

Volume 19: No. 7, July-August 2005

How Experience Shapes Vision

by Michael J. Tarr

Michael J. Tarr received his PhD from M.I.T. in 1989. He is currently the Fox Professor of Ophthalmology and Visual Sciences and a Professor of Cognitive and Linguistic Sciences at Brown University. He is the recipient of the 1997 APA Distinguished Scientific Award for Early Career Contribution to Psychology in the Area of Cognition/Human Learning and the 2003 Troland Research Award from the National Academy of Sciences. His research focuses on human visual processing and cognition using a wide variety of methodologies. More information about his lab, as well as downloadable stimuli such as the Greebles, is available at <http://www.tarrlab.org>

The study of human visual object recognition has a relatively short and somewhat controversial history. My interest in how experience shapes both object representations and the processes applied to such representations has involved me in two spirited debates concerning: i) how we recognize objects across changes in viewpoint or other sources of variation; and, ii) the functional interpretation of category selectivity in the primate visual cortex. In both cases our work has successfully challenged appealing, widely held theories, prompting reinterpretation of how the human visual system accomplishes the task of recognizing objects.

Invariance

My 20-month old son is fond of pointing to two similar objects and declaring “same thing!” So-called “basic-level” recognition involves categorizing visually-similar, yet distinct objects as members of the same class. Thus, one form of invariance requires our visual systems to perform a many-to-one mapping between individual exemplars and object categories. At the same time, individual exemplars of three-dimensional objects rarely appear the same from one moment to the next. Variation in the two-dimensional images falling on our retinæ arises from almost any change in viewing conditions, including changes in position, changes in object pose, changes in lighting, or changes in object configuration. Thus, a second form of invariance requires our visual systems to perform a many-to-one mapping between individual “views” of objects and their unique identities. Theories of object recognition have often addressed these twin challenges by positing three-dimensional, volumetric representations that are invariant over both class and viewing variation (Marr & Nishihara, 1978; Biederman, 1987).

One of the most salient characteristics of such models is that they are viewpoint invariant. That is, the same three-dimensional representation is derived over a wide range of viewing orientations. The behavioral implication of this is that recognition performance should be independent of the particular viewpoint from which the object is seen. This prediction is also consistent with our intuitions – our recognition of familiar objects from unfamiliar viewpoints feels effortless. However, psychophysical tests of this prediction by me (Tarr & Pinker, 1989; Tarr, 1995) and my colleagues (e.g., Bülthoff & Edelman, 1992) suggest otherwise. Using a wide range of stimuli, we have found that experience with particular views is a critical factor in achieving invariance. Several studies have found that if observers learn to recognize novel objects from specific viewpoints, they are both faster and more accurate at recognizing these same objects from those familiar viewpoints relative to unfamiliar viewpoints (Tarr & Pinker, 1989; Bülthoff & Edelman, 1992; Tarr, 1995). Moreover, recognition performance at unfamiliar viewpoints is systematically related to those views that are familiar, observers taking progressively more time and being progressively less accurate as the distance between the unfamiliar and the familiar increases. These and related results (e.g., Hayward & Tarr, 1997; Lawson & Humphreys, 1996; Tarr et al., 1998b) suggest that human object recognition relies on multiple “views,” where each view encodes the appearance of an object under specific viewing conditions, including viewpoint, pose, configuration, and lighting (Tarr et al., 1998a) and a collection of such views constitutes the mental representation of a given object.

