

Visual Object Recognition: How Your Brain Identifies Reality

Authored by
mohammad looti

June 16, 2026

RECOMMENDED CITATION

mohammad looti (2026). *Visual Object Recognition: How Your Brain Identifies Reality*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=38120>

Object recognition is the ability to perceive an object's physical properties (such as shape, colour and texture) and apply semantic attributes to it (such as identifying the object as an apple). This process includes the understanding of its use, previous experience with the object, and how it relates to others. Regardless of an object's position or illumination, humans possess the ability to effectively identify and label an object. Humans are one of the few species that possess the ability of invariant visual object recognition. Both "front end" (knowledge/goal driven) and "back end" (sensory driven) processing are required for a species to be able to recognize objects at varying distances, angles, lighting, etc....

Basic Stages of Object Recognition

One model of object recognition, based on neuropsychological evidence, provides information that allows us to divide the process into four different stages.

Stage 1 Processing of basic object components, such as colour, depth, and form.

Stage 2 These basic components are then grouped on the basis of similarity, providing information on distinct edges to the visual form. Subsequently, figure-ground segregation is able to take place.

Stage 3 The visual representation is matched with structural descriptions in memory.

Stage 4 Semantic attributes are applied to the visual representation, providing meaning, and thereby recognition.

It should be noted that, within these stages, there are more specific processes that take place to complete the different processing components. In addition, other existing models propose integrative hierarchies (top-down and bottom-up), as well as parallel processing, as opposed to this general bottom-up hierarchy.

Hierarchical Recognition Processing

Visual recognition processing has been typically viewed as a bottom-up hierarchy in which information is processed sequentially with increasing complexities, where lower-level cortical processors, such as the primary visual cortex, are at the bottom of the processing hierarchy and higher-level cortical processors, such as the inferotemporal cortex (IT), are at the top, where recognition is facilitated. A most recognized bottom-up hierarchical theory is David Marr's theory of vision. In contrast, an increasingly popular recognition processing theory, is that of top-down processing. One model, proposed by Moshe Bar (2003), describes a "shortcut" method in which early visual inputs are sent, partially analyzed, from the early visual cortex to the prefrontal cortex (PFC). Possible interpretations of the crude visual input is generated in the PFC and then sent to the inferotemporal cortex (IT) subsequently activating relevant object representations which are then incorporated into the slower, bottom-up process. This "shortcut" is meant to minimize the amount of object representations required for matching thereby facilitating object recognition.

Lesion studies have supported this proposal with findings of slower response times for individuals with PFC lesions, suggesting use of only the bottom-up processing.

Object Constancy and Theories of Object Recognition

A significant aspect of object recognition is that of object constancy: the ability to recognize an object across varying viewing conditions. These varying conditions include object orientation, lighting, and object variability (size, colour, and other within-category differences). For the visual system to achieve object constancy, it must be able to extract a commonality in the object description across different viewpoints and the retinal descriptions. Participants who did categorization and recognition tasks while undergoing a functional magnetic found as increased blood flow indicating activation in specific regions of the brain. The categorization task consisted of participants placing objects from canonical or unusual views as either indoor or outdoor objects. The recognition task occurs by presenting the participants with images that they had viewed previously. Half of these images were in the same orientation as previously shown, while the other half were presented in the opposing viewpoint. The brain regions implicated in mental rotation, such as the ventral and dorsal visual pathways and the prefrontal cortex, showed the greatest increase in blood flow during these tasks, demonstrating that they are critical for the ability to view objects from multiple angles. Several theories have been generated to provide insight on how object constancy may be achieved for the purpose of object recognition including, viewpoint-invariant, viewpoint-dependent and multiple views theories.

Viewpoint-Invariant Theories

Viewpoint-invariant theories suggest that object recognition is based on structural information, such as individual parts, allowing for recognition to take place regardless of the object's viewpoint. Accordingly, recognition is possible from any viewpoint as individual parts of an object can be rotated to fit any particular view. This form of analytical recognition requires little memory as only structural parts need to be encoded, which can produce multiple object representations through the interrelations of these parts and mental rotation. Participants in a study were presented with one encoding view from each of 24 preselected objects, as well as five filler images. Objects were then represented in the central visual field at either the same orientation or a different orientation than the original image. Then participants were asked to name if the same or different depth-orientation views of these objects presented. The same procedure was then executed when presenting the images to the left or right visual field. Viewpoint-dependent priming was observed when test views were presented directly to the right hemisphere, but not when test views were presented directly to the left hemisphere. The results support the model that objects are stored in a manner that is viewpoint dependent because the results did not depend on whether the same or a different set of parts could be recovered from the different-orientation views.

