

COMPARATIVE COGNITION & BEHAVIOR REVIEWS

2010

Volume 5, pp 136 - 138

Do animals recognize pictures as representations of 3D objects?

Masako Jitsumori
Chiba University

Keywords: picture-object equivalence, picture recognition, picture perception, motion perception, categorization

A picture is something in which objects can be recognized, even though the objects themselves are not actually present when it is observed. For observers to recognize objects in a picture, visual inputs have to be matched to the representations of the visual objects in memory. Patients who suffer from visual agnosia often exhibit greater impairments in recognizing line-drawings and photographs than real objects, while leaving abilities of perceiving 2D features intact. Hiraoka, Suzuki, Hirayama, and Mori (2009) reported a patient who does not show apparent impairment with real objects and even their photographs, but has difficulties in recognizing line-drawings and silhouettes (i.e., the stimuli are missing much of 2D features of the real objects). In contrast, Turnbull, Driver, and McCarthy (2004) reported the performance of a patient on a variety of experimental tasks that investigated the patient's ability to extract 2D and 3D information from 2D stimuli. His performance was relatively intact on the 2D tasks (e.g., picture-matching across rotations in the picture plane), but was greatly impaired on the 3D tasks (mental rotation in depth, relative depth judgments within line-drawings and shaded photographs, and the discrimination of line-drawings depicting "possible" and "impossible" 3D objects). The finding suggested that his difficulties in recognizing visually presented objects was due to the deficit in deriving 3D structure from the pictorial depth-cues within 2D images. It has also been reported

that patients with agnosic deficits often show a greater impairment in the recognition of visually presented living things (e.g., animals) as compared to non-living things (e.g., tools). Wolk, Coslett, and Glosser (2005) argued that such category-specific visual agnosia suggests that information about an object's form, as well as sensory-motor information specifying the manner of manipulation, may contribute to object recognition. Generally, visual agnosia is assumed to be derived from impairments in the processes by which visual information contacts stored knowledge of objects or representations of 3D objects in memory.

In the case of humans, we can ask them to verbally describe pictures. If they can or cannot correctly describe the objects depicted in the pictures, we can tell that whether they do or do not recognize the correspondence between the pictures and real objects. In contrast, the question of picture-object correspondence in animals is far trickier. Weisman and Spetch (2010) warn us about the potential risks of assuming picture-object correspondence in the studies of animal behavior. The problem arises not only because mental images are impenetrable (i.e., private events) and the visual system of animals, such as birds, are different from ours, but also because it is very difficult – if not impossible – to rule out the possibility that animals respond to pictures merely on the basis of 2D characteristics (e.g., brightness, color, oriented lines, corners, local shape, and motion). Even if animals respond to pictures in the same way as to real objects, there is the possibility that they are responding to the pictures much in the same way as Turnbull et al.'s patient,

Correspondence concerning this article should be addressed to Masako Jitsumori, Department of Cognitive and Information Sciences, Faculty of Letters, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan. E-mail: mjitsu@L.chiba-u.ac.jp

who could copy line drawings without recognizing what the original or the duplicate depicted.

When can we say that animals, like birds, perceive a correspondence between pictures and objects? In an effort to equate the retinal images of real objects and their pictures, Weisman and Spetch (2010) propose several display methodologies that would facilitate the correspondence between real objects and their pictures to the birds' eyes. If the retinal images of a real object and its picture are to the same, their percepts will naturally be more similar. For studies that use visual stimuli as substitutions for real objects, it is important to accurately reproduce the stimuli as though they were real (e.g., the video image must be life-like in size). Studies that have measured animal preference for certain video images and those that observed animals' natural behaviors to pictures of biologically significant stimuli allow us to assess the extent to which the animals confuse real objects with their 2D representations. However, the improvement of video equipment itself may not provide a complete resolution as to whether animals perceive (or do not perceive) the picture as a representation of the real-world object in memory but not the object itself (see Fagot, Martin-Malivel, & Dépy, 1999, for the distinction between the confusion and equivalence modes of picture processing).

