

Viewpoint Dependency in Visual Object Recognition Does Not Necessarily Imply Viewer-Centered Representation

Moshe Bar

Abstract

■ The nature of visual object representation in the brain is the subject of a prolonged debate. One set of theories asserts that objects are represented by their structural description and the representation is “object-centered.” Theories from the other side of the debate suggest that humans store multiple “snapshots” for each object, depicting it as seen under various conditions, and the representation is therefore “viewer-centered.” The principal tool that has been used to support and criticize each of these hypotheses is subjects’ performance in recognizing objects under novel viewing conditions. For example, if subjects take more time in recognizing an object from an unfamiliar viewpoint, it is common to claim that the representation of that object is

viewpoint-dependent and therefore viewer-centered. It is suggested here, however, that performance cost in recognition of objects under novel conditions may be misleading when studying the nature of object representation. Specifically, it is argued that viewpoint-dependent performance is not necessarily an indication of viewer-centered representation. An account for the neural basis of perceptual priming is first provided. In light of this account, it is conceivable that viewpoint dependency reflects the utilization of neural paths with different levels of sensitivity *en route* to the same representation, rather than the existence of viewpoint-specific representations. New experimental paradigms are required to study the validity of the viewer-centered approach. ■

INTRODUCTION

The appearance of an object is affected by factors such as viewing distance, viewing angle, illumination, and shading. In fact, the difference in appearance is so pronounced that the same object will rarely project the same image in two different encounters. Nevertheless, humans demonstrate an impressive ability to overcome those changes. Studying the mechanisms by which we cope with such transformations can reveal properties of the representation that subserve object recognition.

There are two main proposals for shape-based visual object representation.¹ The most elaborated and extensively tested theory of the object-centered “camp,” recognition-by-components (RBC; Biederman, 1987), asserts that objects are represented by their structural description, consisting of the object parts (termed “geons”) and the spatial relations between them. This description is constructed from viewpoint invariant properties such as whether a contour is curved or straight, whether two contours are parallel or not, and what type of a vertex is created from the cotermination of two contours (Lowe, 1984). Hummel and Biederman (1992) proposed that these invariant properties are extracted from edges in the input image, and that their reconstruction is used as indexing for activating a geon

structural description of an object in memory, leading to recognition. As a result of building structural descriptions from invariant components and unique relationships, the representation is not only characteristic of the object but it also remains unchanged under many transformations.

Viewer-centered theories, on the other hand, suggest that humans store multiple instances of each object (Edelman & Bülthoff, 1992; Poggio & Edelman, 1990; Ullman, 1989), representing different possible appearances of the same object. For example, the representation of an elephant would include “snapshots” of this elephant viewed from different angles and distances. According to these theories, object recognition is accomplished by performing template matching between the input image and the “snapshots” stored in memory. Consequently, we are able to recognize an object because its novel image could be matched to existing templates (Ullman, 1989; Poggio & Edelman, 1990) or matched to a collection of pooled features that comprises a template (Riesenhuber & Poggio, 1999).

The central experimental paradigm that has been used to try to distinguish between these two theories of object representation is to study recognition performance under novel viewing conditions. The prediction of object-centered theories is that if the characteristic invariant properties are visible and as easily extractable in the novel image of a familiar object, its recognition will

Harvard University, Massachusetts General Hospital

be as efficient as the recognition of the original, untransformed image (Biederman & Gerhardstein, 1993). On the other hand, viewer-centered theories predict that due to the computational effort involved in template matching (e.g., alignment, interpolation and extrapolation, pooling between multiple inputs), humans will be slower and less accurate when recognizing a familiar object under novel conditions. Consequently, performance advantage in recognizing objects from familiar viewpoints compared with novel viewpoints has been taken to support viewer-centered representation.

As will be argued here, however, an alternative interpretation of viewpoint-dependency is that it reflects the use of neural paths with different sensitivity *en route* to the activation of the same representation. The sensitivity of these paths is determined by experience, resulting in different activation and propagation characteristics (i.e., speed and reliability).

