

Visual object categorization in birds and primates: Integrating behavioral, neurobiological, and computational evidence within a “general process” framework

Fabian A. Soto · Edward A. Wasserman

© Psychonomic Society, Inc. 2011

Abstract Previous comparative work has suggested that the mechanisms of object categorization differ importantly for birds and primates. However, behavioral and neurobiological differences do not preclude the possibility that at least some of those mechanisms are shared across these evolutionarily distant groups. The present study integrates behavioral, neurobiological, and computational evidence concerning the “general processes” that are involved in object recognition in vertebrates. We start by reviewing work implicating error-driven learning in object categorization by birds and primates, and also consider neurobiological evidence suggesting that the basal ganglia might implement this process. We then turn to work with a computational model showing that principles of visual processing discovered in the primate brain can account for key behavioral findings in object recognition by pigeons, including cases in which pigeons’ behavior differs from that of people. These results provide a proof of concept that the basic principles of visual shape processing are similar across distantly related vertebrate species, thereby offering important insights into the evolution of visual cognition.

Electronic supplementary material The online version of this article (doi:10.3758/s13415-011-0070-x) contains supplementary material, which is available to authorized users.

F. A. Soto (✉)
Department of Psychological and Brain Sciences,
University of California Santa Barbara,
Santa Barbara, CA 93106, USA
e-mail: fabian.soto@psych.ucsb.edu

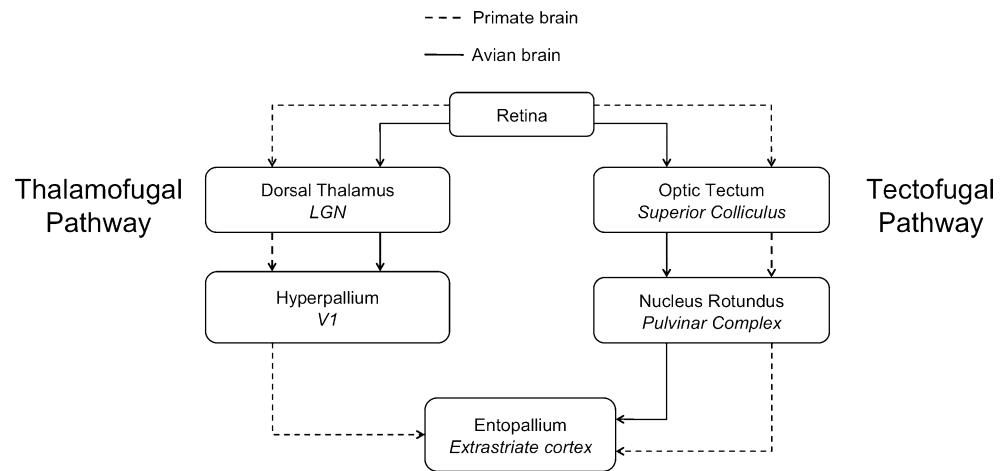
E. A. Wasserman
Department of Psychology, University of Iowa,
Iowa City, IA, USA

Keywords Object categorization · Object recognition · Error-driven learning · Hierarchical model · Feedforward processing · Comparative cognition · Avian vision · Pigeon · Computational model · Animal models · Visual cortex

Many species must visually recognize and categorize objects to successfully adapt to their environments. Considerable comparative research has been conducted in object recognition, especially involving pigeons and people, whose visual systems have independently evolved from a common ancestor, from which their lineages diverged more than 300 million years ago. The results of behavioral studies have sometimes disclosed striking similarities between these species, and at other times have disclosed notable disparities, especially pointing toward a lower ability of pigeons to recognize transformed versions of familiar objects (for reviews, see Kirkpatrick, 2001; Spetch & Friedman, 2006).

Similarly, the results of neurobiological studies have revealed both similarities and disparities in the structures that underlie visual object processing. The overall organization of the two visual systems is quite similar, with the most notable shared feature being their subdivision into parallel pathways. All amniotes (mammals, birds, and reptiles) have two main visual pathways from retina to telencephalon: the thalamofugal and tectofugal pathways (see Fig. 1; Shimizu & Bowers, 1999; Wylie, Gutierrez-Ibanez, Pakan, & Iwaniuk, 2009). Furthermore, the avian tectofugal pathway and its pallial targets seem to be separated into at least two parallel subdivisions, one specialized for processing motion and the other specialized for processing shape (Fredes, Tapia, Letelier, Marín, & Mpodozis, 2010; Laverghetta & Shimizu, 2003; Nguyen et al., 2004; Shimizu & Bowers, 1999; Wang, Jiang, & Frost, 1993)—similar to the organization of the

Fig. 1 Diagram of the two main visual pathways from the retina to the extrastriate cortex in primates (italicized letters) and from the retina to the entopallium in birds (nonitalicized letters)



primate thalamofugal system and its cortical targets (Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Haxby, 1994).

