# Research Article

## A NEURAL BASIS FOR EXPERT OBJECT RECOGNITION

James W. Tanaka[1] and Tim Curran[2]

*[1]Oberlin College and [2]Case Western Reserve University*

**Abstract—***Although most adults are considered to be experts in the identification of faces, fewer people specialize in the recognition of other objects, such as birds and dogs. In this research, the neurophysiological processes associated with expert bird and dog recognition were investigated using event-related potentials. An enhanced early negative component (N170, 164 ms) was found when bird and dog experts categorized objects in their domain of expertise relative to when they categorized objects outside their domain of expertise. This finding indicates that objects from well-learned categories are neurologically differentiated from objects from lesser-known categories at a relatively early stage of visual processing.*

The term *jizz* is used by veteran birdwatchers to describe their flash of instant recognition of a bird based on its color, shape, and movement. Similarly, dog-show judges and breeders can discern in a single glance the specific breed and attributes of a canine from its facial structure, gait, and posture. Although the subtle perceptual cues that differentiate species of birds and breeds of dogs frequently go unnoticed by the novice, detection of these cues seems obvious and automatic to the expert.

What is the neural basis of this perceptual expertise? Although relatively few people specialize in the recognition of particular objects (e.g., birds, cars, dogs), it has been suggested that virtually all people are experts in the recognition of faces (Carey, 1992; Tanaka & Gauthier, 1997). Electrophysiological studies employing event-related potentials (ERPs) have been informative for understanding the temporal aspects of face processing. Results from these experiments indicate that the magnitude of an early ERP component, referred to as the N170, is significantly larger when participants view face stimuli than when they view other natural and human-made objects (Bentin & Deouell, 2000; Eimer, 2000). Moreover, patients with prosopagnosia, the inability to recognize faces, either fail to demonstrate an enhanced N170 to faces (Eimer & McCarthy, 1999) or demonstrate a nonselective enhanced N170 to both face and nonface stimuli (Bentin, Deouell, & Soroker, 1999). Thus, the presence of the N170 component during viewing of faces in normal participants and its absence or the presence of a nonselective N170 during viewing of faces in prosopagnosic patients indicate that the N170 is a good neurophysiological index of face perception processes. More generally, this evidence suggests that the visual system can differentially respond to specific and important kinds of visual information at a relatively early stage of processing (Bentin & Deouell, 2000). However, whether the enhanced N170 component is exclusive to faces or whether it can be extended to other important objects in the environment (i.e., objects of expertise) is an open question.

In the current study, we investigated the neural basis of object expertise by monitoring brain wave activity of bird and dog experts while they categorized pictures of common birds and dogs. The experiment was designed so that participants served as their own experimental controls in that they were expected to perform as experts when categorizing objects in their domain of expertise (e.g., bird experts categorizing birds) and novices when categorizing objects outside their domain of expertise (e.g., bird experts categorizing dogs). We expected that if the increased N170 reflects a general form of expert processing that is not unique to faces, experts would exhibit an enhanced N170 when categorizing objects in their domain of expertise relative to when they categorized objects outside their domain of expertise.

Address correspondence to James W. Tanaka, Department of Psychology, Severance Lab, Oberlin College, Oberlin, OH 44074, e-mail: tanaka@cs.oberlin.edu, or to Tim Curran, Department of Psychology, University of Colorado, Campus Box 345, Boulder, CO 80309-0345, e-mail: curran@psych.colorado.edu.

## METHOD

### Participants

Fifteen bird and 15 dog experts participated in the experiment. Participants were selected on the basis of their active membership in local bird and dog organizations and on the basis of personal recommendations from other organization members. All participants had a minimum of 10 years of experience in their area of expertise, with the majority of experts having more than 20 years of experience. The group of bird experts was composed of 9 males and 6 females, ranging in age from 32 years to 57 years old, with a mean age of 44.0 years. The group of dog experts was composed of 9 females and 6 males, ranging from 37 years to 54 years of age, with a mean age of 49.7 years. All participants had normal or corrected-to-normal vision.

