Behavioural Processes xxx (2015) xxx–xxx

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 xxx

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/

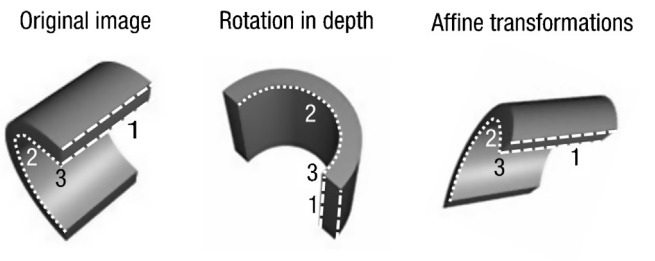

Fig. 1. Effects of rotation in depth and three affine transformations (changes in size, planar rotation, and shear) on different properties of geon images. Both transformation types change metric properties, such as the degree of edge curvilinearity (2) and the angular degree of a co-termination (3). 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).

Please cite this article in press as: Soto, F.A., Wasserman, E.A., Promoting rotational-invariance in object recognition despite experience with only a single view. Behav. Process. (2015), http://dx.doi.org/10.1016/j.beproc.2015.11.005

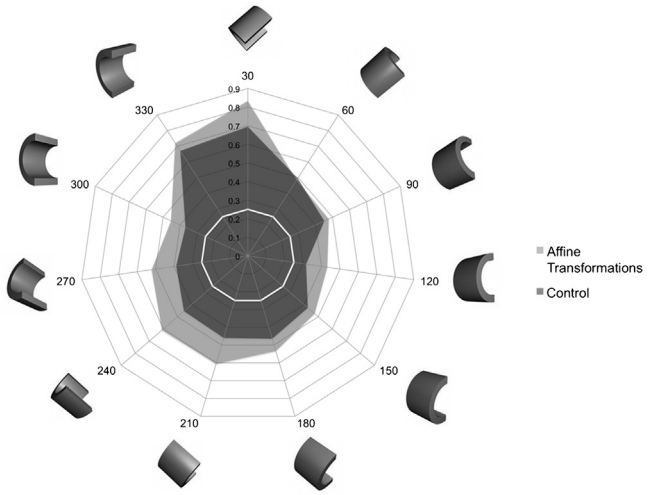

Fig. 2. Results of the generalization test with pigeons. Plotted in the polar spikes (labels represent degrees of rotation) is the mean proportion of correct responses for different novel views of the training objects. The white line represents chance performance (.25 correct responses).

Two measures of image similarity were calculated, each widely used in earlier research (Kayaert et al., 2003; Vogels et al., 2001; Wang et al., 2005; Yamashita et al., 2010). The first measure was the pixel-by-pixel dissimilarity in luminosity between images with adjustment for position, which controls for image similarities as they are present in the retinal input (Vogels et al., 2001). For a pair of images, the position of the object within one of the displays was shifted 441 times, resulting from the combination of 21 horizontal and 21 vertical image translations in 5-pixel steps; each time, the Euclidean distance between the pixel luminosity values of the images was calculated. The smallest of the distances was kept as the pixel dissimilarity measure.

To obtain the second measure, the wavelet transform of images was computed using a biorthogonal (6.8) wavelet and the Euclidean distance between the wavelet coefficients of image pairs was taken as an index of their dissimilarity. This measure controls for image similarities as they are represented in the earlier stages of shape processing in biological visual systems (Vogels et al., 2001), which are thought to implement a wavelet-like decomposition of images (Field, 1999; Stevens, 2004).

