

High-Level Vision

Aspects of vision that reflect influences from memory, context, or intention are considered "high-level vision," a term originating in a hierarchical approach to vision. In currently popular interactive hierarchical models, however, it is almost impossible to distinguish where one level of processing ends and another begins. This is because partial outputs from lower-level processes initiate higher-level processes, and the outputs of higher-level processes feed back to influence processing at the lower levels (McClelland and Rumelhart 1986). Thus, the distinctions between processes residing at high, intermediate, and low levels are difficult to draw. Indeed, substantial empirical evidence indicates that some high-level processes influence behaviors that are traditionally considered low-level or MID-LEVEL VISION. With this caveat in mind, the following topics will be considered under the heading "high-level vision": object and face recognition, scene perception and context effects, effects of intention and object knowledge on perception, and the mental structures used to integrate across successive glances at an object or a scene.

One major focus of theory and research in high-level vision is an attempt to understand how humans manage to recognize and categorize familiar objects quickly and reliably. An adequate theory of OBJECT RECOGNITION must account for (1) the accuracy of object recognition over changes in object size, location, and orientation (preferably, this account would not posit a different memory record for each view of every object ever seen); (2) the means by which the spatial relationships between the parts or features of an object are represented (given that objects and spaces seem to be coded in different VISUAL PROCESSING STREAMS, with object processing occurring in ventral pathways and space processing occurring in dorsal pathways); and (3) the attributes of both basic-level and subordinate-level recognition (e.g., recognition of a finch as both a bird and as a specific kind of bird). Current competing object recognition theories differ in their approach to each of these factors (see Biederman 1987; Tarr 1995). According to Biederman (1987), objects are parsed into parts at concave portions of their bounding contours, and the parts are represented in memory by a set of abstract components (generalized cylinders); the claim is that these components can be extracted from an image independent of changes in orientation (up to an accidental view rendering certain component features invisible). On Biederman's view, (1) object recognition should be robust to orientation changes as long as the same components can be extracted from the image; and (2) very few views of each object need be represented in memory. Tarr (1995) adopts a different theoretical approach, proposing that specific views of objects are represented by salient features, and that object recognition is orientation-dependent. On Tarr's approach, multiple views of each object are stored in memory, and objects seen in new views must undergo some time-consuming process before they are recognized. The empirical evidence suggesting that object recognition is orientation-dependent is accumulating, favoring the multiple-views approach. However, evidence indicates

that the concave portions of bounding contours are more important for recognition than other contour segments, supporting the idea that part structure is critically important for object recognition, consistent with an approach like Biederman's.

A related, but independent, research focus is FACE RECOGNITION. Behavioral evidence obtained from both normal and brain damaged populations suggests that different mechanisms are used to represent faces and objects, and in particular, that holistic, configural processing seems to be more critical for face than for object recognition (e.g., Farah, Tanaka, and Drain 1995; Moscovitch, Winocur, and Behrmann 1997).

A second major problem in high-level vision is the question of how scenes are perceived and, in particular, how the semantic and spatial context provided by a scene influences the identification of the individual objects within the scene. Any effects of scene context require the interaction of spatially local and spatially global processing mechanisms; the means by which this is accomplished have yet to be identified. Research indicates that scene-consistent objects are identified faster and more accurately when placed in a contextually appropriate spatial location rather than one that is contextually inappropriate (Biederman, Mezzanotte, and Rabinowitz 1982). In addition, recent evidence (Diwadkar and McNamara 1997) suggests that scene memory is viewpoint dependent, just as object memory is orientation-dependent. Such dependencies and similarities in the processing of scenes and objects raise questions about the extent to which the mechanisms for processing scenes and objects overlap, despite the apparent specialization of the two different visual processing streams. Nevertheless, much research continues to argue for fundamental differences in the representation of spaces and objects. An example is evidence that when no semantic context is present, memory for spatial configuration is excellent under conditions in which memory for object identity is impaired (Simons 1996). It is worth pointing out that whereas context effects are prevalent in visual perception, their influence may not extend to motor responses generated on the basis of visual input (Milner and Goodale 1995). Experiments measuring motor responses raise the possibility that the different visual processing streams associated with ventral and dorsal anatomical pathways are specialized for vision and action, respectively, rather than for the visual perception of objects and spaces, as originally hypothesized.