### Materials

Stimuli for the experiment consisted of pictures of common birds and dogs. The birds selected for the experiment were the robin, sparrow, cardinal, oriole, pigeon, blue jay, hawk, and crow. These birds are among the 15 most frequently mentioned birds according to Battig and Montague's (1969) category norm study. The dogs selected for the study were the German shepherd, Doberman pinscher, beagle, chow chow, schnauzer, golden retriever, collie, and dachshund. These dogs are among the 10 most popular dogs as determined by the American Kennel Club's list of registered dogs. For each bird and dog, five exemplar pictures (e.g., five different robin pictures) were selected from field guides, handbooks, and manuals. The exemplar pictures were digitized with a MicroTek Z Scanner; half of the exemplars were scanned in a left-facing orientation, and the other half in a right-facing orientation. Each image was scaled to fit within a 100- × 100-pixel array. Additional foil pictures of plants were selected from magazine and book sources.

### Procedure

After reading a list of the eight birds and eight dogs included in the study, participants were seated in front of a computer monitor at a

  

viewing distance of approximately 60 cm. At this distance, picture stimuli subtended a visual angle of approximately 1.6° in the horizontal and vertical dimensions. At the beginning of each trial, a fixation point (a plus sign) appeared on the computer monitor for a random interval that varied between 1,000 and 1,500 ms. The fixation point was replaced by a category name (superordinate, basic, or subordinate) for 255 ms and was then replaced by the fixation point for 570 ms. The picture stimulus was then presented for 255 ms and was replaced by the fixation point for 735 ms. At the end of the trial, the true/false screen provided the prompt for the participant's response. Subjects were instructed to press the "true" key if the picture matched the category word; otherwise, they were to press the "false" key. For example, in the *true* condition, the category labels "animal," "bird," and "robin" preceded the picture of a robin in the superordinate, basic, and subordinate trials, respectively. In the *false* condition, the category labels "plant," "dog," and "sparrow" preceded the robin picture at the superordinate, basic, and subordinate levels, respectively. The foil category labels at the three levels of abstraction were drawn from the contrast category that was at the same level of abstraction as the target. For subordinate-level foils, a different false subordinate label preceded each picture presentation.

An electroencephalogram (EEG) was recorded for 2 s during each trial, beginning 195 ms prior to the onset of the category word, and was terminated prior to the subject's response at the onset of the true/false screen. Subjects were instructed to withhold their response until the true/false screen appeared. Although the delayed-response paradigm rendered reaction time data uninformative, it was used in order to minimize EEG contamination from overt motor movements.

The intertrial interval was randomly varied from 2 to 2.5 s. Each exemplar (40 birds, 40 dogs) was tested across the three category conditions (superordinate, basic, subordinate) and two response conditions (true, false), yielding a total of 480 critical trials. Additionally, there were 80 catch trials (40 true trials with plants and 40 false trials with animals). The 560 trials were presented randomly, with rest breaks provided every 40 trials.

### EEG Recording and ERP Averaging

Scalp EEG was collected with a 128-channel Geodesic Sensor Net™ connected to an AC-coupled, 128-channel, high-input impedance amplifier (200 MΩ, Net Amps™, Electrical Geodesics, Inc., Eugene, Ore.). Individual sensors were adjusted until impedances were less than 50 kΩ. Amplified analog voltages (0.1- to 100-Hz bandpass) were digitized at 250 Hz. Recorded voltages were initially referenced to a vertex channel. The EEGs were averaged into ERPs, separately for each condition, after incorrect trials were removed. Trials were also removed from ERP averaging if they contained eye movements (vertical electro-oculogram channel differences greater than 70 μV) or more than five bad channels (changing more than 100 μV between samples, or reaching amplitudes over 200 μV). Data from individual channels that were consistently bad for a given subject were replaced using a spherical interpolation algorithm. After incorrect trials and trials containing movement artifacts were eliminated, the mean number of acceptable trials retained for ERP averaging per condition per subject was 34 (range: 31–36). Voltages were rereferenced off-line into an average-reference representation to minimize the effects of reference-site activity and accurately estimate the scalp topography of the measured electrical fields (Dien, 1998). ERPs were baseline-corrected for the 100-ms interval prior to the presentation of the picture stimulus and digitally low-pass filtered at 40 Hz. A final grand average was obtained by averaging across the subject averages for each experimental condition.