“View-Based” Models

Given that we represent three-dimensional objects as collections of viewpoint-specific representations, how do we manage to attain both class and view invariance? One clue may be found in the systematic pattern of performance seen for the recognition of familiar objects in unfamiliar viewpoints. Steven Pinker and I (1989) believed that this pattern arose due to the use of mental rotation (Shepard & Metzler, 1971) or a continuous alignment process (Ullman, 1989) to transform unfamiliar viewpoints to familiar views encoded in visual memory (with familiar viewpoints being recognized with the need for a transformation). The strongest evidence

favoring this interpretation is the nearly identical linear reaction time pattern across viewpoint obtained for the same objects in naming and left-right handedness discrimination tasks (Tarr & Pinker, 1989). However, in a nice example of how neuroimaging can inform us regarding cognitive processes, Gauthier et al. (2002) found that entirely different brain systems exhibited viewpoint-dependent activity for recognition and mental rotation tasks (i.e., Pinker and I were wrong). Consistent with current thinking on the “division of labor” in the primate visual system (Goodale & Milner, 1992), the recognition of objects in unfamiliar viewpoints preferentially recruited the fusiform region along the ventral pathway, while handedness discriminations recruited the superior parietal lobe along the dorsal pathway (Gauthier et al., 2002). Thus, the computational mechanism underlying viewpoint-dependent recognition behavior is not the continuous transformation process of mental rotation.

Accumulation of Evidence

How then do we explain the fact that it takes more time to recognize objects in unfamiliar views? Perrett, Oram, and Ashbridge (1998) presented a novel solution to this problem based on the well-established finding that individual object-selective neurons tend to preferentially respond to particular object views (Perrett et al., 1985; Logothetis & Pauls, 1995). This sort of “view-tuning” appears puzzling when considered at the single neuron level – if objects are represented by individual neurons tuned to specific views, how can any sort of invariance be achieved? The answer may lie in considering populations of neurons as the actual neural code for objects. In this context, individual neurons may be considered as coding – from a familiar viewpoint – the complex features or parts of which objects are composed. Recognition then takes the form of an “accumulation of evidence” across all neurons selective for some aspect of a given object. During recognition the particular rate of accumulation will depend on the similarity between visible features/parts in the present viewpoint and the view-specific features/parts to which individual neurons are tuned (Perrett et al., 1998). Across a population of object-selective neurons, sufficient neural “evidence” (summed neuron activity) will accumulate more slowly when the current appearance of an object is dissimilar from its learned appearance. In contrast, when an object’s appearance is close to previously-experienced views, evidence across the appropriate neural population will accumulate more rapidly. Thus, systematic behavioral changes in recognition performance with changes in viewpoint may be explained as a consequence of how similarity is computed between new object percepts and their previously-learned neural representations.

Returning to the question raised above, one appealing element of the accumulation of evidence approach is that class invariance may be achieved using the same mechanism as view invariance. In this model, recognition amounts to reaching a threshold of sufficient evidence across a neural population. One consequence of this is that unfamiliar views of objects will require more time to reach threshold, but will be successfully recognized given some similarity between input and known viewpoints. A second consequence is that unfamiliar exemplars within a familiar class will be likewise recognized given some similarity (e.g., similar configurations and viewpoints) with known exemplars from within that class. One behavioral implication is that familiarity with individual objects should facilitate the viewpoint-dependent recognition of other, visually-similar objects; a prediction borne out by several studies (Edelman, 1995; Tarr & Gauthier, 1998). A second implication is that object viewpoints or class exemplars that are significantly different from known views or objects should be instantiated as distinct representations; again, a prediction that seems to be supported (Jolicoeur, Gluck, & Kosslyn, 1984). Whether the same mechanism can account for all forms of object invariance remains unknown, although it seems likely that configuration and lighting variation present unique challenges that may require the inclusion of structural information (e.g., Biederman, 1987; Bienenstock & Geman, 1995).

