces show the same qualitative entry-level shift as other categories of expertise (Tanaka, 2001; Tanaka & Taylor, 1991); specifically, pictures of faces are identified uniquely as quickly as they are categorized as people, much in the same way that for bird experts pictures of birds are identified as quickly as they are categorized as birds. We acknowledge we are confounding expertise with category in this discussion. But as a first pass, using faces for an expert category is far more efficient than recruiting bird or car experts or training people to become experts.

Our first experiment attempts to replicate the entry-level shift with faces reported by Tanaka (2001). The second experiment uses a signal-to-respond (STR) technique to examine the detailed time-course of basic-level categorization and subordinate-level identification. STR probes decisions at various time-points after stimulus onset. Of particular interest is whether we can observe a delay in the initial buildup of subordinate-level decisions relative to basic-level decisions, a potential temporal marker for a delayed stage of subordinate-level categorization. Finally, we consider the results from these experiments from the perspective of the EBRW model.

Experiment 1

The first experiment attempts to replicate the entry-level shift for faces reported by Tanaka (2001).

Methods

Participants Fifteen Vanderbilt University undergraduates participated in two sessions for course credit or monetary compensation.

Stimuli Images of objects from three categories (faces, dogs, birds) were used. Each category consisted of about 320 images from eight different subordinate-level categories: faces - Arnold Schwarzenegger, Jennifer Aniston, Britney Spears, Nicole Kidman, George W. Bush, Mel Gibson, Hillary Clinton, Bill Clinton; dogs - sharpei, beagle, chihuahua, chow chow, golden retriever, german shepherd, weimaraner, poodle; birds - robin, dove, crow, hawk, duck, penguin, ostrich, owl. Images were presented in grayscale and subtended approximately 5.2° x 5.2° of visual angle.

Procedure Participants performed in a speeded category verification task. Participants were seated approximately 60 centimeters from the computer display. Each trial began with a basic- or subordinate-level category label displayed for 1000 ms, followed immediately by the test image. The image remained on the screen until the participant responded. Participants responded by hitting a "yes" key if the label matched the object shown in the test image, and a "no" key if it did not. Half of the category verifications were made at the basic level (face, dog, or bird), and half were made at the subordinate level (Jennifer Aniston, sharpei, robin, etc.). On true trials, the category label and the object in the test image matched. On false trials of basic-level trials, another basic-level category was shown (e.g., a label BIRD for the image of a german shepard). On false trials of subordinate-level trials, and another category label from the same basic-level category was displayed (e.g., a label BEAGLE for the image of a german shepherd); for faces, the label on false trials was a person of the same gender as the one depicted in the image. Participants were instructed to respond as quickly and accurately as possible. Participants completed a short practice session before beginning the experimental trials; the practice stimuli was drawn from other basic-level categories. Each session consisted of 960 trials and lasted approximately one hour.

Results

Verification response times and accuracy for true trials from each of the object categories are shown in Figure 2. A basic-level advantage was found for birds and dogs but not faces. Planned comparisons were conducted on the difference between the basic- and subordinate-level verifications. For both birds and dogs, responses were faster and more accurate for basic than subordinate verifications [birds – RT t(14) = 2.91, accuracy t(14) = 3.93; dogs –RT t(14) = 2.29, accuracy t(14) = 5.04]. For faces, no significant difference was found for either response time [t(14) = 1.81, p = 0.10] or accuracy [t(14) < 1.0].

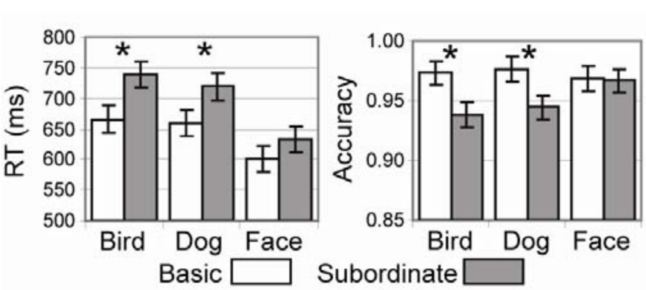

Figure 2: RT and accuracy for speeded verification. White and gray bars represent basic- and subordinate-level performance, respectively. Asterisks (*) represent significant differences ($p < 0.05$) between basic- and subordinate-level performance and error bars represent 95% confidence intervals.