3-D Model Representation

This model, proposed by Marr and Nishihara (1978), states that object recognition is achieved by matching 3-D model representations obtained from the visual object with 3-D model representations stored in memory as veridical shape precepts. Through the use of computer programs and algorithms, Yi Yungfeng (2009) was able to demonstrate the ability for the human brain to mentally construct 3D images using only the 2D images that appear on the retina. Their model also demonstrates a high degree of shape constancy conserved between 2D images, which allow the 3D image to be recognized. The 3-D model representations obtained from the object are formed by first identifying the concavities of the object, which separate the stimulus into individual parts. Recent research suggests that an area of the brain, known as the caudal intraparietal area (CIP), is responsible for storing the slant and tilt of a plan surface that allow for concavity recognition. Rosenburg et al. implanted monkeys with a scleral search coil for monitoring eye position while simultaneously recording single neuron activation from neurons within the CIP. During the experiment, monkeys sat 30 cm away from an LCD screen that displayed the visual stimuli. Binocular disparity cues were displayed on the screen by rendering stimuli as green-red anaglyphs and the slant-tilt curves ranged from 0 to 330. A single trial consisted of a fixation point and then the presentation of a stimulus for 1 second. Neuron activations were then recorded using the surgically inserted microelectrodes. These single neuron activations for specific concavities of objects lead to the discovery that each axis of an individual part of an object containing concavity are found in memory stores. Identifying the principal axis of the object assists in the normalization process via mental rotation that is required because only the canonical description of the object is stored in memory. Recognition is acquired when the observed object viewpoint is mentally rotated to match the stored canonical description.