In the present context, a “neural path” includes the neurons that are directly involved in the perception and recognition of the specific image. In other words, this path may start in the primary visual cortex, V1, and include all the activated neurons that lead from V1 to the activation of the object’s identity. For example, a path that may be used for activating the representation of a *box* would include neurons that represent straight lines in orthogonal orientations, cells that represent vertices, cells that represent squares and cells that represent the binding of the squares. These paths are likely to be confined to neurons in the ventral visual stream (V1, V2, V4, IT), which is known to process object shape for the purpose of recognition.

Consider the case of rotation in depth, which is the most common transformation studied in this debate. Rotation in depth modifies the two-dimensional image of the object. Therefore, a different set of visual features is available in different views of the same object. Because the activation of a specific path is determined by the features in the input image, the activated path is, at least partially, different when the object is seen from different viewpoints.

Next, consider priming. It has been shown repeatedly that a single presentation of a stimulus typically improves the recognition of that stimulus in a subsequent encounter. It will be assumed here that the activation of a certain neural path potentiates this path such that it is more sensitive, resulting in priming that facilitates future recognition attempts. Therefore, encountering an object from a viewpoint that has been experienced previously takes advantage of a sensitized path, whereas recognition from a new viewpoint will utilize a less sensitized path, possibly to activate the same representation. Consequently, an object seen from a familiar viewpoint is expected to be recognized faster and more accurately than when the same object is seen under novel viewing conditions.

In the following sections I will: (1) describe briefly experiments that studied visual recognition of depth-rotated objects, (2) discuss perceptual priming, (3) outline a semi-hypothetical neural mechanism for priming, and (4) apply this account of priming to argue that viewpoint-dependency is not necessarily indicative of the nature of visual object representation.

STUDYING VIEWPOINT-INVARIANCE

Initially, the debate over the nature of object representation was concentrated around the question of whether the representation is object-centered or viewer-centered (Ullman, 1989; Marr, 1982). This terminology captures the difference in the nature of object representation and, indirectly, the computations required for recognition. Over the years, the debate has shifted to concentrate on whether object recognition is viewpoint-dependent or viewpoint-invariant, assuming that this is an analogous question to the previous one, and that studying viewpoint dependency will also provide a resolution to the original debate regarding the representation. The main reason for this shift, it seems, was an apparent pragmatism: while contrasting object-centered with viewer-centered representations directly seemed experimentally difficult, studying viewpoint-invariance is a well-defined question that can be, and has been, addressed behaviorally. As will be suggested next, however, reports of viewpoint-dependency cannot support viewer-centered representations unequivocally.

RBC is a representative theory that supports the object-centered perspective. One of its predictions is that recognizing an object from a novel viewpoint is expected to be as quick and as accurate as recognizing that object from a familiar viewpoint, as long as the same characteristic components are visible in the two views and the relations between them remain identical. Indeed, when these conditions were met, object recognition was found to be generally viewpoint-invariant both in behavioral experiments (Biederman & Bar, 1999; Biederman & Gerhardstein, 1993) and in physiological recordings (Booth & Rolls, 1998).

At the same time, several studies of object recognition have reported viewpoint-dependency (e.g., Tarr, Williams, Hayward, & Gauthier, 1998; Tarr & Bülhoff, 1995; Bülhoff & Edelman, 1992; Poggio & Edelman, 1990; Rock & DiVita, 1987). In those experiments, subjects were often required to judge whether a pair of sequentially presented images was of the same or different objects, where the images could differ in viewing position. The stimuli used were either familiar, nameable objects or novel, nameless objects. The general finding of these reports was that object recognition from unfamiliar viewpoints requires more time and results in more mistakes than recognition from familiar viewpoints. The typical conclusion was that this cost is a manifestation of template matching processes.

It is proposed here that because of the effect of experience on the neural paths that lead to object representations, viewpoint-dependent object recognition is a necessary but insufficient support for viewer-centered representation. To make the case for this argument, a neural mechanism for priming will be outlined first.