A third question central to investigations of high-level vision concerns the mechanisms by which successive glances at an object or a scene are integrated. Phenomenologically, perception of objects and scenes seems to be holistic and fully elaborated rather than piecemeal, abstract, and schematic. Contrary to the phenomenological impressions, evidence indicates that perception is not "everywhere dense" (Hochberg 1968); instead, visual percepts are largely determined by the stimulation obtained at the locus of fixation or attention, even when inconsistent information lies nearby (Hochberg and Peterson 1987; Peterson and Gibson 1991; Rensink O'Regan, and Clark 1997). It has been shown that the structures used to integrate the information obtained in successive glances are abstract and schematic in

nature (Irwin 1996); hence, they can tolerate the integration of inconsistent information. Similarly, visual memories, assessed via mental IMAGERY research, are known to be schematic compared to visual percepts (Kosslyn 1990; Peterson 1993). One of the abiding questions in high-level vision is, given such circumstances, how can one account for the phenomenological impressions that percepts are detailed and fully elaborated? A recent appealing proposal is that the apparently richness of visual percepts is an illusion, made possible because eye movements (see EYE MOVEMENTS AND VISUAL ATTENTION) can be made rapidly to real world locations containing the perceptual details required to answer perceptual inquiries (O'Regan 1992). On this view, the world serves as an external memory, filling in and supplementing abstract percepts on demand.

Other research in high-level vision investigates various forms of TOP-DOWN PROCESSING IN VISION. Included in this domain are experiments concerning the effects of observers' intentions on perception (where intentions are manipulated via instructions; Hochberg and Peterson 1987) and investigations of how object knowledge affects the perception of moving or stationary displays. For example, detection thresholds are lower for known objects than for their scrambled counterparts (Purcell and Stewart 1991). In addition, object recognition cues contribute to DEPTH PERCEPTION, along with the classic depth cues and the configural cues of GESTALT PERCEPTION (Peterson 1994). For moving displays, influences from object memories affect the direction in which ambiguous displays appear to move (McBeath, Morikowa, and Kaiser 1992). Moreover, although apparent motion typically seems to take the shortest path between two locations, Shiffrar and Freyd (1993) found that, under certain timing conditions, object-appropriate pathways are preferred over the shortest pathways. Much early research investigating the contributions to perception from knowledge, motivation, and intention was discredited by later research showing that the original results were due to response bias (Pastore 1949). Hence, it is important to ascertain whether effects of knowledge and intentions lie in perception *per se* rather than in memory or response bias. One way to do this is to measure perceptual processes on-line; another way is to measure perception indirectly by asking observers to report about variables that are perceptually coupled to the variable to which intention or knowledge refers (Hochberg and Peterson 1987). Many of these recent experiments have succeeded in localizing the effects of intention and knowledge in perception *per se* by using one or more of these methods; hence, representing an advance over previous attempts to study top-down effects on perception.

It is important to point out that not all forms of knowledge or memory can influence perception and not all aspects of perception can be influenced by knowledge and memory. Consider the moon illusion, for example. When the moon is viewed near the horizon, it appears much larger than it does when it is viewed in the zenith; yet the moon itself does not change size, nor does it cover areas of different size on the viewer's retina in the two viewing conditions. The difference in apparent size is an illusion, most likely caused by the presence of many depth cues in the horizon

condition and by the absence of depth cues in the zenith condition. However, knowledge that the apparent size difference is an illusion does not eliminate or even reduce the illusion; the same is true for many illusions. The boundaries of the effects of knowledge and intentions on perception have yet to be firmly established. One possibility is that perception can be altered only by knowledge residing in the structures normally accessed in the course of perceptual organization (Peterson et al. 1996).