Stimuli were selected in six steps: first, images were created using a candidate set of affine transformations. Second, the two similarity measures were computed for all pairs of training-testing images. Third, each testing image was classified as generated by a particular object according to two criteria: (a) the object with the lowest dissimilarity between a single training image and the testing image or (b) the object with the lowest mean dissimilarity between all training images and the testing image. Fourth, the classification accuracies obtained using these two classification methods and the two similarity measures were computed. Fifth, similar accuracies were computed using only the non-transformed training image to estimate similarity-based performance in the control condition. Sixth, if accuracies obtained using the affinely transformed images were lower than those obtained using the non-transformed image alone, then the transformations were chosen. So, transformations were used only if their classification by similarity led to the *opposite* prediction from that tested by our experiment.

2.1.3. Apparatus

Four 36- × 36- × 41-cm operant chambers (Gibson et al., 2004) were used. The stimuli were presented on a 15-in. LCD monitor located behind an AccuTouch® resistive touchscreen (Elo TouchSystems, Fremont, CA) covered by mylar. Rewards were 45-mg food pellets dispensed into a cup on the rear wall of the chamber. Each chamber was controlled by an Apple® iMac® computer.

The experimental procedure was programmed using Matlab 7.9 (MathWorks, Natick, MA) with the Psychophysics Toolbox extensions (Brainard, 1997).

2.1.4. Procedure

Pigeons were randomly assigned to control and affine transformations groups. All stimuli were shown on a 7.5- × 7.5-cm display screen positioned in the middle of a computer monitor.

A trial began by presenting a white square with a black cross in the middle. A single peck in the square led to the presentation of the trial stimulus. The pigeon had to peck the screen 5–45 times before four response keys (black-and-white square icons) were displayed, one next to each corner of the center display screen. The pigeon had to identify the object presented by pecking its associated key. A correct choice led to 1–3 food pellets and an intertrial interval of 5 s. An incorrect choice led to a timeout of 5–30 s and to the repetition of the trial. Data from the first instance of each trial were analyzed.

During training, daily sessions consisted of a single block of 108 randomly ordered trials. Each of the 108 stimuli was presented once in the affine transformations group, and each of the four stimuli was presented 27 times in the control group. Training sessions continued until pigeons achieved accuracies of .85 correct responses overall and .80 correct responses for each object; then, pigeons were given 20 testing sessions.

Testing sessions started with four presentations of each unmodified training geon, followed by a testing block consisting of the presentation of each geon from 11 novel viewpoints, randomly intermixed within a block of training trials. All responses were reinforced on testing trials, thus eliminating the need for correction trials.

2.2. Results and discussion

It took the birds in group affine transformations between 16 and 102 training days ($M = 44.25$) to reach criterion. It took the birds in the control group between 21 and 53 training days ($M = 38.5$) to reach criterion. The difference was not significant according to an independent samples t-test, $t(6) = .28$, $p > .5$.

Fig. 2 is a polar plot showing the mean proportion of correct responses to the testing stimuli in both groups. Data from the affine transformation group are plotted using a lighter shade of gray than

Please cite this article in press as: Soto, F.A., Wasserman, E.A., Promoting rotational-invariance in object recognition despite experience with only a single view. Behav. Process. (2015), http://dx.doi.org/10.1016/j.beproc.2015.11.005

data from the control group. The white line in the center of the plot represents chance performance (.25 correct responses). As predicted, pigeons showed a mean proportion of correct responses that was higher in the affine transformations group ($M = .56$; $SE = .03$; $n = 4$) than in the control group ($M = .47$; $SE = .02$; $n = 4$). The testing data were analyzed with a 2 (Group) × 11 (Rotation) Analysis of Covariance (ANCOVA), using average performance with training stimuli as the covariate to control for the possibility that differences in performance with novel views might be due to disparities in performance to the training images. The ANCOVA revealed significant main effects of Group, $F(1, 5) = 11.76$, $p = .0186$, $d_{unbiased} = 2.03$, and Rotation, $F(10, 60) = 12.57$, $p = 1.982 \times 10^{-11}$, $\eta^2 = .68$, but no other effects were significant.