A NEURAL MECHANISM FOR PRIMING

Priming may be defined as facilitation in recognition of a perceptual stimulus that stems from previous experience. In a typical recognition priming experiment, for example, subjects are exposed to a set of briefly presented stimuli (the primes) and their performance in object naming is recorded (usually reaction times and correct responses). In a second block, subjects are presented with either the same stimuli or stimuli that have some defined relationship to the primes. Any performance difference that is specific to the stimulus is taken as a measure of priming.

Priming is considered to be an implicit type of memory because it does not require explicit recollection of previous experience. It is believed to exist independently of other types of memory, while closely interacting with them (Tulving & Schacter, 1990). Priming can be long lasting. For example, priming of object naming can last 48 weeks after a single exposure to a picture (Cave, 1997) and priming of word completions can be evidenced even after 16 months (Sloman, Hayman, Ohta, Law, & Tulving, 1998). Priming has been shown to be a useful tool for studying the representation of visual objects. For example, repeating a picture of an object at different sizes typically does not affect the magnitude of the priming, suggesting that shape representation is size invariant (Biederman & Cooper, 1992).

Little is known about the neural mechanisms that subserve priming. It is reasonable to assume that different neural substrates mediate different types of priming (e.g., perceptual, conceptual) and that visual recognition priming, specifically, takes place in the ventral pathway of the visual cortex. Indeed, results from various imaging studies suggest that visual (or perceptual) priming is associated with changes in the visual cortex, whereas semantic (or conceptual) priming is associated with changes in the prefrontal cortex and to some extent also the medial temporal lobe (see Schacter & Buckner, 1998; Wiggs & Martin, 1998, for reviews).

One phenomenon, termed "repetition suppression," may be a central component of the neural mechanisms underlying priming. Converging evidence suggests that repeated exposure to similar stimuli results in reduced cortical activity.² Single-unit recordings in the inferior temporal cortex (IT) (e.g., Ringo, 1996; Li, Miller, & Desimone, 1993; Brown, Wilson, & Riches, 1987) and in the prefrontal cortex (e.g., Rainer & Miller, 2000) of monkeys demonstrated that cells reduce their activity as a response to a familiar stimulus compared with novel

stimuli. Similar effects are reported in studies of humans. For example, functional magnetic resonance imaging signals are often reduced when familiar images are repeated on different occasions, compared with the activity elicited by novel images (Buckner et al., 1998). Results from positron emission tomography studies (Blaxton et al., 1996; Squire et al., 1992) and from event-related potential imaging of verbal and visual tasks (Rugg, Soardi, & Doyle, 1995) also suggest a correlation between priming and reduced neural activity. While the connection between the physiological repetition suppression and the behavioral priming is somewhat speculative, the similarities between them are compelling. They are both stimulus-specific, long-lasting, robust in spite of intervening images, modulated by number of repetitions, and invariant to transformations such as size and position (see also Wiggs & Martin, 1998).

This repetition suppression may be a manifestation of a fine-tuning process during which the neurons that best represent the object remain active, while those cells that are only coarsely sensitive to the features in the image gradually stop responding (see also Desimone, 1996). Subsequently, modifying synaptic connections such that only the final outcome of this fine-tuning process is primed would result in a more efficient recognition in future encounters with the same image.