In summary, research in high-level vision focuses on questions regarding how context, memory, knowledge, and intention can influence visual perception. In the course of investigations into the interaction between perception and these higher-order processes, we will undoubtedly learn more about both. The result will be a deeper understanding of high-level vision and its component processes.

See also PICTORIAL ART AND VISION; SHAPE PERCEPTION; SPATIAL PERCEPTION; STRUCTURE FROM VISUAL INFORMATION SOURCES; VISUAL OBJECT RECOGNITION, AI

—Mary A. Peterson

References

- Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review* 94: 115–147.
- Biederman, I., R. J. Mezzanotte, and J. C. Rabinowitz. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology* 14: 143–177.
- Diwadkar, V. A., and T. P. McNamara. (1997). Viewpoint dependence in scene recognition. *Psychological Science* 8: 302–307.
- Farah, M. J., J. W. Tanaka, and H. M. Drain. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance* 21: 628–634.
- Hochberg, J. (1968). In the mind's eye. In R. N. Haber, Ed., *Contemporary Theory and Research in Visual Perception*. New York: Holt, Rinehart, and Winston, pp. 309–331.
- Hochberg, J., and M. A. Peterson. (1987). Piecemeal organization and cognitive components in object perception: perceptually coupled responses to moving objects. *Journal of Experimental Psychology: General* 116: 370–380.
- Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science* 5: 94–100.
- Kosslyn, S. M. (1990). Mental imagery. In D. N. Osherson, S. M. Kosslyn, and J. M. Hollerbach, Eds., *Visual Cognition and Action: An Invitation to Cognitive Science*, vol. 2. Cambridge, MA: MIT Press.
- McBeath, M. C., K. Morikowa, and M. Kaiser. (1992). Perceptual bias for forward-facing motion. *Psychological Science* 3: 362–367.
- McClelland, J. L., and D. E. Rumelhart. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2. Cambridge, MA: MIT Press.
- Milner, A. D., and M. Goodale. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.
- Moscovitch, M., G. Winocur, and M. Behrmann. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience* 9: 555–604.
- O'Regan, D. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology* 46: 461–488.

- Pastore, N. (1949). Need as a determinant of perception. *The Journal of Psychology* 28: 457-475.
- Peterson, M. A. (1993). The ambiguity of mental images: Insights regarding the structure of shape memory and its function in creativity. In B. Roskos-Ewoldsen, M. J. Intons-Peterson, and R. Anderson, Eds., *Imagery, Creativity, and Discovery: A Cognitive Perspective*. Amsterdam: North Holland, pp. 151-185.
- Peterson, M. A. (1994). Shape recognition can and does occur before figure-ground organization. *Current Directions in Psychological Science* 3: 105-111.
- Peterson, M. A., and B. S. Gibson. (1991). Directing spatial attention within an object: Altering the functional equivalence of shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance* 17: 170-182.
- Peterson, M. A., L. Nadel, P. Bloom, and M. F. Garrett. (1996). Space and Language. In P. Bloom, M. A. Peterson, L. Nadel, and M. F. Garrett, Eds., *Language and Space*. Cambridge, MA: MIT Press, pp. 553-577.
- Purcell, D. G., and A. L. Stewart. (1991). The object-detection effect: configuration enhances perception. *Perception and Psychophysics* 50: 215-224.
- Rensink, R. A., J. K. O'Regan, and J. J. Clark. (1997). To see or not to see: the need for attention to perceive changes. *Psychological Science* 8: 368-373.
- Shiffrar, M., and J. J. Freyd. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science* 4: 379-384.
- Simons, D. (1996). In sight, out of mind: when object representations fail. *Psychological Science* 5: 301-305.
- Tarr, M. J. (1995). Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review* 2: 55-82.

Hippocampus

The hippocampus is a brain structure located deep within the temporal lobe, surrounded by the lateral ventricle, and connected to subcortical nuclei via the fornix and to the neocortex via the parahippocampal region. Considerations of the information-processing functions of the hippocampus highlight its position as the final convergence site for outputs from many areas of the CEREBRAL CORTEX, and its divergent outputs that return to influence or organize cortical memory representations (figure 1).