On the other hand, in birds, the tectofugal pathway is more developed and plays a dominant role in visual discrimination tasks, whereas in primates, the thalamofugal pathway plays the dominant part (Shimizu & Bowers, 1999; Wylie et al., 2009). Thus, the visual pathways that are responsible for object recognition and categorization are not homologous in primates and birds. As well, the optic tectum and other midbrain structures play a central role in visual processing in birds, but not in mammals.

How should these similarities and disparities be interpreted? Is it possible that similar cognitive and perceptual processes underlie object recognition by birds and primates, regardless of these behavioral and neuroanatomical disparities? Given the disparities, can studies in object recognition by birds provide any insights into human visual cognition? The present study integrates behavioral, neurobiological, and computational evidence within an evolutionary framework to offer answers to these three questions. We propose that differences between birds and primates should not be taken as evidence of completely specialized visual systems, that these two taxa likely share common processes of object recognition, and that birds can and should be used by the cognitive neuroscience community as an animal model for the study of *some* of the processes that are involved in object recognition by people.

To support these three claims, we begin by presenting an overview of contemporary thinking in comparative cognition, proposing that neither behavioral nor neuroanatomical differences between birds and primates should stop us from using these different animals to pursue the study of general processes of object categorization. Then, we review evidence that error-driven associative learning is involved in object categorization learning by birds and primates. Finally, we present new computational evidence that principles of computation discovered in the primate visual

system can explain many aspects of avian object perception, including cases in which pigeons' behavior differs from people's behavior, thereby suggesting that processes other than categorization learning and visual shape processing might underlie these disparities.

A comparative approach to the study of object categorization

According to a comparative approach, cognition, like any other biological process, should be studied and understood within the context of evolutionary theory (Shettleworth, 2010b). Two different perspectives have arisen in comparative cognition as the consequence of such evolutionary thinking: one focusing on *general processes*, and a second focusing on *adaptive specializations* (Riley & Langley, 1993).

According to the general-processes approach (Bitterman, 2000; Macphail & Bolhuis, 2001; Wasserman, 1993), some principles of cognitive processing are both widely distributed across species and highly useful for the solution of various environmental tasks. The main evolutionary argument for this proposal is that even evolutionary diversity implies shared ancestry and that some selective pressures are spread so widely across environments that they affect the selection of characters above the species level (see Papini, 2002).

According to the specialized-adaptations approach (Gallistel, 1999; Shettleworth, 1993, 2000), many of the cognitive processes underlying adaptive behavior are better understood as species-specific mechanisms. The main evolutionary argument for this proposal is that there is so much variability in the ways in which information must be used by different species to adapt to their different environments that it is impossible for general processing mechanisms to solve all of these environmental tasks (Shettleworth, 2000).

Note that, according to these definitions, the issue of whether a process is relatively general or specialized has to

do with how widespread the process is across species and with the kind of computational problem that the process might solve. General processes are relatively widespread across species; they are likely to solve computational problems that are shared by the environments to which such species have adapted. Adaptive specializations are relatively specific to one or a few species; they are likely to solve computational problems that are idiosyncratic to the particular environments to which such species have adapted. Importantly, the issue of generality and adaptive specialization, as it is defined here, is different from the issue of whether the mind is composed of domain-general or domain-specific mechanisms (e.g., Atkinson & Wheeler, 2004; Cosmides & Tooby, 1994).

Currently, it is accepted that the study of cognition from an evolutionary standpoint should encompass both general processes and adaptive specializations (Papini, 2002; Riley & Langley, 1993; Shettleworth, 2010b). The complex behavior and cognition that animals exhibit in their natural environments can be broken down into more basic processes, with some of them being quite general and others quite specific. Even if a complex behavior is unique to a single species, the subprocesses that underlie this behavior may be more widespread (Shettleworth, 2000). Thus, the study of complex cognition should be carried out within an atomistic and “bottom-up” approach (de Waal & Ferrari, 2010; Shettleworth, 2010b) focused on identifying and investigating the basic mechanisms that underlie complex cognition and how they may have evolved.