## RESULTS

Analysis for the present study focused on the N170 component in response to the picture stimulus. The channels selected for the N170 analysis were determined by identifying the electrode sites in the left and right hemispheres where the amplitude of the grand-averaged N170 was maximal across all conditions. As shown in Figure 1, the N170 was identified to be maximal at channels 59 and 92. Analyses were conducted on ERPs averaged across these maximal channels and the six immediately adjacent channels within each hemisphere.

The peak latency of the N170 occurred 164 ms after onset of the picture stimulus (see Fig. 2). We performed an analysis of variance (ANOVA) of the mean amplitudes within the time window ±24 ms (±2 *SD*) around the peak latency (i.e., 140 to 188 ms). The ANOVA included the following within-subjects factors: 2 object domains (expert, novice) × 3 category levels (superordinate, basic, subordinate) × 2 responses (true, false) × 2 hemispheres (left, right). Overall, the amplitude of the N170 was more negative in the right hemisphere than the left hemisphere, $F(1, 29) = 5.466$, $MSE = 56.124$, $p < .05$. This laterality effect is apparent in Figure 3, which shows the topographic distribution of the N170 effect for the expert and novice conditions. More important, the critical main effect of object domain (expert vs. novice) was significant, $F(1, 29) = 22.921$, $MSE = 43.287$, $p < .001$, demonstrating that the magnitude of the N170 was greater when participants categorized objects in the expert domain than when they categorized objects in the novice domain (see Fig. 3). The main effects of category level and response were not significant, $p > .05$.[1]

To more closely examine the expertise effect, we performed a second ANOVA using expertise (bird experts, dog experts) as a between-groups factor and stimuli (birds, dogs) as a within-groups factor. As reflected in the wave plots shown in Figure 2, there was a significant interaction between expertise and stimuli, $F(1, 29) = 24.413$, $MSE = 6.915$, $p < .001$. Bird experts exhibited a more negative N170 in response to bird stimuli than dog stimuli, whereas dog experts showed the reverse pattern of effects. Separate analyses of bird and dog experts revealed that the stimulus differences were significant for both bird experts, $F(1, 14) = 8.122$, $MSE = 15.035$, $p < .05$, and dog experts, $F(1, 14) = 14.912$, $MSE = 29.451$, $p < .01$.

## DISCUSSION

The foregoing analysis revealed that the N170 component was larger when experts categorized objects in their domain of expertise relative to when they categorized objects outside their domain of expertise. Because the expertise effect was doubly dissociated (i.e., bird experts demonstrated the N170 effect for bird stimuli, whereas dog experts demonstrated the N170 effect for dog stimuli), the effect

1. It is possible that the effect of category level in the current experiment was eliminated by testing only two kinds of basic-level categories (birds and dogs). In contrast, a recent study by one of us (Tanaka, Luu, Weisbrod, & Kiefer, 1999) found a reliable categorization effect when participants classified objects across a range of 12 artifactual and 8 natural basic-level categories.