What neuronal mechanisms may mediate such a long-term effect of experience? It is reasonable to assume that priming is mediated by mechanisms of synaptic plasticity. One candidate is long-term potentiation (LTP), which, put simply, is an increased sensitivity of cells due to their prior activation. A brief, high frequency stimulation of afferent fibers induces LTP of synaptic transmission that is manifested by a facilitation of response in subsequent stimulations. For example, the threshold for inducing LTP can be reduced by previous stimulation (Christie, Stellwagen, & Abraham, 1995). LTP lasts for hours *in vitro* and months *in vivo*, and is at present the most extensively studied form of activity-dependent plasticity. It has been mainly studied in the hippocampus, but LTP has been also demonstrated in the visual cortex (Kirkwood & Bear, 1997; Hirsch & Gilbert, 1993; Komatsu, Toyama, Maeda, & Sakaguchi, 1981; Tsumoto & Suda, 1979). The opposite phenomenon, long-term depression (LTD), may also play a role in priming. For example, it is conceivable that the process of object recognition involves an initial activation of multiple possible interpretations before convergence to the correct identity is possible (Desimone, 1996; Ullman, 1995). LTD may then be used to inhibit the representations that were activated initially and rejected, and by this facilitate subsequent recognition attempts of the same object. These forms of synaptic modifications may therefore provide the cellular basis for the plasticity involved in priming, and may explain how the cortex changes dynamically based on use and experience (Buonomano & Merzenich, 1998).

In light of the evidence reviewed above, it is possible that the presentation of a prime image sensitizes the cells that represent information specific to this particular image by potentiating the corresponding synapses. In subsequent encounters with the same object, at least some of the “detectors” that are activated and lead to the representation of that object are already more sensitive due to a decreased activation threshold, and recognition will therefore proceed faster and with fewer errors compared with the recognition of a novel image.

In the present account, the primed image “leaves” a neural trace that renders the specific path more sensitive than non-primed paths (see also Ullman, 1995). A mechanism in which cortical paths become sensitized with experience is capable of explaining the reduced activity evidenced in physiological and imaging studies, as well as the reduced reaction times in psychophysical experiments of priming. Indeed, the cortical response to familiar objects not only reduces (Ringo, 1996; Miller, Li, & Desimone, 1991), but it also peaks earlier with repetitions (Li et al., 1993). For example, in the study of Li et al., the activity of the relevant cell population in IT initially distinguished between novel and familiar objects after 100 msec from the onset of the cortical response (approximately 180 msec from stimulus onset). The delay of this diagnostic activity, however, was reduced to only 10 msec following additional presentations. In other words, the activity suppression begins earlier with repetitions. This shortening of the response to a familiar stimulus may be a manifestation of the efficiency involved in priming.

APPLYING THE PROPOSED ACCOUNT OF PRIMING TO EXPLAIN ROTATION COSTS

Sensitizing cells that represent a certain visual feature of the object (e.g., an oriented bar, a certain spatial frequency, an L-vertex, parts, spatial relations) will facilitate the detection of the same feature in subsequent encounters. In other words, when the same feature is repeated, its processing will proceed faster because the neurons that represent this feature are more sensitive. Therefore, the more features that two consecutive presentations of an object have in common, the more gain in performance is expected, and the more priming will be demonstrated. This idea is elaborated next, following a brief overview of the functional organization of cells in the visual cortex.

The visual ventral pathway is widely believed to be responsible for shape processing for the purpose of object recognition. Neurons along this pathway differ in their feature selectivity and the size and location of their receptive fields (RFs). Cells in V1 are organized by columns of mutual preference to basic features such as orientation and retinal location (Hubel & Wiesel, 1962). Cells in the intermediate ventral area V4 appear to respond maximally to polar and hyperbolic stimuli (Gal-

lant, Connor, Rakshit, Lewis, & Van Essen, 1996) as well as to vertices (Pasupathy & Connor, 1999). Cells in IT are more sensitive to multipart patterns (Tanaka, 1993), viewpoint-invariant properties (Vogels, Biederman, Bar, & Lorincz, 2001), and faces (Perrett, Rolls, & Caan, 1982). This hierarchical organization results in an increasingly complex processing towards the output of the ventral pathway such that cells in TE (the anterior region of IT) respond to very complex features, over large regions of the visual field.