What this consideration means for interpreting studies of object recognition in birds is that this form of cognition is likely to arise from a number of processes, including mechanisms of visual shape extraction, attention, associative learning, perceptual learning, and decision making. Adaptive specializations in any of these systems could be responsible for behavioral differences between species, but it is likely that many other processes are shared, especially if they provide solutions to problems that are general to many visual environments. As for the neuroanatomical differences between birds and primates, two points must be considered. First, many processes other than visual perception are involved in high-level visual cognition. For example, object categorization might also involve the selection of those outputs from the visual system that are useful for the recognition of a class of objects. In the following section, we will propose that a common mechanism, implemented in homologous structures, carries out this selection process in birds and primates.

Second, even when the structures underlying object perception are not homologous in people and birds, it is still possible that they implement similar processes. If what defines a general process is that it is shared across many species and that it solves a common environmental

problem, then any nondivergent form of evolutionary change can lead to the evolution of a general process. In contrast, adaptive specializations can be produced only by divergent evolution. Avian and primate visual systems have evolved from a common ancestor, through exposure to a similar pressure to solve the computational problems that are posed by object recognition (see Rust & Stocker, 2010). Thus, homoplasies may have evolved through either parallel or convergent evolution, for example (see Papini, 2002). If similar mechanisms have evolved for object recognition in birds and primates, regardless of whether they are due to homology or homoplasy, then studying the neurobiological substrates of behavior in birds offers an opportunity to increase our understanding of the same processes in humans. A good example of the successful application of such an approach is using song learning in birds as an animal model of human language acquisition (for a recent review, see Bolhuis, Okanoya, & Scharff, 2010). Although most similarities between song learning and language acquisition are likely to be the result of convergent evolution, considerable insight has been gained about the latter by studying the former.

The neuroscience community has focused almost exclusively on the macaque monkey for studying the neuroscience of visual cognition, since this species has higher face validity as a good animal model for human vision than do birds. From the point of view of comparative cognition, however, birds are just as useful as nonhuman primates for the study of the general processes of object categorization. Only in the study of more specialized processes should nonhuman primates be considered a better animal model than birds. We next review evidence implicating the involvement of general processes in object recognition by primates and birds.

Object categorization and general learning processes

The psychological principles of associative learning have proven to be quite general (Bitterman, 2000; Macphail & Bolhuis, 2001; Siegel & Allan, 1996), regardless of the various tasks and species that have been used for their investigation. Studying object recognition from a general-processes view leads naturally to the question of whether associative-learning processes participate in this form of cognition.

Researchers generally agree that prediction error plays an important part in associative learning. We have recently proposed a model of object categorization that is based on error-driven associative learning (Soto & Wasserman, 2010a). The idea behind this model is simple: Each image is represented as a collection of “elements” that vary in their levels of specificity and invariance with respect to the

stimuli that they represent. Some elements tend to be activated by a single image, representing its *stimulus-specific* properties; other elements tend to be activated by several different images depicting objects from the same category, representing *category-specific* properties. These two kinds of elements are associated with responses in any categorization task depending on each of their abilities to predict reward via an error-driven learning rule, in which the change in the strength of the association between a stimulus element s_i and an action a_j , or Δv_{ij} , is determined by an equation similar to the following:

$$\Delta v_{ij} = s_i a_j \delta_j, \quad (1)$$

where δ_j represents the reward prediction error for action a_j , equal to the difference between the actual reward received and the prediction of reward estimated through the sum of the associative values of all of the active elements:

$$\delta_j = r - \sum_i s_i v_{ij}. \quad (2)$$

A consequence of making learning proportional to prediction error is that stimulus-specific and category-specific properties enter into a competition for learning; only those properties that are predictive of rewarded responses are selected to control performance.