Experiment 2

Replicating Tanaka (2001), in Experiment 1, we observed a basic-level advantage for novice categories (birds and dogs) but not for an expert category (faces). Are these results a consequence of information critical for verifying a basic-level category being available *before* information critical for verifying a subordinate-level category, at least for novice categories? To answer this question, we contrasted the time-course of categorizing expert and novice objects at basic and subordinate levels using a signal-to-respond (STR) task (Corbet & Wickelgren, 1978; Dosher, 1981; Hintzman, Caulton, & Curran, 1994). This task uncovers the relationship between performance and processing time by varying the amount of time the participant has to process a test item in order to make a decision. We used the same category verification tasks used in Experiment 1, but after the image was presented to be verified (at basic or subordinate levels), a tone signaled when the participant was required to make a response. Tones – signals to respond – were presented at varying times after stimulus onset. Varying the lag from image onset to signal allows us to examine how verification accuracy changes over time. To quantitatively compare the temporal dynamics of the functions associated with basic- and subordinate-level categorization, d' values were fitted with the following exponential function (Wickelgren & Corbett, 1977), a function used widely to analyze STR data,

$$d' = \lambda(1 - e^{-\beta(t-\delta)}),$$

where t is the lag until the response signal plus the response time after the signal, λ is the asymptote, β is the growth rate, and δ is the intercept. The last three parameters can be mapped onto particular elements of information processing. The asymptote represents an expected maximum accuracy for the task given unlimited time, the growth rate represents the rate at which relevant information is extracted, and the intercept represents when during the time course of processing begins to grow above chance.

In addition to simply fitting the exponential functions to data, we also tested hypotheses by fitting special cases of the function. For example, we could test whether the intercept for basic and subordinate decisions is the same by constraining the asymptote to be identical for basic and subordinate decisions but allowing the growth rate and asymptote to vary. Indeed, this is a key prediction we will test. We then contrast the fit of the "full model", one with three parameters for basic and three parameters for subordinate, with a "restricted model", such as one with one common intercept parameter for both basic and subordinate but separate growth and asymptote parameters for basic and subordinate. If the restricted model fits significantly worse than the full model, we reject the hypothesis, such as the hypothesis that the intercept is the same for basic and subordinate decisions. The specific approach we used to do this statistical model testing is described, for example, in Dosher (1981).

Methods

Participants Five of the participants that participated in the first experiment completed sixteen sessions of this experiment and were paid $12 per session.

Stimuli The same stimuli were used in experiments 1 and 2.

Procedure Participants completed a category verification task like Experiment 1, but with the inclusion of a signal-to-respond manipulation. On each trial, a category label was displayed for 1000ms, then a premask was displayed for a variable duration, followed by the presentation of the stimulus image for 200ms, followed by a postmask. An auditory signal to respond was presented to the participants after a variable duration (12, 24, 35, 47, 94, 188, 376, 753, or 1506 ms) from image onset. Masking was used to limit the amount of perceptual processing in order to make the task more difficult than unmasked viewing; the same limits from masking were imposed at all signal-to-respond levels. As in Experiment 1, participants verified the match or mismatch between the category label (basic or subordinate) and object in the stimulus image. But they could only respond after hearing the auditory signal. A warning message was presented if the participants responded before the signal or if the response time after the signal was smaller than 180ms or greater than 350ms. Participants responded by pressing keys marked as "yes" and "no" on a keyboard. Each session consisted of 864 trials and lasted approximately 1 hour.