In summary, as predicted by our theory of object recognition and categorization in pigeons (Soto and Wasserman, 2010a, 2012a, 2014), the results of this experiment show that that experience with affine transformations of geons shown from a single view demonstrably enhances the recognition of those geons from novel viewpoints. Because all of the cognitive mechanisms proposed in our theory are thought to be present across all amniote vertebrates (see Soto and Wasserman, 2014), an additional prediction is that the same effect demonstrated here for pigeons should be present in humans as well. The following experiment tested this prediction.

3. Human experiment

3.1. Materials and methods

3.1.1. Participants

Thirty-eight undergraduates from the University of California, Santa Barbara, participated in exchange for course credit.

3.1.2. Stimuli and apparatus

Stimuli were the same as those used in Experiment 1. People were tested with five Apple® eMac® computers. The experiment was programmed using Matlab 7.9 (MathWorks, Natick, MA) with the Psychophysics Toolbox extensions (Brainard, 1997). Images were displayed on the monitor using the same parameters as with pigeons. Responses were recorded via keyboard. Auditory feedback was provided through headphones.

3.1.3. Procedure

At the start of the experiment, participants were given instructions indicating that they would have to identify three-dimensional shapes by pressing one of four keys. Participants were told that the same shapes would be shown later from new viewpoints, that no feedback would be provided during those trials, and that they should keep responding as accurately and as quickly as they could.

A trial began by presenting a white square with a black cross in the center of the screen for 500 ms. The stimulus was then presented for 40 ms, followed by a mask for 200 ms, randomly chosen from a set of 20 masks consisting of 16 image segments randomly selected from all of the training images. The short presentation time and masks were included to increase task difficulty and to enhance the sensitivity of the experiment. The experiment was based on the assumption that a robust decrease in response time in the recognition of novel views of a geon would be observed, as in previous research (Tarr et al., 1998). Pilot studies suggested that such an effect might be small and insufficiently robust to corroborate our main hypothesis, thereby prompting us to increase task difficulty.

Participants responded by pressing the numerals 1, 2, 3, or 4 on their keyboard. After a key press or a maximum period of 2000 ms, visual and auditory feedback was provided for 500 ms. On correct trials, "correct" was displayed in the middle of the screen, together with a pleasant chime presented through headphones. On incorrect trials, "incorrect" was displayed in the middle of the screen, together with an unpleasant buzzer presented through headphones.

The experiment was completed in a single 50-min session divided into three training blocks and one testing block, all having the same structure as blocks in the pigeon experiment. No feedback was provided during testing trials and participants were given unlimited time to respond. Only trials in which response times were lower than 2000 ms were included in the reported analysis to eliminate the influence of outliers and to have all of the data (from training and testing trials) fall within the same range. The same results were obtained when response times longer than 2000 were included.

3.2. Results and discussion

Fig. 3 is a polar plot displaying the mean correct response times obtained during testing trials of the present experiment. As before, data from the affine transformation group are plotted using a lighter shade of gray than data from the control group. Note that the scale in this plot has been inverted, with slower response times toward the center of the figure, to make the results more easily comparable to those obtained from pigeons and displayed in Fig. 2. In both cases, a higher score represents better performance (more accurate or faster).

People showed mean correct response times that were faster in the affine transformations group ($M = 621.16$; $SE = 15.82$; $n = 19$) than in the control group ($M = 688.83$; $SE = 17.22$; $n = 19$). A 2 (Group) × 11 (Rotation) ANCOVA, using mean correct response times with training stimuli during testing as the covariate, revealed a significant main effect of Group, $F(1, 35) = 8.96$, $p = .005$, $d_{unbiased} = .95$. No other effects were significant.

4. Scrambled control

One possible explanation for the human results reported in the previous section is that, in the control group, the surprising presentation of transformed stimuli for the first time during the test slowed people's response times. Such slower response times would not be observed in the affine transformations group, due to the familiarity of this group with transformed versions of the training images.