The common-elements model can explain many key empirical results in the literature on object categorization by birds (as reviewed in Soto & Wasserman, 2010a). More importantly, the model generates precise predictions as to the conditions that should foster or hinder categorization learning. For example, the error-driven learning rule predicts that learning about stimulus-specific properties can block subsequent category learning. In agreement with this prediction, when pigeons (Soto & Wasserman, 2010a) and people (Soto & Wasserman, 2010b) are trained to solve a discrimination task by memorizing individual objects in photographs and their assigned responses, both species are

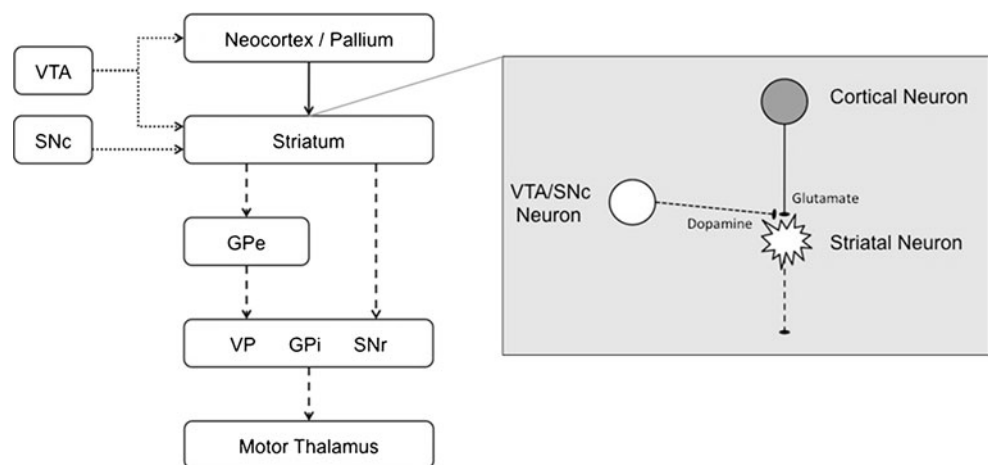
impaired in detecting a change in the training circumstances in which all of the presented objects are sorted according to their basic-level categories. This impairment arises because learning of the identification task promotes good performance and low prediction error in the subsequent categorization task.

What we now know about the neurobiology of associative learning accords with the idea that birds and people rely on an evolutionarily conserved learning mechanism to solve object categorization tasks. There is considerable evidence of error-driven learning in the basal ganglia, which are homologous structures in birds and mammals (Reiner, 2002; Reiner, Yamamoto, & Karten, 2005). A diagram of the main structures in the basal ganglia is shown on the left of Fig. 2.

The main input site to the basal ganglia is the striatum, which receives projections from most areas of the primate neocortex (Bolam, Hanley, Booth, & Bevan, 2000) and from most areas of the avian pallium (Veenman, Wild, & Reiner, 1995). Neurons in the striatum send projections to the output nuclei of the basal ganglia: ventral pallidum, the internal segment of globus pallidus, and substantia nigra pars reticula. These output nuclei influence motor control through descending projections to the motor thalamus.

The basal ganglia might learn, through error-driven learning, which responses are appropriate given a pattern of sensory stimulation in visual categorization tasks (see Ashby & Ennis, 2006; Seger, 2008; Shohamy, Myers, Kalanithi, & Gluck, 2008). Work by Schultz and colleagues (Montague, Dayan, & Sejnowski, 1996; Schultz, 1998, 2002; Schultz, Dayan, & Montague, 1997; Waelti, Dickinson, & Schultz, 2001) has shown that the activity of dopaminergic neurons in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) is well described through error-driven learning algorithms. The striatum is highly enriched in dopaminergic terminals arising from VTA/SNc (Durstewitz, Kröner, & Güntürkün, 1999; Nicola, Surmeier, & Malenka,

Fig. 2 Main structures of the basal ganglia (left) and a diagram of the cortical and nigral inputs to striatal neurons (right). A solid line represents an excitatory/glutamatergic projection, a dashed line represents an inhibitory/GABAergic projection, and a dotted line represents a modulatory/dopaminergic projection. VTA, ventral tegmental area; SNc, substantia nigra pars compacta; GPe, external globus pallidus; VP, ventral pallidum; GPi, internal globus pallidus; SNr, substantia nigra pars reticula



2000), and it has been shown that the plasticity of cortical–striatal synapses depends on the presence of dopamine in the synapse (Centonze, Picconi, Gubellini, Bernardi, & Calabresi, 2001; Reynolds & Wickens, 2002). These observations suggest that cortical–striatal synapses may mediate error-driven learning of stimulus–response associations (Schultz, 1998, 2002), such as those proposed by the three-term learning algorithm in our model (see Eq. 1). Object classification learning in the striatum would require activity of the presynaptic visual neurons (s_i), activity of the postsynaptic striatal neurons (a_j), and the presence of a dopaminergic signal coming from VTA/SNc (δ —see the subpanel in Fig. 2).

