Definition: **depth perception** from *3D A-to-Z: An Encyclopedic Dictionary*

The ability to register depth in a 3D image or judge distance between objects thanks to a combination of binocular depth cues (made possible by disparity and vergence) and monocular depth cues, such as the relative size of different objects, occlusion, geometric perspective, motion parallax, etc. See **stereopsis**.

Summary Article: **Depth Perception**
From *Encyclopedia of the Mind*

Depth perception is the ability of observers to discriminate the distances of objects, particularly their relative distances, and identify the three-dimensional shape of surfaces. This entry concentrates on sources of visual information contributing to depth perception and separates these into two main types: those that apply when the viewer stands still with one eye closed (pictorial cues) and those that come into play when the viewer moves or opens the other eye.

It is hard to imagine perceiving the world without a perception of depth. Animals move around continually, receiving changing images as they do. If they were unable to relate these images to a stable scene they could not operate successfully in a 3D world. This may explain why clinical disorders of depth perception *per se* are so rare (as opposed to malfunctions of systems that contribute to depth perception, such as stereoscopic vision).

Generally, in nature, depth vision is a consequence of moving through the world. Most animals have eyes positioned on the side of their head so that they can look out for predators; some have a 360° field of view. They are able to recover depth information about the scene by moving. Binocular stereopsis (seeing depth with two eyes) is not necessary for depth perception and acute stereoscopic vision is almost uniquely the preserve of carnivores. It allows them to remain perfectly still in the grass and yet gain the advantages of depth vision.

**Pictorial Cues**

Of course, it is possible to perceive depth and the 3D structure of a scene without either moving or using binocular stereopsis. When people look at a photograph, which is taken from a single vantage point, they perceive depth in the scene. There are rare examples when this perception can be thoroughly misleading. For example, Figure 1 shows the Ames room, which appears to be a normal shaped room even when we know, from seeing the people in different parts of the room, that this cannot be the case. Assumptions about floors being perpendicular to walls and windows or tiles being rectangular are so strong that they affect our perception of depth and size. The Ames room makes clear that we cannot deduce the depth structure of a scene in a photograph unless we make assumptions about the scene. Remarkably often, those assumptions prove to be correct but occasionally they can fail.

**Perspective**

One pictorial cue is perspective. Parallel lines in the world project to straight lines in the image that eventually come together at a vanishing point, as shown in Figure 1, for example. This is one of the

[https://search.credoreference.com/content/topic/depth_perception](https://search.credoreference.com/content/topic/depth_perception)
reasons that the viewer is so convinced by the Ames room: It is highly unlikely that a set of lines in the image would all converge to a point unless the lines in the world were in fact parallel (but, of course, in the Ames room they are not). Over many centuries, artists have discovered how to use perspective in paintings. First, they learned to paint lines that recede toward a single vanishing point (in the 15th century, e.g., Masolino), then much later, they painted lines that recede toward two points (in the 18th century, e.g., Canaletto's pictures of buildings viewed from one corner), and finally, they drew lines that recede toward three vanishing points (in the 20th century, e.g., M. C. Escher's pictures). For these perspective cues to be informative, the visual system must assume that converging lines in the image correspond to parallel lines in the real world. This is true sufficiently often for perspective to be a powerful cue to depth.

Other Pictorial Cues

Other pictorial cues include aerial perspective (more distant objects have lower contrast and a bluish tinge), familiar size (the retinal size of a familiar object is smaller when it is further away), interposition (near objects occlude more distant ones), shading (a cue to local surface shape rather than distance), relative height in the image (very distant objects, if attached to the ground, are at the height of the horizon), texture gradients (a cue to surface slant and shape), and so on. When artists paint a picture they use many of these cues to create an illusion, fooling the viewer into believing that there is depth in the scene when in fact the picture is flat.

Multiple-View Cues

A large trompe l'oeil picture can be so realistic that if the viewer shuts one eye and keeps their head still, it can appear extremely realistic. Opening the other eye and moving their head breaks the illusion. Some pavement artists can create a similar effect: Viewed from one place, the picture is compelling, but walk around the picture and the vivid perception of depth disappears. This example, like the Ames
room, illustrates the fact that multiple views of a scene give much more information about the 3D structure of a scene than a single view. If the scene were real, with objects at different depths, then the image would change radically as the viewer moves. Objects would be dynamically occluded (move behind one another) as the viewer moves, but this does not happen with a painting. The change in relative position of features in the image is often described as motion parallax. Binocular stereopsis is just a special case of parallax, with the left and right eyes providing two samples from the much larger set of images that the observer receives as they move.

