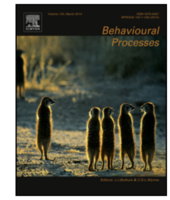
Behavioural Processes 123 (2016) 107–113

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc



Promoting rotational-invariance in object recognition despite experience with only a single view

Fabian A. Soto[a,*], Edward A. Wasserman[b]

[a] Department of Psychology, Florida International University, Miami, FL 33199, USA
[b] Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, USA

ARTICLE INFO

Article history:
Received 17 July 2015
Received in revised form 3 November 2015
Accepted 4 November 2015
Available online 28 November 2015

Keywords:
Object recognition
View invariance
Learning
Evolution of visual cognition

ABSTRACT

Different processes are assumed to underlie invariant object recognition across affine transformations, such as changes in size, and non-affine transformations, such as rotations in depth. From this perspective, promoting invariant object recognition across rotations in depth requires visual experience with the object from *multiple* viewpoints. One learning mechanism potentially contributing to invariant recognition is the error-driven learning of associations between relatively view-invariant visual properties and motor responses or object labels. This account uniquely predicts that experience with affine transformations of a *single* object view may also promote view-invariance, if view-invariant properties are also invariant across such affine transformations. We empirically confirmed this prediction in both people and pigeons, thereby suggesting that: (a) the hypothesized mechanism participates in view-invariance learning, (b) this mechanism is present across distantly-related vertebrates, and (c) the distinction between affine and non-affine transformations may not be fundamental for biological visual systems, as previously assumed.

1. Introduction

Keen interest exists in discovering how organisms achieve object recognition that is invariant across changes in identity-preserving variables, such as distance and viewpoint. Such identity-preserving variables are aspects of the viewing situation that modify the image that an object projects to the retina, without changing the object's identity. Object identity depends on more stable properties, such as its three-dimensional shape, which must be extracted by any visual system in order to achieve accurate object recognition. Most recent research has been motivated by the idea that, because objects change more slowly than do their retinal images, the brain can, without supervision, learn invariant representations from different retinal images which are merely presented in close temporal contiguity (Cox et al., 2005; Földiák, 1990; Li and DiCarlo, 2008; Stringer et al., 2006; Wallis and Bülthoff, 2001; Wiskott and Sejnowski, 2002).

A second learning mechanism that is potentially involved in invariance learning has received far less attention. Visual features that are common to multiple views of an object may come to control recognition because they reliably predict object identity (Soto et al., 2012; Wang et al., 2005; Yamashita et al., 2010). According to the Common Elements Model of object categorization and recognition in pigeons (Soto and Wasserman, 2010a, 2012a, 2014; Soto et al., 2012), the image of an object shown from a particular viewpoint is represented by the activation of a set of "elements," which can be interpreted as encoding visual properties in the image. Importantly, these properties vary widely in the level to which they are repeated across different images showing the same object. Properties can be relatively view-invariant, being repeated across many views of the same object, or they can be relatively view-specific, being idiosyncratic to a single view of an object. Several experiments have shown that pigeons do extract relatively view-invariant properties from images and rely on them for object recognition (e.g., Gibson et al., 2007; Lazareva et al., 2008). On the other hand, the fact that pigeons do not show view-invariant object recognition after training with a single object view (Peissig et al., 1999, 2000; Spetch et al., 2001; Wasserman et al., 1996) suggests that they are sensitive to rather view-specific information in the training images.

The model also proposes that the selection of which elements come to control responding in an object categorization or identification task is carried out through associative error-driven learning (see Soto and Wasserman, 2010a) implemented in the circuitry of the basal ganglia (Soto and Wasserman, 2012a, 2014). Many predictions of the model regarding the role of error-driven learning in object categorization and recognition have recently been

* Corresponding author at: Department of Psychology, Florida International University, Modesto A. Maidique Campus, 11200 SW 8th St, AHC4 460, Miami, FL 33199, USA.
E-mail address: fabian.soto@fiu.edu (F.A. Soto).

http://dx.doi.org/10.1016/j.beproc.2015.11.005
0376-6357/

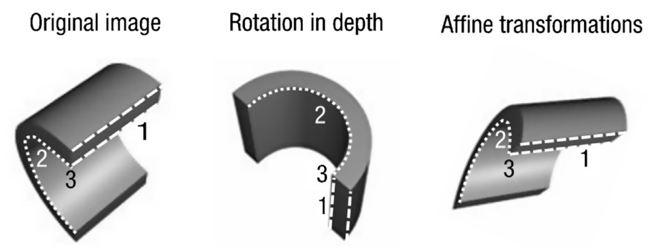


Fig. 1. Effects of rotation in depth and three affine transformations (changes in size, planar rotation, and shear) on different properties of geon images. Both transformation types change metric properties, such as the degree of edge curvilinearity (2) and the angular degree of a co-termination (3). Both transformation types leave nonaccidental properties intact: parallel edges remain parallel (1), curved edges remain curved (2), and coterminations are unchanged (3).

empirically confirmed (Soto and Wasserman, 2010a,b, 2012b; Soto et al., 2012).