Most of the reviewed research has been done with mammals. But the avian basal ganglia may also play a role in visual discrimination learning (Izawa, Zachar, Aoki, Koga, & Matsushima, 2002) and in the prediction of reward from visual cues (Izawa, Aoki, & Matsushima, 2005; Yanagihara, Izawa, Koga, & Matsushima, 2001).

Although some researchers (Freedman, Riesenhuber, Poggio, & Miller, 2003; Jiang et al., 2007; Serre, Oliva, & Poggio, 2007) have suggested that the critical site for category learning in primates might be the prefrontal cortex (PFC), recent evidence has suggested that lateral PFC, the PFC area that has been more strongly related to categorization learning, is not necessary for learning and generalization in some object categorization tasks (Minamimoto, Saunders, & Richmond, 2010). The basal ganglia might be a more important site for categorization learning in primates than PFC, especially for learning processes that are shared with other species. An interesting possibility is that learning of abstract category representations in PFC is trained by the output of the basal ganglia (Antzoulatos & Miller, 2011).

In summary, evidence indicates that general principles of associative learning can explain object categorization learning in birds and people. We suggest that the basal ganglia may be a candidate site to implement such general processes. We now turn to the question of whether there might in addition be general visual processes involved in object categorization.

General visual processes? Hierarchical shape processing in the primate and avian brains

The ventral stream of the primate visual cortex is specialized for shape processing and object recognition (Mishkin et al., 1983; Ungerleider & Haxby, 1994). The ventral stream comprises several sequentially organized areas, starting at the primary visual cortex (V1), going to extrastriate areas V2 and V4, and then to the inferior temporal lobe (IT).

Several observations have indicated that the ventral pathway is hierarchically organized (Felleman & Van Essen, 1991; Grill-Spector & Malach, 2004). Areas along the pathway implement visual processing over information of increasing complexity, with each level integrating information from the previous one, discarding irrelevant information, and generating a more complex and abstract representation that is passed to the next level for further processing.

Hubel and Wiesel (1962, 1968) first proposed a hierarchical functional architecture to explain the neural response properties of cells in primary visual cortex. This model has received substantial empirical support over the years (Ferster & Miller, 2000; Martinez & Alonso, 2003; Reid & Usrey, 2004). Furthermore, it has been extended to explain the response properties of cells across the whole ventral pathway, including V2 (Anzai, Peng, & Van Essen, 2007; Boynton & Hegde, 2004), V4 (Cadieu et al., 2007), and IT neurons (Riesenhuber & Poggio, 1999). Consequently, a family of primate object recognition models has arisen (e.g., Fukushima, 1980; Perrett & Oram, 1993; Riesenhuber & Poggio, 1999, 2000; Rolls & Milward, 2000; Serre, 2006; Serre et al., 2005; Serre, Oliva, & Poggio, 2007; Wersing & Körner, 2003), each proposing that processing in the ventral stream is hierarchical and feedforward, involving an increase in the complexity of features to which neurons are selective and in the invariance of their responses to several variables from early to late stages of processing.

Although birds and primates use nonhomologous structures for visual pattern discriminations, there is direct evidence of hierarchical feedforward visual processing in the avian tectofugal pathway. The receptive fields of many neurons in the tectum have a center–surround structure similar to that shown by LGN neurons (Frost, Scilley, & Wong, 1981; Gu, Wang, & Wang, 2000; Jassik-Gerschenfeld & Guichard, 1972). Li, Xiao, and Wang (2007) found that the receptive fields of neurons in the nucleus isthmus, a structure that has reciprocal connections with the tectum, are constructed by feedforward convergence of tectal receptive fields, as proposed by the hierarchical model of Hubel and Wiesel (1962, 1968). Also in accord with hierarchical processing, there is a large increase in receptive field size from the optic tectum to the entopallium (Engelage & Bischof, 1996).

Thus, the main mechanism that is thought to be implicated in visual shape processing in the primate brain also seems to be at work in the avian tectofugal pathway. However, as noted earlier, comparative studies have found important behavioral disparities between people and pigeons in object recognition tasks. Here, we argue that these disparities arise from specialization of processes other than categorization learning and visual shape processing. Our argument is that it is plausible and in accord with the

available behavioral and neurophysiological data that the avian tectofugal pathway, like the primate ventral stream, implements hierarchical and feedforward processing of shape information.