Results

The behavioral time courses along with the exponential curve fits from each object category are shown in Figure 3. The average curve fit parameters from each of the object categories are also shown in Figure 3. Planned comparisons testing for difference in asymptote, growth rate, and intercept parameters between basic- and subordinate-level decisions were conducted. For objects from novice categories (birds and dogs), no significant difference was observed for the intercepts [$t(4) < 1$]. For birds, the asymptote [$t(4) = 2.87$] and the growth rate [$t(4) = 6.95$] was significantly higher in the basic-level condition. For dogs, a marginally significant difference was observed in growth rate only [$t(4) = 2.26$, $p = 0.08$]. For objects from the expert category (faces), planned comparisons revealed a marginally significant difference in the growth rate [$t(4) = 2.56$, $p = 0.06$]. Interestingly, a small but significant difference in intercept [$t(4) = 2.88$, $p < 0.05$] was observed, with the basic-level condition having the smaller intercept. Note that this is entirely opposite to what a "fastest means first" hypothesis would predict: It was the expert category that demonstrated a significant difference in the intercept for basic vs. subordinate decisions, not the novice category.

These conclusions were also confirmed by comparing the fits of the full exponential model (independent curves for basic and subordinate decisions) with those of various restricted models. For objects from novice categories (birds and dogs), the restricted models with equal intercepts and asymptotes fit as well as the full six parameter model. Only the restricted model with equal growth rates fit worse than the full model. For objects from the expert category (faces), the restricted model with equal asymptotes fit as well as the full six parameter model. The restricted models with equal growth rates or intercepts fit worse than the full model.

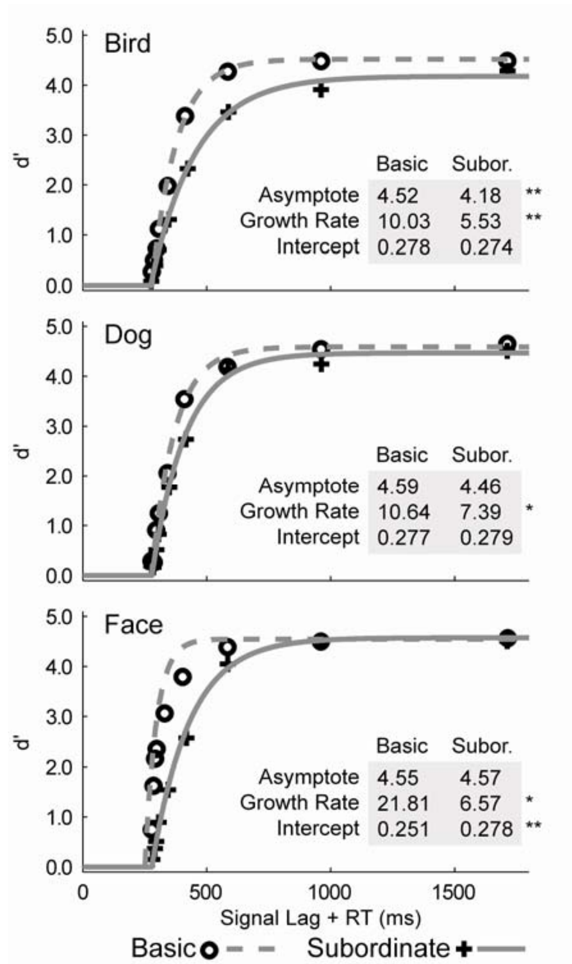

Figure 3: Behavioral time course data (circles and crosses), exponential curve fits (gray lines), and parameters values from exponential fits for STR task (text box). Performance (d') is plotted along the y-axis and reaction time plus lag before the signal is on the x-axis. Average parameter values are shown in the tables with significant differences labeled at $p < 0.05$ (**) and $p < 0.10$ (*).

Discussion

Examination of the results in Figure 3 offers some interesting observations. Across novice and expert categories, the time-courses of basic- and subordinate-level categorization are qualitatively quite similar. This suggests that differences between novice and expert object processing may not be due to qualitatively different processing mechanisms, but rather quantitative differences in processing efficiency (Palmeri et al., 2004). Specifically, there is no delay in the growth of subordinate-level decisions compared to basic-level decisions for novice categories. Basic-level decisions may be made faster than subordinate-level decisions, but faster does not mean first.