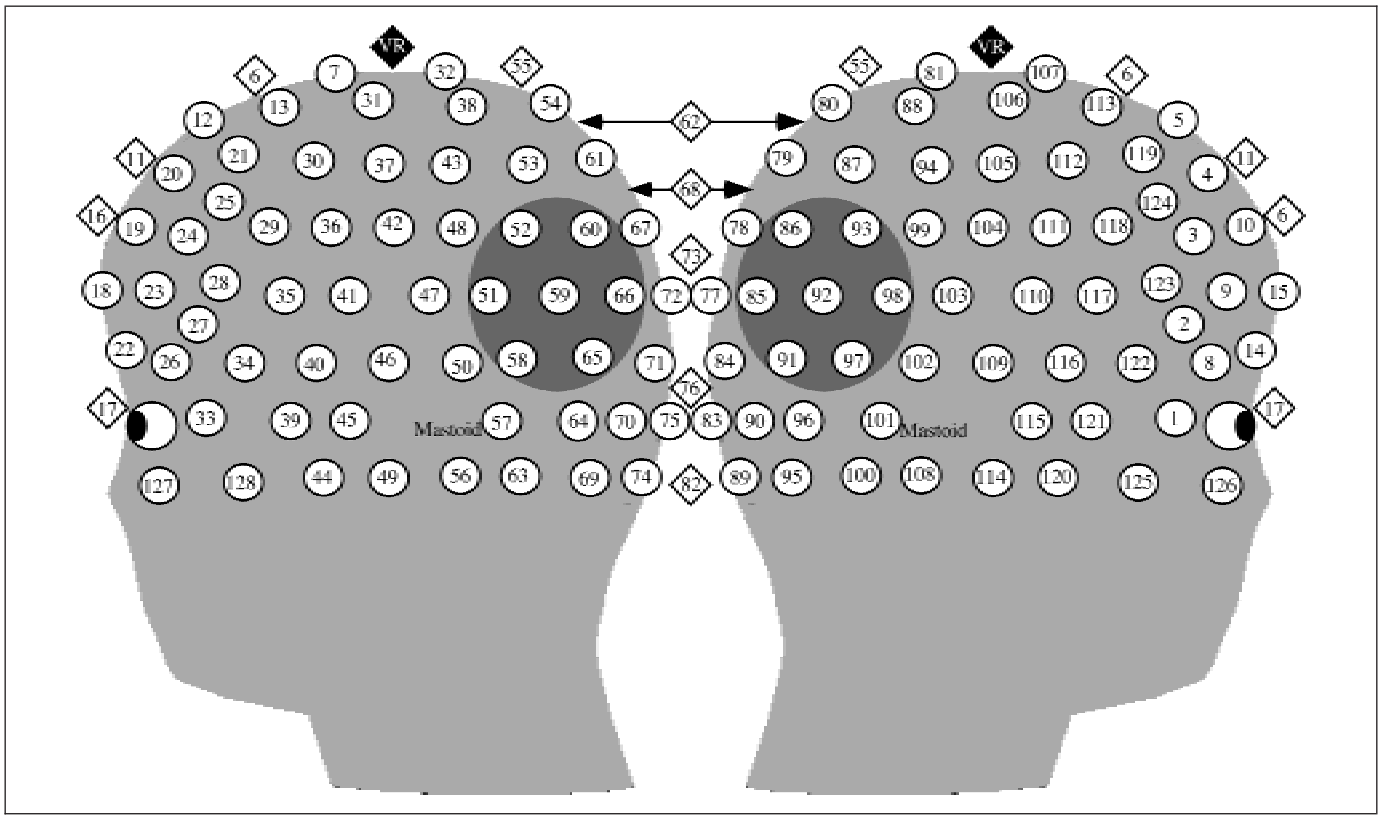


**Fig. 1.** Approximate scalp locations on the 128-channel Geodesic Sensor Net™. The N170 was identified to be maximal at channels 59 (located between T5 and P3 of the International 10-20 System; Jasper, 1958) and 92 (between T6 and P4). Analyses were conducted on event-related potentials averaged across the maximal channel and the 6 immediately adjacent channels within each hemisphere (denoted by dark circular regions). VR = vertex reference.

cannot be attributed to low-level, image properties (e.g., spatial frequencies, color) associated with a particular object category. Nor can the obtained differences be attributed to group differences given that the same participants were tested in both the expert and the novice conditions. Furthermore, it was revealing that the magnitude of the expertise effect was not affected by the expectation primed by the category (superordinate, basic, subordinate) or response (true, false) condition. For example, bird experts exhibited an N170 of equivalent magnitude in response to a picture of a robin regardless of whether the picture was preceded by the correct category labels, "animal," "bird," or "robin," or by the misleading category labels, "plant," "dog," or "sparrow." The degree to which the N170 was not affected by the participants' conscious expectations suggests that the neurological response to an expert object is obligatory and automatic.