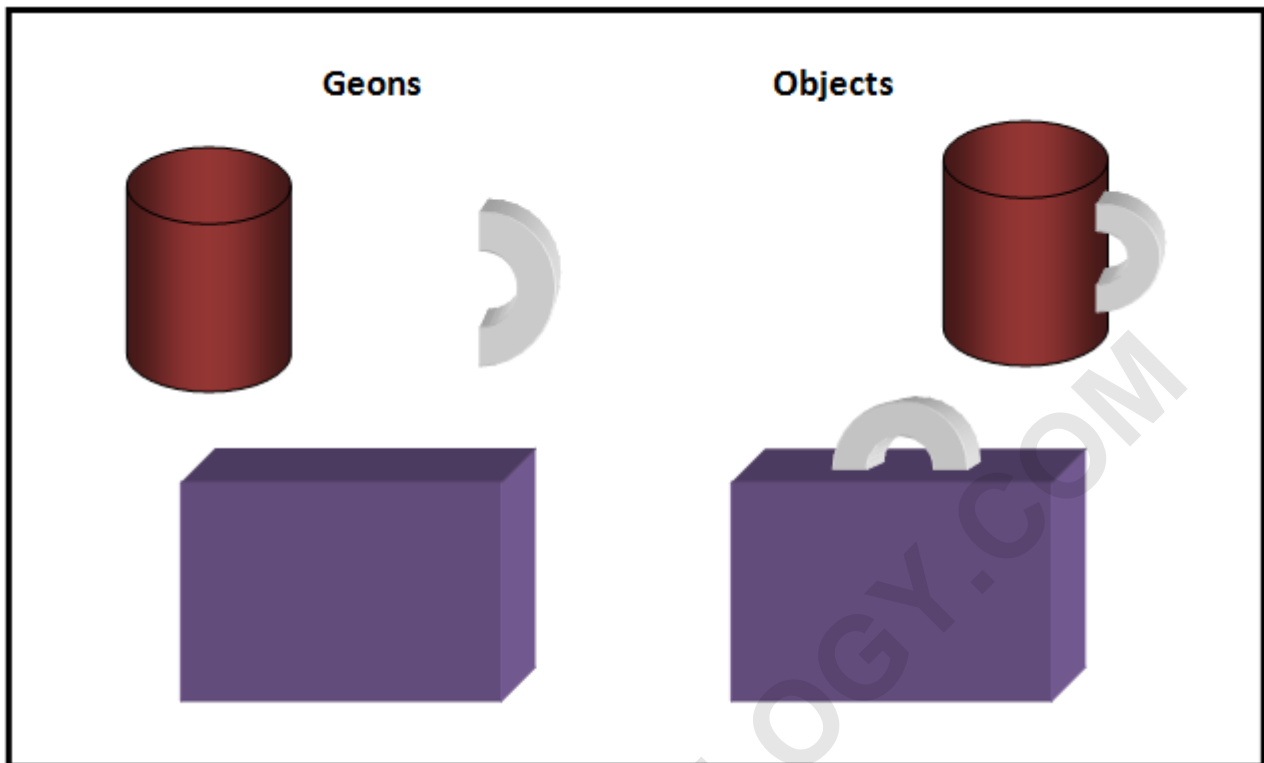


Figure 1. This image, created based on Biederman's (1987) Recognition by Components theory, is an example of how objects can be broken down into Geons.

Recognition by Components

An extension of Marr and Nishihara's model, the recognition-by-components theory, proposed by Biederman (1987), proposes that the visual information gained from an object is divided into simple geometric components, such as blocks and cylinders, also known as "geons" (geometric ions), and are then matched with the most similar object representation that is stored in memory to provide the object's identification (see Figure 1).