Note that the novelty of transformed testing images is unlikely to affect the results from the pigeon experiment, for two reasons. First, although the surprising presentation of a novel stimulus might affect response times, it seems less likely to affect the proportion of correct responses, particularly in an easy discrimination like the one given to the control group in the previous experiments. Second, unlike people, who were tested in a single block of trials lasting only minutes, pigeons were tested in several separate daily sessions. Any surprising effect of novel stimuli is likely to vanish very quickly with the procedure used for pigeons.

To test the hypothesis that training with any transformation would produce results like those observed in the affine transformations group from the previous experiment, an additional control group was trained with transformations of the training view in which image fragments were spatially scrambled. Examples of these scrambled stimuli are shown in Fig. 4. Note how the level of scrambling in the transformed images still leaves the object easily recognizable.

Training with scrambled versions of the geon images should familiarize the participants with the presentation of several transformed versions of such images, making the testing procedure unsurprising. However, the scrambling transformation does not have the properties of rotation and affine transformations highlighted in Fig. 1. That is, scrambling does not selectively preserve

Please cite this article in press as: Soto, F.A., Wasserman, E.A., Promoting rotational-invariance in object recognition despite experience with only a single view. Behav. Process. (2015), http://dx.doi.org/10.1016/j.beproc.2015.11.005

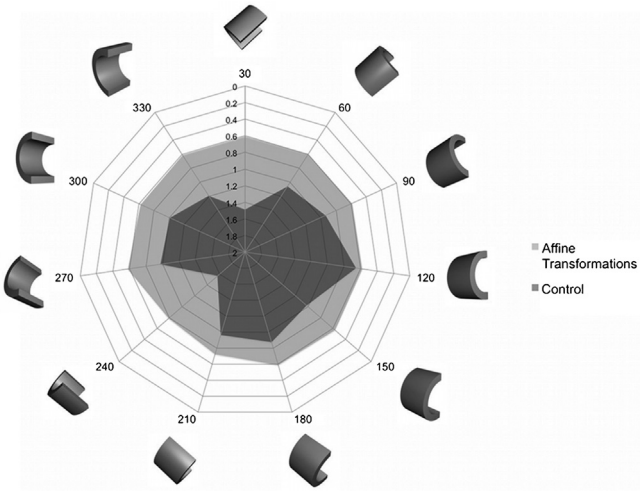

Fig. 3. Results of the generalization test with people. Plotted in the polar spikes (labels represent degrees of rotation) is the mean correct response times (second) for different novel views of the training objects. Note that the scale has been inverted, with slower response times toward the center of the figure, to make the results more easily comparable to those from Experiment 1 (see Fig. 2).

nonaccidental properties while disrupting metric properties in the geon images. Instead, scrambling simply disrupts all image properties at the points in which image segments are replaced. For this reason, we predicted that the scrambling control from the present experiment would show lower rotational invariance than the affine transformations group from the previous experiment.

4.1. Materials and methods

4.1.1. Participants

Thirteen people with the same characteristics as those in Experiment 2 were studied in the scrambled control.

4.1.2. Stimuli and apparatus

Twenty-six scrambled versions of each of the four training images used in the control group of the previous experiments were created (for examples, see Fig. 4). The area containing the object in each image was divided into 16 squares of equal size (i.e., in a 4×4 grid). The location of six of these squares was randomized to create each scrambled training stimulus.

4.1.3. Procedure

The training and testing procedure was the same as in the affine transformations group from Experiment 2. The only difference was that the 26 affinely transformed versions of the training images were replaced by 26 scrambled versions.

4.2. Results and discussion

The mean correct reaction times were slower in the scrambled control group ($M = 701.45$; $SE = 38.46$; $n = 13$) than in the affine group from the previous experiment ($M = 621.16$; $SE = 15.82$; $n = 19$). A 2 (Group) $\times$ 11 (Rotation) ANCOVA, using mean correct reaction time with the training stimuli as the covariate, revealed a significant main effect of Group, $F(1, 29) = 4.45$, $p = .0436$, $d_{unbiased} = .74$. No other effects were significant. These results suggest that the main effect of Group in human participants shown in Fig. 3 was not due to slower response times in the control group because of the surprising presentation of transformed stimuli for the first time during testing.