Category Selectivity in Visual Cortex

While the question of invariance has often dominated thinking on object recognition, recent neuroimaging results have focused more on subordinate, rather than basic, level recognition (e.g., face recognition). In this vein, a third challenge to the human visual system is discriminating between individuals within a homogeneous object class, the most salient example being face recognition. Rather than exploring the computational principles underlying this ability, the specific question addressed within this domain is often whether faces are “special” or not (Farah et al., 1998). Although the form of this debate has varied, neuroimaging studies logically speak to the issue of neural specialization, that is, whether there are distinct regions of the visual system specialized for and exclusive to face recognition. Neuroimaging studies using both PET (Sergent, Ohta, & MacDonald, 1992) and fMRI (Kanwisher, McDermott, & Chun, 1997) reveal a small region in the fusiform gyrus of the ventral-temporal lobe that is more active when we view faces as compared to other objects. One interpretation of this finding is that this brain area, dubbed the “fusiform face area” or FFA (Kanwisher et al., 1997), is a face-specific neural module (Fodor, 1983); that is, its function is to process and/or recognize faces and only faces. An alternative explanation is that this and other forms of putatively face-specific processing (e.g., Farah, 1990; Yin, 1969) are actually by-products of our extensive experience which makes us face experts (Diamond & Carey, 1986). Thus, the recognition of individual faces exhibits qualities that should be true for any domain of visual expertise for a homogeneous object class; faces being

processed this way by default due to their social importance, but not as a result of anything intrinsic to them as visual objects.

Greebles, Cars, and Birds, Oh My!

Isabel Gauthier and I have collaborated with others (Gauthier & Brown, 2004) to explore these competing accounts using several different approaches. In the laboratory we have created experts for novel objects called "Greebles" and observed the behavioral (Gauthier & Tarr, 1997) and neural (Gauthier et al., 1999; Rossion et al., 2000) changes that occur with the onset of expertise. We have also examined the neural bases of expert-level recognition in extant experts (Gauthier et al., 2000; Righi & Tarr, 2004; Tanaka & Curran, 2001).

Several of our findings speak directly to the question "Are faces special?" First, Greeble experts, but not Greeble novices, show behavioral effects – notably configural processing – that are often taken as markers for specialized face processing (Gauthier & Tarr, 1997; Gauthier et al., 1998). Second, Greeble experts, but not Greeble novices, show category-selectivity for Greebles in the right fusiform gyrus (Gauthier et al., 1999). Similarly, bird experts show category-selectivity for birds, but not cars, in the right fusiform, while car experts show category-selectivity for cars, but not birds (Gauthier et al., 2000). Reinforcing the generality of this result, chess experts, but not chess novices, likewise show category-selectivity in right fusiform for valid, but not invalid, chess game boards (Righi & Tarr, 2004). Third, across Greeble expertise training, subjects show a significant positive correlation between a behavioral measure of holistic processing (sensitivity to the presence of the correct parts for that object) and neural activity in the right fusiform (Gauthier & Tarr, 2002). Similarly, bird and car experts show a significant correlation between their relative expertise measured behaviorally (birds minus cars) and neural activity in the right fusiform (Gauthier et al., 2000). Behaviorally measured chess playing ability also shows a significant correlation with right fusiform response (Righi & Tarr, 2004). Fourth, the N170 potential (as measured by event-related potentials) shows face-like modulation in Greeble (Rossion et al., 2000), bird and dog experts (Tanaka & Curran, 2001), but only for a given expert's domain of expertise.

These and other findings (e.g., Gauthier, in press; Tarr & Gauthier, 2000) suggest that putatively face-specific effects may be obtained with non-face objects, but only when subjects are experts for the non-face object domain. Thus, our answer to the question "Are faces special?" is yes and no. There is no doubt that faces are special in terms of their centrality to social interaction, yet our data suggest that this is a characteristic of our environment and not an intrinsic property of our brains. This argument is based on studies using both Greeble and extant experts in domains as diverse as cars, birds, and chess. Across these domains we find a pattern of behavioral and neural effects consistent with those seen for face recognition. In particular, category-selective activation in the fusiform gyrus has, of late, been taken as the hallmark of face specificity. We and others see similar selectivity for many other object domains, in particular when subjects are experts. Of course, this analysis only addresses the question of spatial specialization, that is, "is a particular piece of neural real estate dedicated to face processing?" (unlikely given current data) and begs the more meaningful question "what are the computational principles underlying processing in this brain region?" (we don't know at present).