The enhanced N170 in response to the objects of expertise is similar in timing and scalp distribution to the enhanced N170 reported elsewhere for faces. The peak latency of 164 ms for objects of expertise in our study is similar to the reported peak latencies for faces, which range from 156 to 189 ms (156 ms in Rossion et al., 1999; 162 ms in Taylor, McCarthy, Saliba, & Degiovanni, 1999; 172 ms in Bentin, Allison, Puce, Perez, & McCarthy, 1996; 189 ms in George, Evans, Fiori, Davidoff, & Renault, 1996). In terms of scalp distribution, most studies have recorded the N170 for faces at channels T5 and T6 (Botzel, Schulze, & Todieck, 1995; Eimer, 2000; George et al., 1996; Rossion et al., 1999; Taylor et al., 1999). Locations T5 and T6 (International 10-20 System; Jasper, 1958) correspond to the channels 58 and 97 analyzed in the present study, but these channels are slightly anterior and inferior to the peak channels, 59 and 92 (see Fig. 1). One study (Bentin & Deouell, 2000) sampling more inferior locations than previous studies, but within the present recording array, has reported the maximal N170 in response to faces at mastoid sites (locations 57 and 101 in Fig. 1). Thus, the N170 for objects of expertise may be distributed slightly more superiorly and posteriorly than the N170 for faces. Establishing a clearer spatiotemporal correspondence between the N170 for faces and other objects of expertise must await further studies recording these potentials under identical conditions. However, the timing and location of the enhanced N170 found for objects of expertise are strikingly similar to the timing and location of the enhanced N170 found for faces.

The current results lend further support to the argument that "different parts of the visual system can learn to tune themselves to respond selectively to specific (probably ecologically important) visual information" (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999, p. 252). In the same manner that it is ecologically important for all people to be experts in face recognition, it is ecologically important for the participants in our study to be experts in the recognition of dogs and birds. According to the expertise account, given the appropriate task demands and learning opportunities, the perceptual system can be modified and tuned to the structural properties of a particular object class. Functionally, it has been suggested that perceptual tuning allows for the efficient classification of objects of expertise, which, in turn, facilitates the subsequent stage of object