One puzzling result was that there was a delay observed for faces, our expert category: basic-level decisions had an earlier intercept than subordinate-level decisions. In this case, there is no difference in overall response time in the speeded task (i.e., there is an entry-level shift), but there is a small but significant difference in the onset of basic-level decisions compared to subordinate-level decisions from chance. It may be that basic-level decisions – is there a "face" in the picture – could be driven by some kind of low-level image properties available very early in visual processing, but that is entirely speculation. Whatever the cause, this interesting effect deserves further attention in future studies. But more importantly, this finding, however puzzling, demonstrates that our task has the statistical power to detect a small but significant intercept difference; it just detected that difference in an unexpected condition.

Simulations

The lack of an intercept difference between basic- and subordinate-level categorization, at least for novice categories, is consistent with most categorization and object recognition models in that these models do not propose a basic-level stage of processing. But why is there a significant response time difference between basic- and subordinate-level categorizations for novice categories that seems to disappear with perceptual expertise? In other words, why is there an entry-level shift? Models like EBRW do not propose any qualitative reconfiguration with learning. Instead they assume gradual quantitative changes. Can these quantitative changes give rise to qualitatively different data patterns across novice and expert categorization?

To begin to answer this question, we simulated the EBRW model using a highly simplified instantiation of the basic and subordinate categories used in our experiments. According to EBRW, objects are represented perceptually as points in a multidimensional psychological space. Like many other categorization models, EBRW does not specify the details of perceptual processing, but multidimensional representations of this sort have been used in extant object recognition models (e.g., Edelman, 1999). To keep things as simple as possible, we assumed a two-dimensional space. As shown in Figure 4*a*, one basic-level category (black points) was represented as a cluster of points in space, with parameter *sw* specifying the within-category similarity. Again, keeping these simulations as simple as possible, we only assumed one other basic-level category (gray points), with the between-category similarity specified by *sb* (Nosofsky, 1988). We assumed that all of these points had a preexisting memory representation, with an associated basic- and subordinate-level label. In these simulations, we did not model any individual exemplar similarity within a single subordinate-level category (either different instance of robins or different images of the same person). Our point was not to model all of the nuances of the experimental situation, or to quantitatively fit data, but simply to provide a theoretical illustration.

According to EBRW, a presented object activates stored memory representations according to the similarity of its perceptual representation to representations stored in memory. Specifically, the similarity between object *i* and stored exemplar *j* is given by $s_{ij} = exp(-c \cdot d_{ij})$, where d_{ij} is the psychological distance between *i* and *j* and *c* is a memory sensitivity parameter. When *c* is large, representations are highly distinct. Evidence for a basic-level category is based on how similar the object is to the target category versus the other category. Evidence for a subordinate-level category is based on how similar the object is to a particular unique object compared to other objects. These evidences drive a stochastic random walk process where an accumulator wanders between a positive ("yes" decision) and a negative ("no" decision) boundary. A response is given when a response boundary is reached. Response time is given by this time plus a constant residual time for perceptual processing and motor responses (*TR*).

Parameters of the simulations include within- and between-category similarity (*sw* and *sb*, respectively), the response boundaries (K_{yes} and K_{no}), the residual time (*TR*), and the sensitivity parameter (*c*). As a simple first step, we assumed that novices would have a smaller value of *c* (lower sensitivity) than experts (see Palmeri et al., 2004). It may be interesting to note that when modeling amnesia, Nosofsky and Zaki (1998) assumed a smaller value of *c* for amnesic individuals compared to nonimpaired controls. So, in this context, where brain damage can cause impaired memory sensitivity, expertise can cause improved memory sensitivity. The qualitative predictions of EBRW remain the same across a wide range of parameter values.