The neocortex provides information to the hippocampus only from the highest sensory processing areas, plus multimodal and LIMBIC SYSTEM cortical areas and the olfactory cortex. These inputs follow a coarse rostral-to-caudal topography arriving in the parahippocampal region, composed of the perirhinal, parahippocampal, and entorhinal cortices (Burwell, Witter, and Amaral 1995). The latter areas project onto the hippocampus itself at each of its main subdivisions, the dentate gyrus, the CA3 and CA1 components of Ammon's horn, and the subiculum (figure 1). The main flow of information through the hippocampus involves serial connections from the dentate gyrus to CA3, CA3 to CA1, and then CA1 to the subiculum (Amaral and Witter 1989). The intrinsic hippocampal pathway partially preserves the topographical gradients of neocortical input, but there is also considerable divergence and associational connections particularly at the CA3 stage. Outputs of the subic-

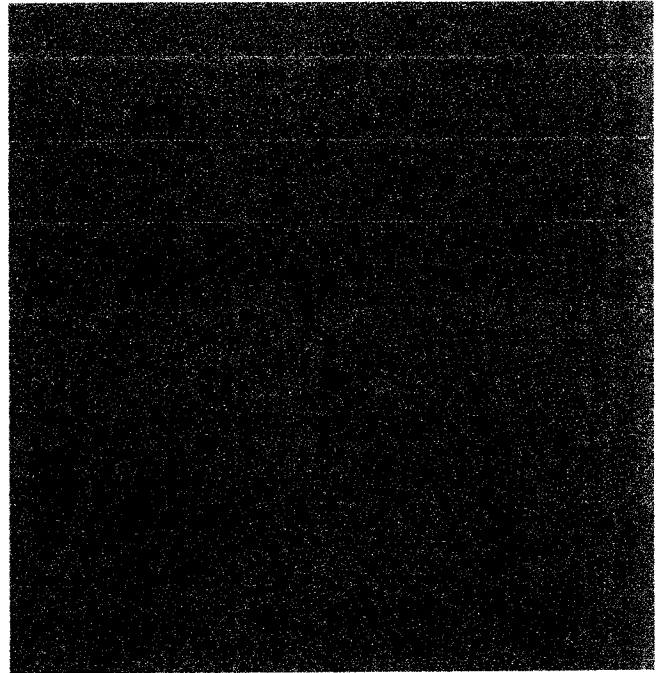


Figure 1.

ulum, and to a lesser extent CA1, are directed back to the parahippocampal region, which in turn projects back onto the neocortical and olfactory areas that were the source of cortical inputs. These aspects of hippocampal organization maximize the potential for association of information from many cortical streams, and the potential for such associations to influence cortical processing broadly. Furthermore, the capacity for associative plasticity in the form of LONG-TERM POTENTIATION at dentate and CA1 synapses is well established, and has been related to normal rhythmic (theta) bursting activity in the hippocampus and to hippocampal memory function.

In 1957 Scoville and Milner described a patient known as H. M. who suffered profound amnesia following bilateral removal of substantial portions of both the hippocampus and parahippocampal region. H. M. demonstrated an almost complete failure to learn new material, whereas his remote autobiographical memories and short term memory were completely intact, leading to the view that the hippocampal region plays a specific role in the consolidation of short term memories into a permanent store. In addition, the amnesic impairment is also selective to declarative or explicit memory (cf. IMPLICIT VS EXPLICIT MEMORY), the capacity for conscious and direct expression of both episodic and semantic memory (Corkin 1984; Squire et al. 1993; see also EPI-SODIC VS SEMANTIC MEMORY). Conversely, amnesiacs demonstrate normal MOTOR LEARNING and CONDITIONING, and normal sensory adaptations and "priming" of perceptual stimuli; such forms of implicit memory occur despite their inability to recall or recognize the learning materials or the events surrounding the learning experience (see MEMORY, HUMAN NEUROPSYCHOLOGY). The development of a nonhuman primate model has demonstrated a parallel dissociation