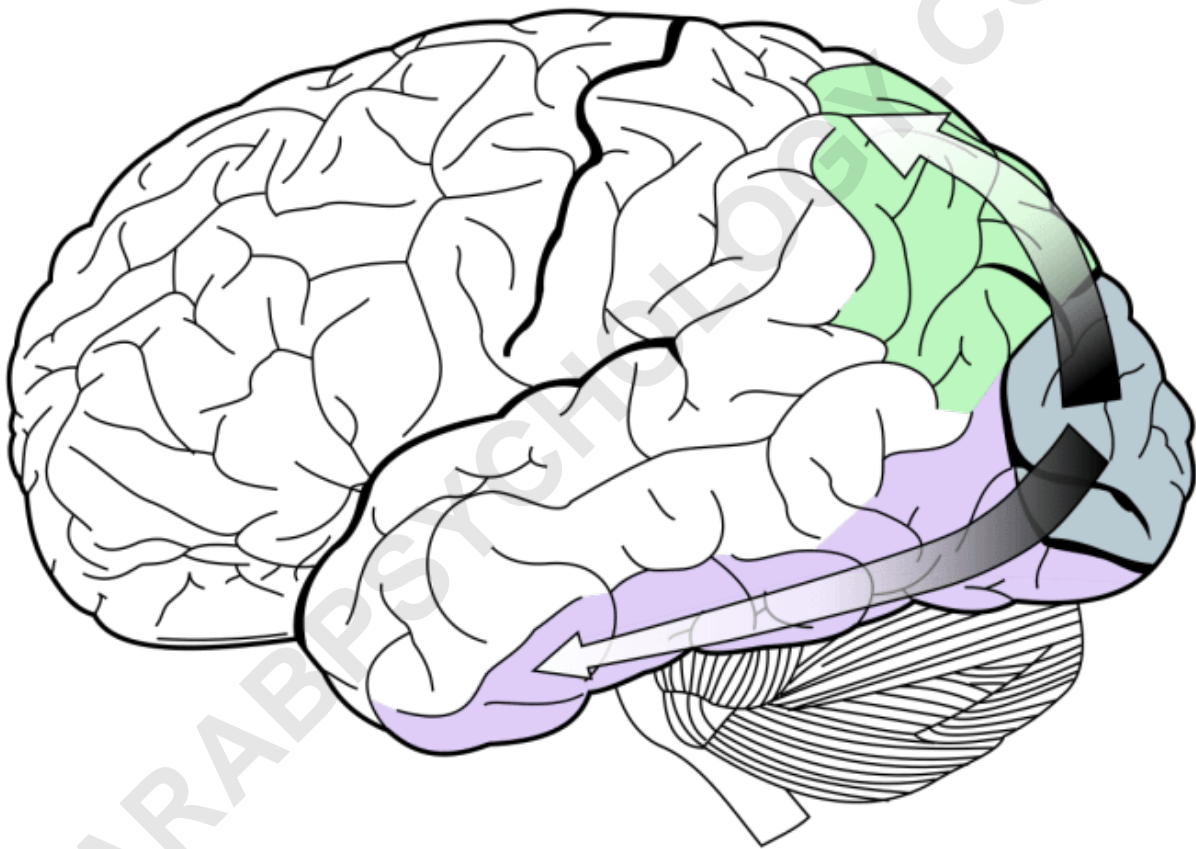
Viewpoint-Dependent Theories

Viewpoint-dependent theories suggest that object recognition is affected by the viewpoint at which it is seen, implying that objects seen in novel viewpoints reduce the accuracy and speed of object identification. This theory of recognition is based on a more holistic system rather than by parts, suggesting that objects are stored in memory with multiple viewpoints and angles. This form of recognition requires a lot of memory as each viewpoint must be stored. Accuracy of recognition also depends on how familiar the observed viewpoint of the object is.

Multiple Views Theory

This theory proposes that object recognition lies on a viewpoint continuum where each viewpoint is recruited for different types of recognition. At one extreme of this continuum, viewpoint-dependent mechanisms are used for within-category discriminations, while at the other extreme, viewpoint-invariant mechanisms are used for the categorization of objects.

Neural Substrates



The Dorsal Stream is shown in green and the Ventral Stream in purple.

The Dorsal and Ventral Stream

The visual processing of objects in the brain can be divided into two processing pathways: the dorsal stream (how/where), which extends from the visual cortex to the parietal lobes, and ventral stream (what), which extends from the visual cortex to the inferotemporal cortex (IT). The existence of these two separate visual processing pathways was first proposed by Ungerleider and

Mishkin (1982) who, based on their lesion studies, suggested that the dorsal stream is involved in the processing of visual spatial information, such as object localization (where), and the ventral stream is involved in the processing of visual object identification information (what). Since this initial proposal, it has been alternatively suggested that the dorsal pathway should be known as the 'How' pathway as the visual spatial information processed here provides us with information about how to interact with objects, For the purpose of object recognition, the neural focus is on the ventral stream.

Functional Specialization in the Ventral Stream

Within the ventral stream, various regions of proposed functional specialization have been observed in functional imaging studies. The brain regions most consistently found to display functional specialization are the fusiform face area (FFA), which shows increased activation for faces when compared with objects, the parahippocampal place area (PPA) for scenes vs. objects, the extrastriate body area (EBA) for body parts vs. objects, MT+/V5 for moving stimuli vs. static stimuli, and the Lateral Occipital Complex (LOC) for discernible shapes vs. scrambled stimuli. (See also: Neural processing for individual categories of objects)

Structural Processing: the Lateral Occipital Complex

The lateral occipital complex (LOC) has been found to be particularly important for object recognition at the perceptual structural level. In an event-related fMRI study that looked at the adaptation of neurons activated in visual processing of objects, it was discovered that the similarity of an object's shape is necessary for subsequent adaptation in the LOC, but specific object features such as edges and contours are not. This suggests that activation in the LOC represents higher-level object shape information and not simple object features. In a related fMRI study, the activation of the LOC, which occurred regardless of the presented object's visual cues such as motion, texture, or luminance contrasts, suggests that the different low-level visual cues used to define an object converge in "object-related areas" to assist in the perception and recognition process. None of the mentioned higher-level object shape information seems to provide any semantic information about the object as the LOC shows a neuronal response to varying forms including non-familiar, abstract objects.

Further experiments have proposed that the LOC consists of a hierarchical system for shape selectivity indicating greater selective activation in the posterior regions for fragments of objects whereas the anterior regions show greater activation for full or partial objects. This is consistent with previous research that suggests a hierarchical representation in the ventral temporal cortex where primary feature processing occurs in the posterior regions and the integration of these features into a whole and meaningful object occurs in the anterior regions.

Semantic Processing

Semantic associations allow for faster object recognition. When an object has previously been associated with some sort of semantic meaning, people are more prone to correctly identify the object. Research has shown that semantic associations allow for a much quicker recognition of an object, even when the object is being viewed at varying angles. When objects are viewed at increasingly deviated angles from the traditional plane of view, objects that held learned semantic associations had lower response times compared to objects that did not hold any learned semantic associations. Thus, when object recognition becomes increasingly difficult, semantic associations allow recognition to be much easier. Similarly, a subject can be primed to recognize an object by observing an action that is simply related to the target object. This shows that objects have a set of sensory, motor and semantic associations that allow a person to correctly recognize an object. This supports the claim that the brain utilizes multiple parts when trying to accurately identify an object.