5. General discussion

The unprecedented finding of the present study was that experience with affine transformations of objects shown at a single view enhances the recognition of those objects at novel viewpoints by both pigeons and people. Critically, these results are difficult to explain in terms of generalization from training views based on image similarities, as the stimuli were expressly created in order to exclude this explanation.

Our results suggest that the distinction between affine and non-affine transformations may not be fundamental for biological visual systems, as has been assumed on computational grounds (Riesenhuber and Poggio, 2000). Quite the contrary, the effects of experience with both types of transformations seem to be interchangeable, at least under some circumstances.

Our findings accord with both psychophysical and neurobiological studies suggesting that the mechanisms of invariant object recognition might be similar across affine and non-affine transformations. Behaviorally, object recognition in primates and birds can show variable levels of tolerance to different types of transformations, but object recognition is not completely invariant across either affine or non-affine transformations (Kravitz et al., 2008; Peissig et al., 2006, 2000; Tarr et al., 1998). In primates, neurons in late stages of the ventral stream show variable levels of invariance to any transformation (Booth and Rolls, 1998; Op de Beeck and Vogels, 2000; Zoccolan et al., 2007). Although a population of neurons with such variable levels of invariance can provide a basis for invariant object recognition at later processing stages (Hung et al., 2005; Li et al., 2009), it is unlikely that behavioral invariance can be achieved without explicit training with variations in irrelevant object dimensions (Goris and Op de Beeck, 2010).

Most recent research into the mechanisms of invariance learning has been motivated by the hypothesis that the primate brain learns invariant representations in an unsupervised manner, from experience with different retinal images which are 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). The current results, as well as previous behavioral results with monkeys (Wang et al., 2005; Yamashita et al., 2010), are difficult to explain by such an unsupervised learning process, because they involve view-invariance learning without experience with temporally contiguous object views. Instead, the results suggest a role for a learning mechanism, in which common

Please cite this article in press as: Soto, F.A., Wasserman, E.A., Promoting rotational-invariance in object recognition despite experience with only a single view. Behav. Process. (2015), http://dx.doi.org/10.1016/j.beproc.2015.11.005

Fig. 4. Examples of the stimuli presented to participants in the control experiment. The leftmost image is the original training object, and the rest are scrambled versions of it.

properties of a number of experienced images are associated with an object representation or common response because they reliably predict object identity (Soto et al., 2012; Wang et al., 2005; Yamashita et al., 2010).

An important aspect of our results is that they are general across distantly related vertebrate species. There were important methodological differences between the pigeon and human experiments (e.g., people but not pigeons were provided with instructions, pigeons were trained for weeks but people were trained only for minutes) and the response variables were not the same in the two species (accuracy in pigeons and response times in people), so it could be argued that the experiments were incomparable.

However, methodological differences are common in all comparative studies involving such distantly related species. For example, training in most human experiments takes in the scale of minutes to hours, whereas training in pigeon experiments takes in the scale of weeks to months. If anything, such methodological differences should be the source of differences in the results across species; yet, we still found in both species an increase in generalization to novel views after training with affine transformations.

Regarding the use of different performance measures in both species, practically all previous relevant research regarding view invariance in object recognition have used proportion of correct responses as the performance measure in pigeons and response times as the performance measure in people. Thus, our selection of response measures permits linking the present results with the previous literature in both pigeons and people.

The generality of our results across species despite methodological differences suggests that, although the brain systems involved in shape perception in the avian and primate brains are not homologous (Shimizu and Bowers, 1999), under at least some circumstances, both species might use similar strategies to achieve invariant object recognition (Soto and Wasserman, 2012a,b, 2014). The learning mechanism suggested by the current results may be of such importance for object recognition that it either has been conserved across millions of years of independent evolution or it has evolved independently on at least two separate occasions.