D'Eath (1998) claimed that video experiments are suitable when animals are using simple stimulus characteristics such as those that do not depend on depth perception. Actually, we do not really experience a visual stimulus presented on a 2D display as being "truly" 3D (unlike with stereoscopic 3D movies and shadow-casting motion in which we have vivid perceptual experience of three dimensionality or depth). Nevertheless, we can "see" real objects in pictures and videos. Representations of 3D objects include not only surface features visible in videos but also additional information about an object's 3D structure.

Pigeons, like most other animals, are consistently faced with the problem of recognizing objects in the natural environment. When a bird looks at an object in the 3D world, the turning and tilting of the head may change the visible shape of the object. Changes in the image as a consequence of locomotion are even more dramatic: as the object changes its size, position, orientation, and relations between parts, and so on. This constant interaction between self movement and changes in visual image presumably enables birds to learn the 3D structure of familiar objects in their world. Temporal associations between the image at one instance with the image at the next instance (Tarr & Bülthoff, 1998) as well as the smoothness of the physical transformation that objects undergo with time (Edelman, 1999) may facilitate the birds integrating the different views as a unified 3D object.

Self-induced motion cues are absent in videos. However, when a 3D object moves to the left or to the right in a video, the visual image of the object changes over time in accordance with its location in the picture plane – one side view disappears and the other side view appears. Such changes are not produced by self movement but yield motion parallax preserved in 2D videos (Rogers & Graham, 1979). Motion parallax could be a powerful cue to extract the appropriate 3D structure of object, so long as the viewer interprets the object as being rigid. We do not know yet whether birds, like pigeons, are able to use motion parallax in 2D video as a cue for learning the 3D structure of an object. It is even more difficult to imagine that pigeons perceive a solid object when it is rotating in depth but keeping its location in the video (but see Cook & Katz, 1999; Koban & Cook, 2009). Jitsumori and Makino (2004) found with pigeons that dynamic 3D rotations of human faces in depth contributed little beyond what would be expected from the additional views provided by the motion. The negative finding suggests that the pigeons treated the static and the dynamic views of the same face differently but does not necessarily exclude the possibility that they perceived a 3D object in the dynamic video. Undoubtedly, processing motion of 3D objects is one of the primary functions of the visual systems of diverse animal species. Comparative studies using a variety of tasks and stimuli are needed to investigate dynamic nature of object perception of the organisms living in the 3D visual world.

Weisman and Spetch (2010) are critical of the evidence that the categorization of photographs of natural objects by pigeons means that they recognize 3D objects in the photographs. I think they are absolutely correct (see Jitsumori, Natori, & Okuyama, 1999) but their criticisms of the studies comparing the categorization of natural and artificial stimuli (e.g., Lazareva, Freiburger, & Wasserman, 2004) are irrelevant. It is reasonable to assume that the structures of natural and artificial categories are different, regardless of the presence or absence of picture-object correspondence. Photographs of natural objects belonging to the same category could be highly variable in local features, reflecting the polymorphous nature of natural categories, compared to photographs of man-made objects. Comparisons in performance with natural and artificial stimuli (flowers and people vs. cars and chairs) allow us to assess the effect of category structure that determines the features used by animals to cope with the different types of categorization. We cannot tell, of course, what pigeons see in photographs of natural objects, but what should still be clarified is the mechanism underlying the robustness of pigeons' visual processing of natural objects that contain much richer information than do the man-made objects or the stimuli that vary only on a few physical dimensions.