Generally, it is assumed that cells that have been activated are subsequently primed and therefore sensitized. Therefore, activating the same or different group of cells on subsequent encounters with the same object is expected to make a difference in the time it takes to analyze the image and access even the same representation. Two identical presentations will activate largely overlapping groups of cells. For example, if the repeated image contains a straight line in a certain orientation and position in space, the cells in V1 that are sensitive to lines in this orientation and whose RF is located at that specific position will be activated in both encounters. If, however, the second image is identical but shifted by 5° relative to the first presentation, then, although the same lines are present in both presentations, they occupy a different position of the visual field. Consequently, most of the V1 cells that responded to the specific straight line on the first encounter will not respond to this line on the second encounter because it is positioned outside of their small RF.³ Therefore, a transformation such as position shift may reduce the overlap between the groups of cells that respond to the first and second presentation of the same picture.

The same analysis can be made with regard to change of feature appearance rather than change of position. Assume that the object is rotated in depth (which is analogous to changing the viewing point). Although it is possible that the same part of the object remains within the same RF on both presentations, the relative rotation may modify the appearance of that part. As a result, different cells will be active in the two encounters; cells that have similar RFs but a different “preference.”

The modifications in shape appearance that are caused by rotation are expected to have a similar effect also on the activity of cells in higher visual areas such as V4 and IT, but with more complex features. Indeed, the cells in IT that respond to complicated patterns (Tanaka, 1996) are believed to be organized in columns such that neighboring cells within a column (but not between columns) respond to slightly different stimuli (e.g., different ellipses). These slight differences may correspond to the various possible appearances of the same feature under various transformations.

It was convenient to describe this hypothesis in terms of response of cells in V1, mainly because this is the most extensively studied visual area and therefore the response properties of V1 neurons are understood

better than in any other visual area. In the real world, however, priming is expected to be less pronounced in V1 than in higher visual areas because of a ceiling effect of sensitivity that is described next.

In our environment, a considerable amount of visual information usually intervenes between two different encounters with the same object. These intervening images typically consist of many oriented lines, which repeat often. Consequently, the extent of overlap between the groups of cells in V1 that respond in different encounters is less relevant because most cells will be primed most of the time. Priming of higher areas, however, is expected to be more object-specific because the representations there are of more complex features, that, unlike oriented lines, appear in fewer contexts (e.g., an elephant's ear).

In general, there seems to be an inverse relation between the complexity of a visual feature and the frequency of its occurrence in the environment. Therefore, given the hierarchical selectivity of cells along the ventral visual cortex, the contribution of priming to recognition performance from novel viewpoints is expected to be more pronounced in higher visual areas, with the repetition of complex features. In other words, rotation in depth that modifies relatively higher-level features, such as those that are analyzed by IT (Tanaka, 1996), is expected to result in a larger cost in recognition performance compared with rotation that only alters low-level features (e.g., orientation of lines).⁴ This may explain why even in studies that applied considerable rotation, recognition was still generally viewpoint invariant (e.g., Biederman & Bar, 1999; Biederman & Gerhardstein, 1993). In those studies, rotation modified low-level features that are analyzed by V1/V2 (e.g., location and orientation of lines), but left intact the higher-level features (termed "nonaccidental properties" in the context of those specific studies). The modified V1-like features did not affect performance because of the sensitivity ceiling effect described above, and the high-level features that have the potential of affecting performance were not modified. Consequently, the rotation did not have a substantial effect on the degree of overlap between the cells that responded to the first and second presentations.

The situation is different in experiments that study viewpoint invariance where subjects have to compare sequentially presented images. Because typically no objects intervened between the two images that had to be compared in those experiments, priming in V1 can make a difference and may account for at least part of the viewpoint dependency reported (e.g., Ellis & Allport, 1986).⁵ Indeed, when the two images to be compared were presented sequentially but in different positions on the screen, and therefore activated different groups of V1 cells altogether, the advantage of using sensitized cells in V1 diminished and performance was less affected by changes in viewpoint (Biederman & Bar, 1999).