Through information provided from neuropsychological patients, dissociations of recognition processing have been identified between structural and semantic processing as structural, colour, and associative information can be selectively impaired. In one PET study, areas found to be involved in associative semantic processing include the left anterior superior/middle temporal gyrus and the left temporal pole comparative to structural and colour information, as well as the right temporal pole comparative to colour decision tasks only. These results indicate that stored perceptual knowledge and semantic knowledge involve separate cortical regions in object recognition as well as indicating that there are hemispheric differences in the temporal regions.

Research has also provided evidence which indicates that visual semantic information converges in the fusiform gyri of the inferotemporal lobes. In a study that compared the semantic knowledge of category versus attributes, it was found that they play separate roles in how they contribute to recognition. For categorical comparisons, the lateral regions of the fusiform gyrus were activated by living objects, in comparison to nonliving objects which activated the medial regions. For attribute comparisons, it was found that the right fusiform gyrus was activated by global form, in comparison to local details which activated the left fusiform gyrus. These results suggest that the type of object category determines which region of the fusiform gyrus is activated for processing semantic recognition, whereas the attributes of an object determines the activation in either the left or right fusiform gyrus depending on whether global form or local detail is processed.

In addition, it has been proposed that activation in anterior regions of the fusiform gyri indicate successful recognition. However, levels of activation have been found to depend on the semantic relevance of the object. The term semantic relevance here refers to "a measure of the contribution of semantic features to the core meaning of a concept." Results showed that objects with high semantic relevance, such as artefacts, created an increase in activation compared to objects with

low semantic relevance, such as natural objects. This is due to the proposed increased difficulty to distinguish between natural objects as they have very similar structural properties which makes them harder to identify in comparison to artefacts. Therefore, the easier the object is to identify, the more likely it will be successfully recognized.

Another condition that affects successful object recognition performance is that of contextual facilitation. It is thought that during tasks of object recognition, an object is accompanied by a "context frame", which offers semantic information about the object's typical context. It has been found that when an object is out of context, object recognition performance is hindered with slower response times and greater inaccuracies in comparison to recognition tasks when an object was in an appropriate context. Based on results from a study using fMRI, it has been proposed that there is a "context network" in the brain for contextually associated objects with activity largely found in the Parahippocampal cortex (PHC) and the Retrosplenial Complex (RSC). Within the PHC, activity in the Parahippocampal Place Area (PPA), has been found to be preferential to scenes rather than objects; however, it has been suggested that activity in the PHC for solitary objects in tasks of contextual facilitation may be due to subsequent thought of the spatial scene in which the object is contextually represented. Further experimenting found that activation was found for both non-spatial and spatial contexts in the PHC, although activation from non-spatial contexts was limited to the anterior PHC and the posterior PHC for spatial contexts.

Recognition Memory

When someone sees an object, they know what the object is because they've seen it on a past occasion; this is recognition memory. Not only do abnormalities to the ventral (what) stream of the visual pathway affect our ability to recognize an object but also the way in which an object is presented to us.

Context

Context allows for a much greater accuracy in object recognition. When an identifiable object is blurred, the accuracy of recognition is much greater when the object is placed in a familiar context. In addition to this, even an unfamiliar context allows for more accurate object recognition compared to the object being shown in isolation. This can be attributed to the fact that objects are typically seen in some setting rather than no setting at all. When the setting the object is in is familiar to the viewer, it becomes much easier to determine what the object is. Though context is not required to correctly recognize, it is part of the association that one makes with a certain object.

Context becomes especially important when recognizing faces or emotions. When facial emotions are presented without any context, the ability to which someone is able to accurately describe the emotion being shown is significantly lower than when context is given. This phenomenon remains

true across all age groups and cultures, signifying that context is essential in accurately identifying facial emotion for all individuals.