**Breaking Camouflage**

In everyday life, pictorial cues and multiple view cues coexist in images, but the two can be separated. Motion parallax and stereopsis give rise to the perception of depth without there being any pictorial cues present. Figure 2, for example, shows a pair of images that contain no pictorial cues or, at least, any pictorial cues that there are indicate that these are flat surfaces. Yet when the two images are viewed together as a stereo pair or as alternating frames in a movie sequence, a shape emerges, in this case a square. The visual system has broken the camouflage that entirely obscured the object. Camouflage in nature is not usually perfect, as it is here, but it can be very subtle. Stereopsis or motion parallax is a crucial way of overcoming it.

**Feature Matching**

To identify depth in the images shown in Figure 2, the visual system must match up corresponding features in the left and right eyes or across time in the case of the movie sequence. Ultimately, the feature matching must occur at a scale as small as the dots in the Figure 2; otherwise, it would not be possible to identify the sharp boundary between the square and its surround. It is likely, however, that the initial matching up is done between coarser scale features than this, which makes the matching process much simpler. The matching process can then be repeated at progressively finer scales to refine the depth estimate.

![Figure 2 Binocular or motion information without pictorial cues](https://search.credoreference.com/content/topic/depth_perception)

**Recovering Object Shape**

Once features have been matched, their depth must be calculated. This can be done in more or less sophisticated ways. For example, working out whether one object is in front of or behind another is
easy enough, allowing the observer to detect the square in Figure 2 or to thread a needle. Some tasks require more information than this, for example, the observer may need to know the ratio of depths of features. This is called *bas relief* depth: For example, observers can recognize characters on the Parthenon frieze despite the fact that the depths of the figures have been squashed. Finally, if the task requires the true shape and size of the object to be known, for example when reaching to pick it up, then more information is needed, including, for example, an estimate of how far away the object is. Interestingly, judgments of the true shape of objects are often most accurate at about grasping distance.

**Combining Depth Information**

Under normal circumstances, pictorial and multiple-view cues to depth agree with one another and support a common interpretation. We have seen examples in which that is not the case, where multiple-view cues (binocular stereopsis and motion parallax) provide the correct answer but pictorial cues are misleading. Nevertheless, there are many cases in which pictorial cues win out over multiple-view cues when a conflict arises. For example, most viewers of a 3D movie see vivid depth in the scene, more so than in a normal movie, but if they were to turn the glasses round so that the left eye sees what the right eye should see and vice versa, the perceived depth does not reverse. This is different from a stereo pair such as that in Figure 2, where reversing the left and right eye's images results in a reversal of perceived depth (the square recedes instead of protruding). Rather, in a stereo or 3D movie, the perception when the glasses are reversed is more like that produced by viewing a nonstereo movie. In this situation, pictorial and motion cues dictate the perception, while binocular cues are ignored or vetoed because they are inconsistent with the most likely interpretation of the scene. These examples emphasize that the goal of depth perception, and of perception in general, is to make an informed guess about the nature and layout of the scene.

**Other Senses**

Although all the examples so far have been visual, depth perception is multimodal. The sound produced by an object at different distances changes in systematic ways, not only in overall loudness but also in spectral cues (high frequencies are attenuated more than low frequencies as a sound source moves further away). People are able to distinguish the distance of objects based on these spectral differences. Touch also plays a key role in depth perception, particularly in determining the shape of objects. There is increasing interest in determining the way in which haptic (touch) and visual information is combined to determine the perceived shape of surfaces.

**Choosing between Interpretations**

Researchers often divide up information about depth into categories, describing multiple depth cues within each sense, such as visual texture, perspective, shading, occlusion, and even subcues such as texture-compression, texture-size-gradient, texture-perspective-convergence and so on. Such subdivisions are, of course, artificial constructs that do not necessarily reflect real distinctions in the way the visual system processes information about depth. If the visual system's task is to choose between competing hypotheses about the scene, then many different types of information may be relevant, all contributing with a greater or lesser weight to the choice. This is a more tractable problem than deciding how the information from different modules should be combined to form a perceptual representation.

**See also**

https://search.credoreference.com/content/topic/depth_perception
Further Readings


Andrew Glennerster

APA

Chicago

Harvard

MLA


© 2013 by SAGE Publications, Inc.
APA

Chicago

Harvard

MLA