IMPLICATIONS AND CONCLUSION

An alternative interpretation of viewpoint-dependent recognition performance was provided. It was suggested that viewpoint-dependency by itself cannot be indicative of viewer-centered object representation in the brain. Instead, it may be a manifestation of the use of neural paths with different levels of sensitization. Rotation in depth makes available different object features, and modifies familiar ones. Therefore, the neural path that will be used when an object is encountered from a novel viewing point will be only partially primed, if at all, because at least some of the neurons that mediate recognition in this instance have not been sensitized by previous experience. Consequently, the possible dependency of recognition performance on the familiarity with a viewpoint may reflect the accessibility of the object representation rather than the existence of multiple representations for each object.⁶ As has been suggested in the context of the mental imagery debate (Anderson, 1978), the nature of a representation should be discussed with relation to the processes acting on that representation.

A related phenomenon that can be explained by the account proposed here is that of *canonical views*. It has been demonstrated (Palmer, Rosch, & Chase, 1981) that objects have a "preferred" viewpoint from which they can be more easily recognized. These canonical views are correlated with participants' judgment. Objects were presented from various viewpoints. The preferred, canonical view was usually a three-quarter view (a view that can be imagined as located on the vector projected from the center of the object to the intersection front-side-top).⁷ It might be that the bias in preference corresponds to the maximal visibility of most surfaces and parts in these views, which is expected to facilitate the extraction of the information necessary for recognition. It is also possible, however, that at least part of the preference stems from previous experience with specific views of these objects, and the subsequent cortical modifications that occurred over time. The objects that Palmer et al. (1981) have used happen to be often viewed also in the real world from their three-quarter view (i.e., chair). Therefore, it might be that the frequency of this view in the real world resulted in stronger connections that consequently made the three-quarter view advantageous for recognition compared with other views. One key test of this hypothesis is to use the same task with objects that are not usually seen from a three-quarter view (e.g., an airplane or an ant, which are more often seen from top/bottom views). The prediction of this hypothesis is that the preferred views for those objects will be the more frequent ones, and not necessarily the three-quarter views.

The account proposed here may also illuminate the counterintuitive fast-"same" phenomenon (Nickerson, 1967). Fast-"same" refers to the finding that in tasks involving "same"- "different" judgments, the "same"

responses are often made more rapidly than the “different” responses. It is conceivable that the fast-“same” effect is a manifestation of increased sensitization of the neural path required to activate the representation of the familiar object. In other words, the object in a “same” trial has been primed by the first presentation and is therefore accessed more readily in the subsequent presentation. To test this idea, one could use a set of highly familiar objects repeatedly, all of which participate both in “same” and in “different” trials. This design minimizes the role of priming and the prediction is that the fast-“same” effect would diminish.

Finally, many of the studies that reported viewpoint-dependent recognition have used either paperclips, wrinkled papers, ropes, or amoebae-like objects as stimuli (Tarr & Bülthoff, 1995; Edelman & Bülthoff, 1992; Poggio & Edelman, 1990; Rock & DiVita, 1987). Objects within these categories are highly similar to each other. The hypothesis presented here can explain why viewpoint-dependency is mostly reported in experiments that used such stimuli and, in general, why it is harder to distinguish between objects that are structurally similar (Joseph et al., 2000).

Highly similar stimuli are expected to be represented by neighboring and overlapping groups of cells, and the neural paths used to recognize similar objects are therefore also expected to be neighboring and overlapping. By definition, a certain visual feature activates maximally the cells that are most sensitive to it, but it also activates to some extent cells that are most sensitive to similar but slightly different features (e.g., two different ellipses; Tanaka, 1996). Consequently, if the neural path that leads to a slightly different object representation is more sensitized than this of the actual object, it may facilitate the propagation of signals from V1 to IT relative to the appropriate but nonsensitized path, and lead to fast activation of the wrong representation. Put simply, in the case of discriminating between very similar stimuli, a neural path that has been previously primed by similar but slightly different object will “attract” propagation and result in false-alarm. The increased reaction time that is also reported in these situations may reflect the time that it takes to verify, and sometimes correctly dismiss, the fast but incorrect interpretation of the input image.