Acknowledgements

We thank Cathleen Moore, Shaun Vecera, Bob McMurray, and Matthew Rizzo for their feedback on this study. This research was supported by National Institute of Mental Health Grant MH47313 to EAW, by National Eye Institute Grant EY019781 to EAW, and by Sigma Xi Grant-in-Aid of Research Grant G20101015155129 to FAS.

References

Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. Psychol. Rev. 94 (2), 115–117.
Booth, M.C., Rolls, E.T., 1998. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb. Cortex 8 (6), 510–523.
Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10 (4), 433–436.
Cox, D.D., Meier, P., Oertelt, N., DiCarlo, J.J., 2005. Breaking position-invariant object recognition. Nat. Neurosci. 8 (9), 1145–1147.
Field, D.J., 1999. Wavelets, vision and the statistics of natural scenes. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 357 (1760), 2527–2542.
Földiák, P., 1990. Forming sparse representations by local anti-Hebbian learning. Biol. Cybern. 64 (2), 165–170.
Gibson, B.M., Wasserman, E.A., Frei, L., Miller, K., 2004. Recent advances in operant conditioning technology: a versatile and affordable computerized touchscreen system. Behav. Res. Methods Instrum. Comput. 36 (2), 355–362.
Gibson, B.M., Lazareva, O.F., Gosselin, F., Schyns, P.G., Wasserman, E.A., 2007. Nonaccidental properties underlie shape recognition in mammalian and nonmammalian vision. Curr. Biol. 17 (4), 336–340.
Goris, R.L.T., Op de Beeck, H.P., 2010. Invariance in visual object recognition requires training: a computational argument. Front. Neurosci. 4 (1), 71–78.
Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J., 2005. Fast readout of object identity from macaque inferior temporal cortex. Science 310 (5749), 863–866.
Kayaert, G., Biederman, I., Vogels, R., 2003. Shape tuning in macaque inferior temporal cortex. J. Neurosci. 23 (7), 3016.
Kravitz, D.J., Vinson, L.D., Baker, C.I., 2008. How position dependent is visual object recognition? Trends Cogn. Sci. 12 (3), 114–122.
Lazareva, O.F., Wasserman, E.A., Biederman, I., 2008. Pigeons and humans are more sensitive to nonaccidental than to metric changes in visual objects. Behav. Processes 77 (2), 199–209.
Li, N., Cox, D.D., Zoccolan, D., DiCarlo, J.J., 2009. What response properties do individual neurons need to underlie position and clutter invariant object recognition? J. Neurophysiol. 102 (1), 360–376.
Li, N., DiCarlo, J.J., 2008. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. Science 321 (5895), 1502–1507.
Op de Beeck, H.P., Vogels, R., 2000. Spatial sensitivity of macaque inferior temporal neurons. J. Comp. Neurol. 426 (4), 505–518.
Peissig, J.J., Kirkpatrick, K., Young, M.E., Wasserman, E.A., Biederman, I., 2006. Effects of varying stimulus size on object recognition in pigeons. J. Exp. Psychol. Anim. Behav. Process. 32 (4), 419–430.
Peissig, J.J., Young, M.E., Wasserman, E.A., Biederman, I., 2000. Seeing things from a different angle: the pigeon's recognition of single geons rotated in depth. J. Exp. Psychol. Anim. Behav. Process. 26 (2), 115–132.
Riesenhuber, M., Poggio, T., 2000. Models of object recognition. Nat. Neurosci. 3, 1199–1204.
Shimizu, T., Bowers, A.N., 1999. Visual circuits of the avian telencephalon: evolutionary implications. Behav. Brain Res. 98 (2), 183–191.
Stevens, C.F., 2004. Preserving properties of object shape by computations in primary visual cortex. Proc. Natl. Acad. Sci. 101 (43), 15524–15529.
Soto, F.A., Siow, J.Y.M., Wasserman, E.A., 2012. View-invariance learning in object recognition by pigeons depends on error-driven associative learning processes. Vision Res. 62, 148–161.
Soto, F.A., Wasserman, E.A., 2010a. Error-driven learning in visual categorization and object recognition: a common elements model. Psychol. Rev. 117 (2), 349–381.
Soto, F.A., Wasserman, E.A., 2010b. Missing the forest for the trees: object discrimination learning blocks categorization learning. Psychol. Sci. 21 (10), 1510–1517.
Soto, F.A., Wasserman, E.A., 2012a. Visual object categorization in birds and primates: integrating behavioral, neurobiological, and computational evidence within a general process framework. Cogn. Affect. Behav. Neurosci. 12 (1), 220–240.
Soto, F.A., Wasserman, E.A., 2012b. A category-overshadowing effect in pigeons: support for the common elements model of object categorization learning. J. Exp. Psychol. Anim. Behav. Process. 38 (3), 322–328.
Soto, F.A., Wasserman, E.A., 2014. Mechanisms of object recognition: what we have learned from pigeons. Front. Neural Circuits 8 (22).
Spetch, M.L., Friedman, A., Reid, S.L., 2001. The effect of distinctive parts on recognition of depth-rotated objects by pigeons (*Columba livia*) and humans. J. Exp. Psychol. Gen. 130, 238–255.
Stringer, S.M., Perry, G., Rolls, E.T., Proske, J.H., 2006. Learning invariant object recognition in the visual system with continuous transformations. Biol. Cybern. 94 (2), 128–142.
Tarr, M.J., Williams, P., Hayward, W.G., Gauthier, I., 1998. Three-dimensional object recognition is viewpoint dependent. Nat. Neurosci. 1 (4), 275–277.
Vogels, R., Biederman, I., Bar, M., Lorincz, A., 2001. Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. J. Cogn. Neurosci. 13 (4), 444–453.
Wallis, G., Bülthoff, H., 2001. Effects of temporal association on recognition memory. Proc. Natl. Acad. Sci. 98 (8), 4800–4804.
Wang, G., Obama, S., Yamashita, W., Sugihara, T., Tanaka, K., 2005. Prior experience of rotation is not required for recognizing objects seen from different angles. Nat. Neurosci. 8 (12), 1768–1775.
Wasserman, E.A., Biederman, I., 2012. Recognition by components: a bird's eye view. In: Lazareva, O.F., Shimizu, T., Wasserman, E.A. (Eds.), How Animals See the World. Oxford University Press, New York.