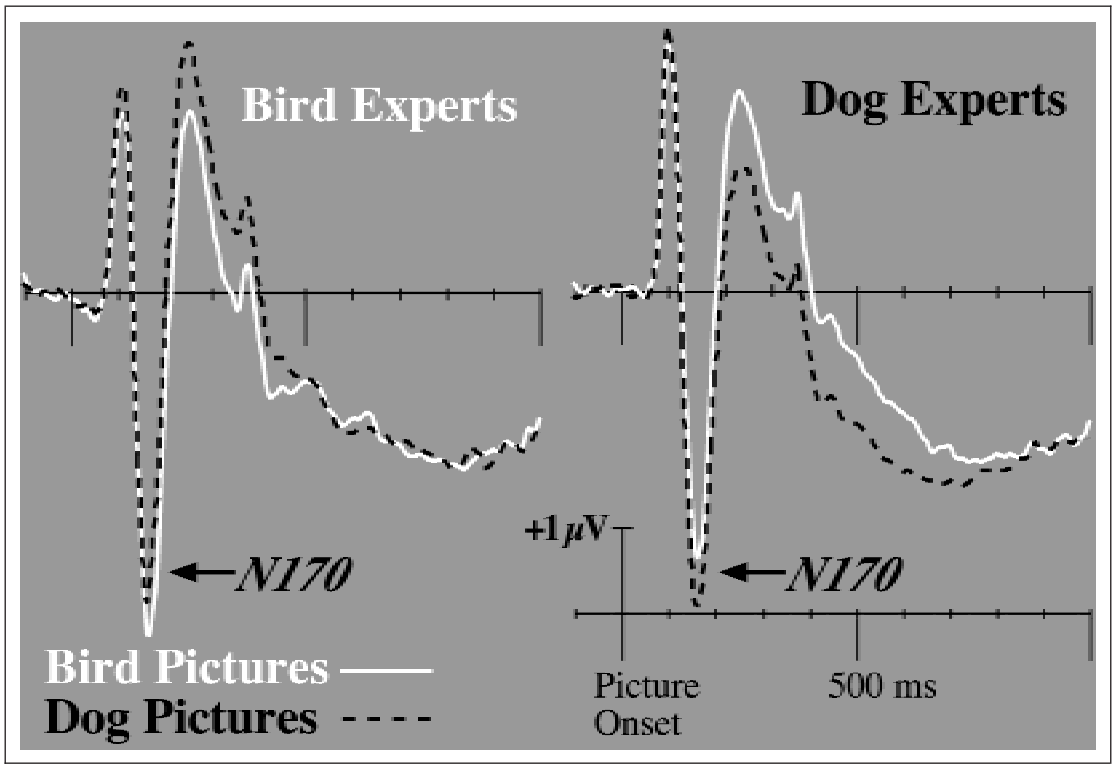


**Fig. 2.** Wave plots of the composite N170 channels for bird experts (left) and dog experts (right). For each group, event-related potentials (ERPs) are plotted separately for bird and dog stimuli. The plotted ERPs were averaged across the channels used in the analyses (51, 52, 58, 59, 60, 65, 66, 85, 86, 91, 92, 93, 97, 98; see Fig. 1).

recognition (Ullman, 1996). This interpretation is consistent with behavioral results demonstrating that experts, relative to novices, show speeded recognition of objects of expertise at subordinate levels of representation (Gauthier & Tarr, 1997; Johnson & Mervis, 1997; Tanaka, in press; Tanaka & Taylor, 1991).

What are the candidate neuroanatomical substrates of object expertise? Studies using functional magnetic resonance imaging methods have shown that extrastriate areas of the visual system, specifically the fusiform gyrus, exhibit more activation to faces than nonface objects (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997). Similarly, intracranial studies recording evoked potentials directly from the cortical surface have found that face stimuli elicit greater negative potentials at the fusiform recording sites than nonface stimuli (Allison, Puce, Spencer, & McCarthy, 1999; Bentin et al., 1996). However, because the intracranial potentials occur at a later latency than the scalp potentials (200 ms vs. 170 ms), have been recorded at multiple extrastriate loci, and can be selective to different types of face information (e.g., isolated eyes, whole face), the precise relation between the N170 scalp potentials and N200 intracranial potentials is not clear. Despite these differences, recent neuroimaging results with object experts provide an interesting parallel between the expert N170 results and the effects of expertise on fusiform activity. In a study in which participants were trained to discriminate perceptually similar, artificial stimuli (i.e., Greebles), the initially unresponsive fusiform area became increasingly activated to Greeble stimuli over the course of training (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). Moreover, Gauthier and colleagues (Gauthier, Skudlarski, Gore, & Anderson, 2000) have shown that the

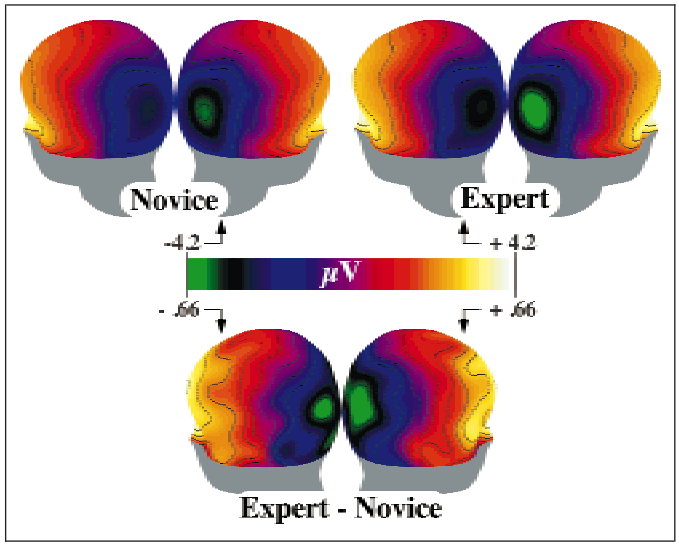


**Fig. 3.** Topographic distribution of the N170 expertise effect. The illustrations at the top show mean voltages from 140 to 188 ms after picture onset, separately for novice and expert domains (0.7 μV between contour lines). The illustration at the bottom shows mean voltage differences between expert and novice domains between 140 and 188 ms after picture onset (0.11 μV between contour lines).