To summarize, viewpoint-dependency in object representation cannot distinguish between the effect of accessibility and representation type. Therefore, new experimental paradigms will have to be devised to study the validity of viewer-centered object representation.

Acknowledgments

I thank two anonymous reviewers, N. G. Kanwisher, S. M. Kosslyn, R. A. Poldrack, N. Rubin, and J. S. Simons. This work was supported by the McDonnell-Pew Program in Cognitive Neuroscience No. 99-6 CNS-QUA.05.

Reprint requests should be sent to Moshe Bar, Department of Psychology, Harvard University, Room 864, William James Hall,

Cambridge, MA 02138, USA, or via e-mail: bar@nmr.mgh.harvard.edu.

Notes

1. See Mel (1997) and Edelman and Intrator (2000) for recent models that do not necessarily fall into these main categories.
2. Under certain conditions, which are beyond the scope of the present analysis, stimulus repetition actually results in activity increase.
3. RFs of cells in monkeys' V1 range from 0.1° to 2.0° (Van Essen, Newsome, & Maunsell, 1984; Dow, Snyder, Vautin, & Bauer, 1981).
4. Note that the complexity of the features that are modified by rotation is not necessarily proportional to the degree of rotation. The same rotation may yield profoundly different changes in the image, depending on the visibility of the diagnostic and familiar visual features both before and after rotation.
5. When a mask containing many lines in various orientations followed each of the images to be sequentially compared (Tarr et al., 1998), recognition performance can still appear viewpoint-dependent. This may indicate that priming of cells is not occurring automatically following a neural activation, but rather may involve context-dependent feedback signals. According to this hypothesis, the lines in the mask did not prime V1 cells because they were less relevant to the task.
6. It is not meant that each object has only a single, all-purpose representation. For example, the side and the top of a bicycle are expected to be represented separately. As long as the same diagnostic features are present in the image, however, a single representation could suffice.
7. In the few cases where the preferred view was not three-quarter (e.g., a clock), subjects preferred frontal view.

REFERENCES

- Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*, 249–277.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, *39*, 2885–2899.
- Biederman, I., & Cooper, E. E. (1992). Size-invariance in visual object priming. *Journal of Experimental Psychology, Human Perception and Performance*, *18*, 121–133.
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for 3D viewpoint invariance. *Journal of Experimental Psychology, Human Perception and Performance*, *19*, 1162–1182.
- Blaxton, T. A., Bookheimer, S. Y., Zeffiro, T. A., Figliozzi, C. M., Gaillard, W. D., & Theodore, W. H. (1996). Functional Mapping of human memory using PET: Comparisons of conceptual and perceptual tasks. *Canadian Journal of Experimental Psychology*, *50*, 42–56.
- Booth, M. C., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, *8*, 510–523.
- Brown, M. W., Wilson, F. A. W., & Riches, I. P. (1987). Neuronal evidence that inferomedial temporal cortex is more important than hippocampus in certain processes underlying recognition memory. *Brain Research*, *409*, 158–162.
- Buckner, R., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D. L., Rosen, B. R., & Dale, A. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation of event-related fMRI. *Neuron*, *20*, 285–296.
- Bülthoff, H., & Edelman, S. (1992). Psychophysical support