To support this claim, we present the results of several simulations with a “general-processes” model of object categorization in pigeons, which implements both the principles of visual processing discovered in studies of the primate ventral stream and the principles of error-driven categorization learning described earlier. The results of these simulations provide a proof of concept for the hypothesis that the basic principles of visual shape processing are alike in primates and birds, despite the known behavioral disparities.

Explaining avian object perception through principles of visual computation in the primate brain

Model

Visual shape processing in the tectofugal pathway Here, we use a hierarchical model of object recognition in primate cortex recently proposed by Serre and colleagues (Serre, 2006; Serre et al., 2005; Serre et al., 2007), which extends previous work by Riesenhuber and Poggio (1999, 2000). This model has several advantages for our purposes.

First, it was developed with the goal of more faithfully reflecting the anatomy and physiology of primate visual cortex than had previous models. Second, the two main operations deployed by the model (see below) can be implemented by biologically plausible neural microcircuits (Carandini & Heeger, 1994; Kouh & Poggio, 2008; Yu, Giese, & Poggio, 2002); evidence suggests that these operations are used in the visual cortex (Gawne & Martin, 2002; Lampl, Ferster, Poggio, & Riesenhuber, 2004). Third, the model includes a dictionary of about 2.3×10^7 shape feature detectors, whose selectivities have been learned through exposure to natural images (see Serre, 2006). Finally, this model can explain empirical results at several levels of analysis, including behavioral data (Serre et al., 2007). This explanatory success is important, because the present work will focus on behavioral studies of avian object recognition.

Despite these advantages, we caution that this is neither a definitive model of the primate ventral pathway nor a realistic model of the avian tectofugal pathway. Some details about the mechanisms that are proposed in the model are debatable and may be incorrect (see Kayaert, Biederman, Op de Beeck, & Vogels, 2005; Kriegeskorte, 2009; Yamane, Tsunoda, Matsumoto, Phillips, & Tanifuji, 2006). However, we do believe that the model grasps some of the processes that are implemented by the primate ventral

pathway. Specifically, the model implements feedforward, hierarchical processing of visual shape information—the process that we wish to test here. If feedforward and hierarchical processing is important for extracting shape information by the avian tectofugal pathway, the model should be able to explain basic behavioral findings in avian object perception.

We will only sketch the model here. More detailed descriptions can be found in the original articles by Serre and colleagues (Serre, 2006; Serre et al., 2005; Serre et al., 2007) and in the [supplementary materials](#). The model (Fig. 3) is organized into eight layers of two types: layers with “simple” units (S) and layers with “complex” units (C). The units in each layer perform different computations over their inputs.

Simple units perform a feature selection operation, responding to higher-level features that result from a combination of inputs from units in the prior layer tuned to various lower-level features. Each simple unit receives inputs from n_{S_k} units in the prior layer, and it has a preferred pattern of stimulation that is represented by the vector of weights \mathbf{w} . Given a vector \mathbf{x} representing the actual input received by the unit, its response y is a Gaussian-like tuning operation:

$$y = \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n_{S_k}} (w_j - x_j)^2\right) \quad (3)$$

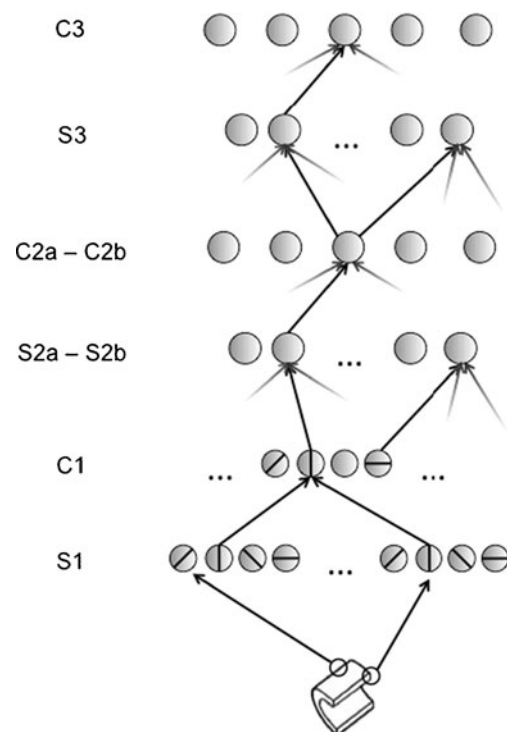


Fig. 3 Global structure of the hierarchical model used in the simulations presented here. Layers of “simple” units are denoted by the letter S, whereas layers of “complex” units are denoted by the letter C. Numbers and noncapital letters index the positions of different layers in the hierarchy

where σ represents the sharpness of the tuning function. Thus, the response of a given unit is a function of the mismatch between the input to the unit and its preferred feature. The weights for units in the first layer of simple units (S_i) take the form of a Gabor function, which has been shown to be a good model for the response properties of V1 simple neurons (Daugman, 1985; Jones & Palmer, 1987). A battery of such Gabor functions is used to directly filter the image that is presented to the model. The values of weights w_j for units in subsequent simple units were learned from exposure to patches of natural images (see Serre, 2006, and the supplementary materials).