Pigeons can only achieve view-invariant object recognition after explicit training with multiple views of an object (Peissig et al., 1999, 2000; Wasserman et al., 1996). According to the Common Elements Model, when an organism experiences multiple views of an object, common properties are presented often and learning about them is faster than learning about properties that are idiosyncratic to each view (Soto et al., 2012). This "repetition advantage" for properties that tend to be invariant across changes in viewpoint should emerge regardless of whether different views are experienced in close temporal contiguity (Wang et al., 2005).

This logic suggests that any manipulation aimed at reproducing a repetition advantage effect for properties that are common across changes in viewpoint should lead to view-invariance learning. In the present study, we designed just such a manipulation by taking advantage of the fact that a class of simple objects, "geons[1]," contains a number of identifiable properties that are shared by most views of a single object, termed "nonaccidental properties" (Biederman, 1987). Rotation in depth of a geon induces changes in several accidental properties (e.g., metric changes in aspect ratio, degree of curvilinearity, departure from parallelism, angle, and line lengths), while keeping parallelism, collinearity, cotermination, and other nonaccidental properties intact (see Fig. 1). Affine transformations (changes in size, planar rotation, shear, and translation) of a geon image also induce changes in metric properties, while keeping nonaccidental properties intact, thus reproducing the same repetition advantage that these properties enjoy during experience with multiple views (Fig. 1). Affine transformations can be applied to a single view of a geon, thereby permitting a critical test of view-invariance learning after training with only one object view.

The prediction that affine transformations of a single object image can foster rotational invariance is quite surprising. The reason is because this prediction argues against the proposal, put forward on computational grounds, that there is a fundamental difference between invariance across affine and non-affine transformations (Riesenhuber and Poggio, 2000). The effects of affine transformations of an image can be estimated from a single object view, which means that it should be possible to show invariance to all affine transformations of an image without the need to collect more than one example in the set. On the other hand, the effects of non-affine transformations, such as rotation in depth, cannot be estimated from a single object view, thereby leading researchers to propose that experience with different object views is necessary to achieve invariant recognition across non-affine transformations. Thus, the prediction that experience with affine transformations of a single image can foster learning of rotational invariance is highly unexpected and goes against the view that the "distinction between types of invariance is more fundamental than the distinction between categorization and recognition" (Riesenhuber and Poggio, 2000).

Furthermore, our prediction was motivated by a theory first developed to explain object categorization in birds using simple associative learning processes, thought to be shared across vertebrates (Soto and Wasserman, 2010a, 2012a, 2014). This theory thus makes the additional striking prediction that the same effect of experience with affine transformations should be observed in even distantly-related vertebrate species, such as people and pigeons.

In the present study, pigeons (Experiment 1) and people (Experiment 2) were each randomly assigned to two groups. In each control group, subjects were trained to discriminate a single view of each of four geons. In both affine transformations groups, subjects were exposed to the same single view of each of four geons, in its original form and after several affine transformations. All of these stimuli were carefully created so that image similarities could not explain the predicted results. After training, all of the groups were tested with novel views of the objects, in order to assess the extent to which the experimental manipulation affected view-invariant recognition. Methods were kept as similar as possible for the two species; however, correct response time was used as the measure of performance for people, whereas proportion of correct responses was used for pigeons.

2. Pigeon experiment

2.1. Materials and methods

2.1.1. Subjects

Subjects were eight pigeons (*Columba livia*) kept at 85% of their free-feeding weights. The birds had previously participated in unrelated research.

2.1.2. Stimuli

The stimuli were obtained from four geons (arch, barrel, horn, and wedge) rendered over a white background. Three-dimensional models were created using Blender 2.49 (The Blender Foundation) and rotated in depth by 30°-intervals, ±10° to avoid accidental views of the objects (Biederman, 1987), along their *x*-axis to yield a total of 12 views. The final stimuli were 7.4×7.4 cm in size. One view was designated the 0° training view for each geon. This was the only view ever seen during training by pigeons in the experiment; all other views were only presented during testing.

In the control group, the 0° training views were the only images shown to each pigeon. In the affine transformations group, additional training stimuli resulted from the application of affine transformations to these 0° training views. The set of 27 stimuli for each object (108 images in total) was obtained by combining three levels of size, planar rotation, and shear ($3 \times 3 \times 3 = 27$ combinations).

To select the magnitudes of all of the affine transformations of the training view, it was necessary to ensure that better performance with the testing views in the affine transformations group could not be explained as the result of low-level image similarities between the training and testing stimuli.

[1] In the context of the present work, geons refer simply to a specific kind of three-dimensional object. Specifically, geons are volumes built by swiping a cross-section through a main axis according to a bevel function (see Biederman, 1987). These objects happen to have properties that are useful for the goals of this study (i.e., they contain nonaccidental properties). The controversial issue of whether or not geons are represented by people or nonhuman animals during object recognition is not addressed by the present study (for a review of this work, see Wasserman and Biederman, 2012).