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

How Experience Shapes Vision

My research on two issues fundamental to human object recognition have led me to the conclusion that experience plays a significant role in shaping our visual behavior. First, I believe that invariance is achieved by deploying our astonishing memory capacities, that is, by encoding a great deal of what we see as it originally appears. Second, I think that there is compelling evidence that our object recognition system is generic, that is, the organization of object classes in ventral-temporal cortex is based on the manner in which we learn to process different categories. At the same time, it should be obvious that evolution has endowed us with the

specific mechanisms that enable such learning through experience. The challenge we face as cognitive scientists is to unravel the synergistic manner in which nature and nurture interact throughout our lifetimes.

References

- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
- Bienenstock, E., & Geman, S. (1995). Compositionality in neural systems. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 223-226). Cambridge, MA: MIT Press.
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl. Acad. Sci. USA*, 89, 60-64.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, 115(2), 107-117.
- Edelman, S. (1995). Class similarity and viewpoint invariance in the recognition of 3d objects. *Biological Cybernetics*, 72, 207-220.
- Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: The MIT Press.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, 105(3), 482-498.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring the face recognition mechanism. *Vision Research*, 37(12), 1673-1682.
- Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research*, 38(15/16), 2401-2428.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2(6), 568-573.
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191-197.
- Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A., Skudlarski, P., & Gore, J. C. (2002). Bold activity during mental rotation and viewpoint-dependent object recognition. *Neuron*, 34(1), 161-171.
- Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 431-446.
- Gauthier, I., & Brown, D. D. (2004). The perceptual expertise network: Innovation on collaboration, *Science Briefs: APA Online*, March.
- Gauthier, I. (in press). Constraints on the acquisition of specialization for face processing. In Y. Munakata & M. Johnson (Eds.), *Attention & performance* (Vol. XXI).
- Goodale, M. A., & Milner, D. A. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15(1), 20-25.
- Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5), 555-562.
- Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23(5), 1511-1521.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425-2430.
- Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, 16, 243-275.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, 17, 4302-4311.
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8, 679-685.
- Lawson, R., & Humphreys, G. W. (1996). View specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 395-416.
- Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representation in the primate. *Cerebral Cortex*, 3, 270-288.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. of Lond. B*, 200, 269-294.
- Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society B*, 223, 293-317.

- Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition*, 67(1,2), 111-145.
- Rhodes, G., Byatt, G., Michie, P. T., & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation or expert individuation? *Journal of Cognitive Neuroscience*, 16, 1-15.
- Righi, G., & Tarr, M. J. (2004). Are chess experts any different from face, bird, or Greeble experts? *Journal of Vision*, 4(8), 504a. [Presentation at the 4th Annual Meeting of the Vision Sciences Society]
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuroreport*, 11(1), 69-74.
- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain*, 115, 15-36.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, 12(1), 43-47.
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2), 233-282.
- Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, 2(1), 55-82.
- Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition*, 67(1-2), 71-108.
- Tarr, M. J., Kersten, D., & Bühlhoff, H. H. (1998a). Why the visual system might encode the effects of illumination. *Vision Research*, 38(15/16), 2259-2275.
- Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998b). Three-dimensional object recognition is viewpoint-dependent. *Nature Neuroscience*, 1(4), 275-277.
- Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8), 764-769.
- Tarr, M. J. (2003). Visual object recognition: Can a single mechanism suffice? In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 177-211). Oxford, UK: Oxford University Press.
- Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32, 193-254.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141-145.

[Return to Psychological Science Agenda Homepage](#)

© 2006 American Psychological Association

Science Directorate

750 First Street, NE • Washington, DC • 20002-4242

Phone: 202-336-6000 • TDD/TTY: 202-336-6123

Fax: 202-336-5953 • [E-mail](#)

[PsychNET®](#) | [Terms of Use](#) | [Privacy Policy](#) | [Security](#)