Categorization tasks require subjects to respond similarly to different objects of the same category and to respond differently to objects in contrasting categories. On the other hand, object identification (and discrimination) tasks require subjects to respond similarly to different views of the same object and to respond differently to individual objects. Although the requirements at the object level are opposite (stability and sensitivity, according to the terminology of Marr & Nishihara, 1978), object identification and object categorization may not be different cognitive processes, but just the different stages of object recognition. Future studies in both areas should help to jointly provide answers to the questions raised by Weisman and Spetch (2010): Do humans and birds view objects in the world similarly? Do humans and birds view objects in pictures similarly? For pigeons and other birds, what is the relationship between their categorizations of 2D pictures and their categorizations of 3D objects in the real world?

References

- Cook, R. G., Katz, J. S. (1999). Dynamic object perception by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 194-210. doi:[10.1037/0097-7403.25.2.194](https://doi.org/10.1037/0097-7403.25.2.194) PMID:10331919
- D'Eath, R. B. (1998). Can video images imitate real stimuli in animal behaviour experiments? *Biological Reviews*, 73, 267-292. doi:[10.1017/S0006323198005179](https://doi.org/10.1017/S0006323198005179)
- Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA:MIT Press.
- Fagot, J., Martin-Malivel, J., Dépy, D. (1999). What is the evidence for an equivalence between objects and pictures in birds and nonhuman primates? *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, 18, 923-949.
- Hiraoka, K., Suzuki, K., Hirayama, K., Mori, E. (2009). Visual agnosia for line drawings and silhouettes without apparent impairment of real-object recognition: A case report. *Behavioural Neurology*, 21, 187-192. PMID:19996516
- Jitsumori, M., Makino, H. (2004). Recognition of static and dynamic images of depth-rotated human faces by pigeons. *Learning & Behavior*, 32, 145-156. PMID:15281387
- Jitsumori, M., Natori, M., Okuyama, K. (1999). Recognition of moving video images of conspecifics by pigeons: Effects of individuals, static and dynamic motion cues, and movement. *Animal Learning & Behavior*, 27, 303-315. <http://psycnet.apa.org/index.cfm?fa=search.displayRecord&uid=1999-11736-006>
- Koban, A. C., Cook, R. G. (2009). Rotational object discrimination by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 35, 250-265. doi:[10.1037/a0013874](https://doi.org/10.1037/a0013874) PMID:19364233
- Lazareva, O. F., Freiburger, K. L., Wasserman, E. A. (2004). Pigeons concurrently categorize photographs at both basic and superordinate levels. *Psychonomic Bulletin and Review*, 11, 1111-1117. PMID:15875984
- Marr, D., Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, 200, 269-294. doi:[10.1098/rspb.1978.0020](https://doi.org/10.1098/rspb.1978.0020)
- Rogers, B., Graham, N. (1979). Motion parallax as an independent cue for depth perception. *Perception*, 8, 125-134. doi:[10.1068/p080125](https://doi.org/10.1068/p080125) PMID:471676
- Tarr, M. J., Bülthoff, H. H. (1998). *Object recognition in man, monkey, and machine*. Cambridge, MA: The MIT Press.
- Turnbull, O. H., Driver, J., McCarthy, R. A. (2004). 2D but not 3D: Pictorial-depth deficits in a case of visual agnosia. *Cortex*, 40, 723-738. doi:[10.1016/S0010-9452\(08\)70167-9](https://doi.org/10.1016/S0010-9452(08)70167-9)
- Weisman, R., Spetch, M. (2010) Determining when birds perceive correspondence between pictures and objects: a critique. *Comparative Cognition and Behavior Reviews*, 5, 117-131. Retrieved from <http://psyc.queensu.ca/ccbr/index.html> doi: [10.3819/ccbr.2010.50006](https://doi.org/10.3819/ccbr.2010.50006)
- Wolk, D. A., Coslett, H. B., Glosser, G. (2005). The role of sensory-motor information in object recognition: Evidence from category-specific visual agnosia. *Brain and Language*, 94, 131-146. doi:[10.1016/j.bandl.2004.10.015](https://doi.org/10.1016/j.bandl.2004.10.015) PMID:15896389