Please cite this article in press as: Soto, F.A., Wasserman, E.A., Promoting rotational-invariance in object recognition despite experience with only a single view. Behav. Process. (2015), http://dx.doi.org/10.1016/j.beproc.2015.11.005

Wasserman, E.A., Gagliardi, J.L., Cook, B.R., Kirkpatrick-Steger, K., Astley, S.L., Biederman, I., 1996. The pigeon's recognition of drawings of depth-rotated stimuli. J. Exp. Psychol. Anim. Behav. Process. 22 (2), 205–221.
Wiskott, L., Sejnowski, T.J., 2002. Slow feature analysis: unsupervised learning of invariances. Neural Comput. 14 (4), 715–770.
Yamashita, W., Wang, G., Tanaka, K., 2010. View-invariant object recognition ability develops after discrimination, not mere exposure, at several viewing angles. Eur. J. Neurosci. 31 (2), 327–335.
Zoccolan, D., Kouh, M., Poggio, T., DiCarlo, J.J., 2007. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. J. Neurosci. 27 (45), 12292.

Please cite this article in press as: Soto, F.A., Wasserman, E.A., Promoting rotational-invariance in object recognition despite experience with only a single view. Behav. Process. (2015), http://dx.doi.org/10.1016/j.beproc.2015.11.005