- for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, U.S.A.*, 89, 60–64.
- Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience*, 21, 149–186.
- Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, 8, 322–325.
- Christie, B., Stellwagen, D., & Abraham, W. C. (1995). Reduction of the threshold for long-term potentiation by prior theta-frequency synaptic activity. *Hippocampus*, 5, 52–59.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, U.S.A.*, 93, 13494–13499.
- Dow, B. M., Snyder, A. Z., Vautin, R. G., & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. *Experimental Brain Research*, 44, 213–228.
- Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3D objects. *Vision Research*, 32, 2385–2400.
- Edelman, S., & Intrator, N. (2000). (Coarse coding of shape fragments) + (retinotopy) ~ representation of structure. *Spatial Vision*, 13, 255–264.
- Ellis, R., & Allport, D. A. (1986). Multilevel levels of representation for visual objects: A behavioral study. In A. Cohn & J. Thomas (Eds.), *Artificial intelligence and its implications* (pp. 245–247). New York: Wiley.
- Gallant, J., Connor, C., Rakshit, S., Lewis, J., & Van Essen, D. V. (1996). Neural responses to polar, hyperbolic, and cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76, 7060–7078.
- Hirsch, J., & Gilbert, C. (1993). Long-term changes in synaptic strength along specific intrinsic pathways in the cat visual cortex. *Journal of Physiology*, 461, 247–262.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106–154.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480–517.
- Joseph, J. E., Jones, K. M., Zeffiro, T. A., & Eden, G. F. (2000). fMRI correlates of structural similarity in object naming. *Society for Neuroscience Abstracts*, 26, Part II, 1502.
- Kirkwood, A., & Bear, M. (1997). LTP and LTD in the visual cortex. In M. Baudry & J. Davis (Eds.), *Long-term potentiation*, vol. 3, (chap. 8, pp. 137–155). Cambridge: MIT Press.
- Komatsu, Y., Toyama, K., Maeda, J., & Sakaguchi, H. (1981). Long-term potentiation investigated in a slice preparation of striate cortex of young kittens. *Neuroscience Letters*, 4, 269–274.
- Li, L., Miller, E. K., & Desimone, R. (1993). The representations of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, 69, 1918–1929.
- Lowe, D. (1984). *Perceptual organization and visual recognition*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.
- Mel, B. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, 9, 777–804.
- Miller, E. K., Li, L., & Desimone, R. (1991). A neuronal mechanism for working and recognition memory in inferior temporal cortex. *Science*, 254, 1377–1379.
- Nickerson, R.S. (1967). "Same"-"different" response times with multi-attribute stimulus differences. *Perceptual and Motor Skills*, 24, 543–544.
- Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135–151). Hillsdale, NJ: Erlbaum.
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, 82, 2490–2502.
- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329–342.
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343, 263–266.
- Rainer, G., & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, 27, 179–189.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition cortex. *Nature Neuroscience*, 2, 1019–1025.
- Ringo, J. L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. *Behavioral Brain Research*, 76, 191–197.
- Rock, I., & DiVita, J. (1987). A case of viewer-centered perception. *Cognitive Psychology*, 19, 280–293.
- Rugg, M., Soardi, M., & Doyle, M. C. (1995). Modulation of event-related potentials by the repetition of drawings of novel objects. *Cognitive Brain Research*, 3, 17–24.
- Schacter, D. L., & Buckner, R. (1998). Priming and the brain. *Neuron*, 20, 185–195.
- Sloman, S. A., Hayman, C. A. G., Ohta, N., Law, J., & Tulving, E. (1998). Forgetting in primed fragment completion. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 14, 223–239.
- Squire, L. R., Ojemann, J., Miezin, F., Petersen, S., Videen, T., & Raichle, M. (1992). Activation of the hippocampus in normal humans: A functional anatomical study of memory. *Proceedings of the National Academy of Sciences, U.S.A.*, 89, 1837–1841.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, 262, 685–688.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19, 109–139.
- Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1995). *Journal of Experimental Psychology, Human Perception and Performance*, 21, 1494–1505.
- Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience*, 1, 275–277.
- Tsumoto, T., & Suda, K. (1979). Cross-depression: An electrophysiological manifestation of binocular competition in the developing visual cortex. *Brain Research*, 18, 190–194.
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247, 301–306.
- Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32, 193–254.
- Ullman, S. (1995). Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, 1, 1–11.
- Van Essen, D. C., Newsome, W. T., & Maunsell, J. H. R. (1984). The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. *Vision Research*, 24, 429–448.
- Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to non-accidental than metric shape differences. *Journal of Cognitive Neuroscience*, 13, 444–453.
- Wiggs, C., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, 8, 227–233.