Familiarity

Familiarity is a mechanism that is context-free in the sense that what one recognizes just feels familiar without spending time trying to find in what context one knows the object. The ventro-lateral region of the frontal lobe is involved in memory encoding during incidental learning and then later maintaining and retrieving semantic memories. Familiarity can induce perceptual processes different from those of unfamiliar objects which means that our perception of a finite amount of familiar objects is unique. Deviations from typical viewpoints and contexts can affect the efficiency for which an object is recognized most effectively. It was found that not only are familiar objects recognized more efficiently when viewed from a familiar viewpoint opposed to an unfamiliar one, but also this principle applies to novel objects. This deduces to the thought that representations of objects in our brain are organized in more of a familiar fashion of the objects observed in the environment. Recognition is not only largely driven by object shape and/or views but also by dynamic information. Familiarity can benefit the perception of dynamic point-light displays, moving objects, the sex of faces, and face recognition.

Recollection

Recollection shares many similarities with familiarity; however, it is context-dependent, requiring specific information from the inquired incident.

Impairments

Loss of object recognition is called visual object agnosia. There are two broad categories of visual object agnosia: apperceptive and associative. When object agnosia occurs from a lesion in the dominant hemisphere, there is often a profound associated language disturbance, including loss of word meaning.

Effects of Lesions in the Ventral Stream

Object recognition is a complex task and involves several different areas of the brain - not just one. If one area is damaged then object recognition can be impaired. The main area for object recognition takes place in the temporal lobe. For example, it was found that lesions to the perirhinal cortex in rats causes impairments in object recognition especially with an increase in feature ambiguity. Neonatal aspiration lesions of the amygdaloid complex in monkeys appear to have resulted in a greater object memory loss than early hippocampal lesions. However, in adult

monkeys, the object memory impairment is better accounted for by damage to the perirhinal and entorhinal cortex than by damage to the amygdaloid nuclei. Combined amygdalohippocampal (A + H) lesions in rats impaired performance on an object recognition task when the retention intervals were increased beyond 0s and when test stimuli were repeated within a session. Damage to the amygdala or hippocampus does not affect object recognition, whereas A + H damage produces clear deficits. In an object recognition task, the level of discrimination was significantly lower in the electrolytic lesions of globus pallidus (part of the basal ganglia) in rats compared to the Substantia-Innominata/Ventral Pallidum which was in turn worse compared to Control and Medial Septum/Vertical Diagonal Band of Broca groups; however, only globus pallidus did not discriminate between new and familiar objects. These lesions damage the ventral (what) pathway of the visual processing of objects in the brain.

Visual Agnosias

Agnosia is a rare occurrence and can be the result of a stroke, dementia, head injury, brain infection, or hereditary. Apperceptive agnosia is a deficit in object perception creating an inability to understand the significance of objects. Similarly, associative visual agnosia is the inability to understand the significance of objects; however, this time the deficit is in semantic memory. Both of these agnosias can affect the pathway to object recognition, like Marr's Theory of Vision. More specifically unlike apperceptive agnosia, associative agnosic patients are more successful at drawing, copying, and matching tasks; however, these patients demonstrate that they can perceive but not recognize. Integrative agnosia (a subtype of associative agnosia) is the inability to integrate separate parts to form a whole image. With these types of agnosias there is damage to the ventral (what) stream of the visual processing pathway. Object orientation agnosia is the inability to extract the orientation of an object despite adequate object recognition. With this type of agnosia there is damage to the dorsal (where) stream of the visual processing pathway. This can affect object recognition in terms of familiarity and even more so in unfamiliar objects and viewpoints. A difficulty in recognizing faces can be explained by prosopagnosia. Someone with prosopagnosia cannot identify the face but is still able to perceive age, gender, and emotional expression. The brain region that specifies in facial recognition is the fusiform face area. Prosopagnosia can also be divided into apperceptive and associative subtypes. Recognition of individual chairs, cars, animals can also be impaired; therefore, these object share similar perceptual features with the face that are recognized in the fusiform face area.

Alzheimer's Disease

The distinction between category and attribute in semantic representation may inform our ability to assess semantic function in aging and disease states affecting semantic memory, such as Alzheimer's disease (AD). Because of semantic memory deficits, persons suffering from

Alzheimer's disease have difficulties recognizing objects as the semantic memory is known to be used to retrieve information for naming and categorizing objects. In fact, it is highly debated whether the semantic memory deficit in AD reflects the loss of semantic knowledge for particular categories and concepts or the loss of knowledge of perceptual features and attributes.

ARABPSYCHOLOGY.COM