Complex units carry out a pooling operation, combining inputs from n_{C_k} units in the prior layer tuned to the same feature over a range of positions and scales. The response y of the complex unit is equal to the maximum value among its inputs:

$$y = \max_{j=1 \dots n_{C_k}} x_j \quad (4)$$

For each simulation, 1,000 units were sampled randomly from each of the four C layers, and their responses were used as input to the classification learning system. Note that, although the pool of feature detectors had a size of 4,000 units, only a very small proportion of these detectors were strongly activated by any image (i.e., a sparse representation), because detectors that were tuned to simple and common features (e.g., C1) had high spatial and scale specificity, whereas detectors exhibiting spatial and scale invariance (e.g., C3) were sensitive to complex feature configurations that were present in only a limited number of objects.

Classification learning in the basal ganglia For classification learning, complex units across the hierarchy project directly to a reinforcement learning stage, which is assumed to be implemented in the basal ganglia. This assumption accords with evidence indicating the existence of projections to the striatum from all areas in the pallium (Veenman et al., 1995).

The classification learning system that was used here is a rendition of our previous model (Soto & Wasserman, 2010a) in terms of a reinforcement learning architecture. Several of the functions that are ascribed to the basal ganglia, together with the role of prediction error in each of these functions, can be captured through reinforcement learning models (Chakravarthy, Joseph, & Bapi, 2010).

Here, we model animal behavior on a trial-by-trial basis. Each trial consists of presentation of a visual stimulus, selection of an action, and updating of weights depending on the reward received. At the beginning of a trial, the stimulus is processed by the visual system, which gives as its output a distributed representation in terms of shape features (the responses of selected complex units, as described earlier). The

response of the i th feature detector to the stimulus is represented as x_i . The weight of the association between x_i and action j is represented as v_{ij} . These values are multiplied by the response of each feature detector to determine the input to the response selection system. As in our prior model (Soto & Wasserman, 2010a), these inputs sum their influence to determine the value of performing action j given the presentation of stimulus s , or Q_{sj} :

$$Q_{sj} = \sum_i x_i v_{ij}. \quad (5)$$

Thus, action values can be computed by the activity of striatal neurons that additively combine the inputs from multiple pallial neurons. Then, these values can be used for action selection in the basal ganglia. In our model, once the value of each action is computed, it is transformed to a response probability through a softmax choice rule:

$$p_j = \frac{e^{\beta Q_{sj}}}{\sum_j e^{\beta Q_{sj}}}, \quad (6)$$

where the parameter β (set to 3.0 in all of our simulations, as in our prior modeling work) represents the decisiveness of the choice rule, with higher values leading to stronger preferences for the choice with the larger incentive value. The probabilities that are obtained from Eq. 6 are used to probabilistically sample an action on a particular trial, in what is known as *Boltzmann exploration* (Kaelbling, Littman, & Moore, 1996).

Once an action is selected, the animal can either receive a reward ($r=1$) if the choice is correct or no reward ($r=0$) if the choice is incorrect. This information is used to compute an action-specific prediction error, which is the discrepancy between the value that is computed for the selected action and the actual reward received:

$$\delta_j = r - \sum_i x_i v_{ij}. \quad (7)$$

It has recently been found that this type of error is indeed signaled by dopaminergic midbrain neurons (Morris, Nevet, Arkadir, Vaadia, & Bergman, 2006; Roesch, Calu, & Schoenbaum, 2007). Given that dopaminergic neurons send projections to striatal neurons, the error signal that they carry could be used for error-driven learning at the pallial–striatal synapses to which they project. This process is captured in our model by updating Q values according to an error-driven learning rule:

$$v_{ij}^{new} = v_{ij}^{old} + \alpha x_i \delta_j, \quad (8)$$

where α is a learning rate parameter that was set to 0.001 in all our simulations. This learning rate was chosen because it

allowed for fast learning of the tasks, with relatively stable trial-by-trial performance at the end of training.