fusiform face area is differentially activated when real-world bird and car experts view stimuli in their domain of expertise relative to when they view stimuli outside this domain. Thus, although a large body of evidence indicates the involvement of the fusiform area in face-processing tasks, it is possible that this area can be recruited to accommodate the processing needs of other kinds of perceptual expertise.

In summary, the present results demonstrate that approximately 164 ms after presentation, objects of expertise are neurologically differentiated from objects from lesser-known categories. These results are similar to the results reported for faces (Bentin et al., 1996). Although previous studies comparing faces with nonface objects are potentially confounded by stimulus differences, the current study avoided possible stimulus artifacts and clearly demonstrates that the enhanced N170 is the direct result of perceptual learning. Collectively, these findings suggest that the pattern of neural activity associated with the early stages of object perception can be modified by real-world experience and learning.

**Acknowledgments—**This research was supported by a Research Status Award from Oberlin College to J.W.T. We would like to thank the dog and bird experts who participated in the study. We want to thank Aaron Daggy, Ruth Spinks, and Jennifer Teichman for their research assistance. We are also grateful to Shlomo Bentin for his helpful comments and advice.

## REFERENCES

Allison, T., Puce, A., Spencer, D., & McCarthy, G. (1999). Electrophysiological studies of human face perception: I. Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex, 9,* 415–430.

Battig, W.F., & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80*(3, Part 2), 1–46.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8,* 551–565.

Bentin, S., & Deouell, L. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology, 17,* 35–54.

Bentin, S., Deouell, L.Y., & Soroker, N. (1999). Selective visual streaming in face recognition: Evidence from developmental prosopagnosia. *NeuroReport, 10,* 823–827.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M.H., Echallier, J.F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience, 11,* 235–260.

Botzel, K., Schulze, S., & Todieck, S. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental Brain Research, 104,* 135–143.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London B, 335,* 95–103.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers, 30,* 34–43.

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology, 111,* 694–705.

Eimer, M., & McCarthy, R.A. (1999). Prosopagnosia and structural encoding of faces: Evidence from event-related potentials. *NeuroReport, 10,* 255–259.

Gauthier, I., Skudlarski, P., Gore, J., & Anderson, A. (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 the face recognition mechanism. *Vision Research, 37,* 1673–1682.

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

George, N., Evans, J., Fiori, N., Davidoff, J., & Renault, B. (1996). Brain events related to normal and moderately scrambled faces. *Cognitive Brain Research, 4,* 65–76.

Jasper, H. (1958). The ten-twenty system of the international federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371–375.

Johnson, K., & Mervis, C. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126,* 248–277.

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.

McCarthy, G., Puce, A., Gore, J., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 16,* 605–610.

Rossion, B., Delvenne, J.F., Dabatisse, D., Goffaux, V., Bruyer, R., Crommelinch, M., & Guerit, J.M. (1999). Spatio-temporal localization of the face inversion effect: An event related potentials study. *Biological Psychology, 50,* 173–189.

Tanaka, J., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *NeuroReport, 10,* 829–835.

Tanaka, J.W. (in press). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General.*

Tanaka, J.W., & Gauthier, I. (1997). Expertise in object and face recognition. In R.L. Goldstone, P.G. Schyns, & D.L. Medin (Eds.), *Psychology of learning and motivation* (Vol. 36, pp. 83–125). San Diego: Academic Press.

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.

Taylor, M.J., McCarthy, G., Saliba, E., & Degiovanni, E. (1999). ERP evidence of developmental changes in processing of faces. *Clinical Neurophysiology, 110,* 910–915.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.

(RECEIVED 11/4/99; REVISION ACCEPTED 5/4/00)