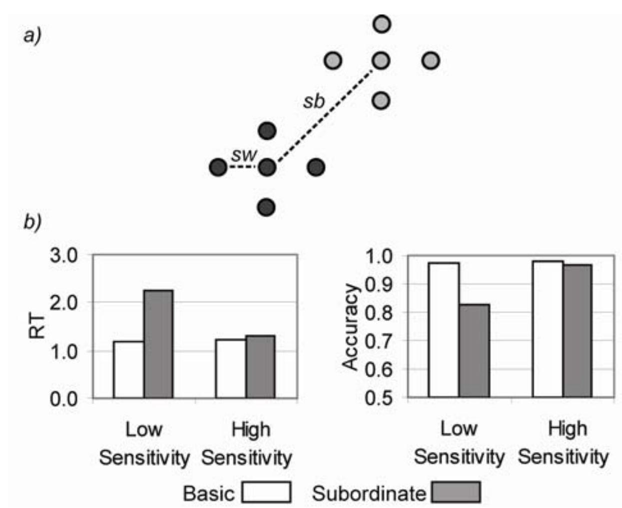

Figure 4: (*a*) Psychological space for EBRW simulations. Parameters *sb* and sw specify similarity between basic-level categories and subordinate-level categories respectively. (*b*) RT and accuracy predictions from simulations. A basic-level advantage and entry-level shift is accounted for with only a quantitative change in the sensitivity parameter.

Results of the simulations are shown in Figure 4*b*. With low sensitivity (novice categorization), a basic-level advantage is predicted for both RT and accuracy. But with high sensitivity (expert categorization), the basic-level advantage is eliminated (note that we made no attempt to scale the RT predictions to the observed range). So, by only adjusting the sensitivity parameter of the model, a quantitative change over learning, the standard basic-level advantage is seen for novice categories and the entry-level shift is seen for expert categories. This brief glimpse at predictions of EBRW demonstrate how extant categorization models can account for qualitative differences in categorization performances at different levels of abstraction across expertise without the need to propose separate stages for basic- and subordinate-level categorization (see also Joyce & Cottrell, 2004).

Conclusions

The results of this study suggest that fastest does not necessarily mean first when it comes to basic- and subordinate-level categorization. While categorization of novice categories is faster at the basic than subordinate level in a speeded category verification task, no qualitative difference was seen in the time-course of decisions in a signal-to-respond paradigm. Our simulation modeling provided a hint at how current categorization models can account for the basic-level advantage and the entry-level shift through quantitative changes of parameters without qualitative changes in processing.

Acknowledgments

This work was supported by NSF SBE-0542013, NSF HSD-DHBS05, and grant from the James S. McDonnell Foundation. Thanks to Nancy Carlisle and Jennifer Richler for fruitful discussion and critical feedback.

References

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions Royal Society London, B, 335*, 395-403.

Carey, S. & Diamond, R. (1994). Are faces perceived as configurations more by adults than children? *Visual Cognition, 1*, 253-274.

Corbett, A. T., & Wickelgren, W. A. (1978). Semantic memory retrieval: Analysis by speed accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology, 30*(1), 1-15.

Dosher, B. A. (1981). The effects of delay and interference: A speed-accuracy study. *Cognitive Psychology, 13*(4), 551-582.

Gauthier, I. & Tarr, M. J., (1997). Becoming a ‘Greeble’ expert: exploring mechanisms for face recognition. *Vision Research, 37*, 1673–1681.

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.

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

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 275-289.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*(2), 243-275.

Joyce, C. & Cottrell G. W. (2004). Solving the visual expertise mystery. In *Connectionist Models of Cognition and Perception II: Proceedings of the Eighth Neural Computation and Psychology Workshop*, Howard Bowman and Christophe Labiouse (Eds.), World Scientific.

Kanwisher. N. (2000). Domain specifcity in face perception. *Nature Neuroscience*, *3(8)*, 759-763.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302-4311.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 700-708.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*(2), 266-300.

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

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382-439.

Tanaka, J. W. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General, 130*(3), 534-543.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*(3), 457-482.

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

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). *Proceedings of the 27th Annual Cognitive Science Conference*. Mahwah, NJ: Lawrence Erlbaum.

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