We did not attempt to fit the free parameters of the model to the data or to perform a systematic search of the parameter space to find those values that would yield the most accurate predictions. Rather, we ran all of the simulations with the model parameters fixed to specific values, with the aim of documenting the model's ability to qualitatively reproduce the behavioral patterns that were observed in the experimental data, even with the constraint of using the same parameters in every simulation.

For each simulation, we present the average of 10 model runs. In general, the results varied little across runs of the model (standard errors for all of the simulations can be found in Supplementary Table 1).

Ideal-observer model An important problem in vision research is determining whether the effect of a particular stimulus manipulation on performance could be due to stimulus and task factors instead of to any processing carried out by a visual system. In our case, if the behavior shown by pigeons in a particular task could be explained as arising from the similarities among the stimuli or the structure of the experimental task, our finding that the general-processes model could explain such a behavioral pattern would be uninformative as to its appropriateness as a model of the avian visual system. To provide a benchmark against which to test the performance of the general-processes model, we also simulated each experiment using an ideal-observer model (Tjan, Braje, Legge, & Kersten, 1995). The behavior of this ideal observer represents the optimal performance in each experimental task for an agent relying on the pixel-by-pixel similarities among the images to solve it. The ideal-observer model allowed us to determine which aspects of an experimental outcome could be explained as resulting directly from task and stimulus demands and not from any visual processing carried out by the pigeons. A more detailed description of this model can be found in the [supplementary materials](#).

Simulations

Here, we present the results of five simulations with our general-processes model. Each simulated experiment represents a sample from a different line of research that has received considerable attention in the pigeon literature: view invariance, image properties that control object recognition, the role of spatial information in object recognition, invariance across changes in surface features, and size invariance. These examples have been chosen either because pigeons behave differently from humans in

that experiment or because of the perceived importance of the experiment to characterize object recognition in birds.

Generalization to stimuli rotated in depth Several experiments have tested whether pigeons can recognize objects that are seen at a single viewpoint when they are later rotated in depth. The results of these experiments have uniformly found that pigeons do not show one-shot viewpoint invariance (Cerella, 1977; Lumsden, 1977; Peissig, Young, Wasserman, & Biederman, 2000; Spetch, Friedman, & Reid, 2001; Wasserman et al., 1996). Yet pigeons do show above-chance generalization of performance to novel views of the training object after training with just one view, and they exhibit generalization behavior that is closer to true viewpoint invariance as the number of training views is increased (Peissig, Wasserman, Young, & Biederman, 2002; Peissig et al., 2000; Wasserman et al., 1996).

As is discussed by Kirkpatrick (2001), these results argue against structural description theories (e.g., Biederman, 1987; Hummel & Stankiewicz, 1998) as a full account of pigeons' recognition of objects across viewpoints. Specifically, structural description theories assume that the identity of some simple volumes, called *geons*, should be recoverable by the visual system from almost any viewpoint. Contrary to this assumption, several studies in which pigeons have been trained to discriminate geons have found strongly viewpoint-dependent behavior (Peissig et al., 2002; Peissig et al., 2000; Spetch et al., 2001). These results accord better with image-based theories (e.g., Poggio & Edelman, 1990; Tarr & Pinker, 1989; Ullman, 1989), which predict both a drop in accuracy to novel views of the training object and a reduction of this generalization decrement as the number of training views is increased.

In a study conducted by Spetch et al. (2001), pigeons were trained to discriminate two objects in four conditions. In the zero-part condition, both objects were composed of five elongated cylinders that were joined in their extremes at different angles (i.e., "paperclip" objects). All of the other conditions involved the replacement of one, three, or all five cylinders in each object by distinctive geons (see Fig. 4a); these were the one-, three-, and five-part conditions, respectively. The birds were trained with two views (0° and 90° of rotation) of each object, and they were tested with six different views (0°, 30°, 90°, 135°, 160°, and 225° of rotation). As discussed by Tarr, Bülthoff, Zabinski, and Blanz (1997), image-based theories predict viewpoint-invariant recognition when a diagnostic feature is available, as in the one-part condition, but not when a diagnostic feature is absent, as in all of the other conditions. Importantly, this is exactly the result that was found with people. Structural description theories predict viewpoint-invariant recognition whenever an object can be easily decomposed into simple volumes that are arranged in a

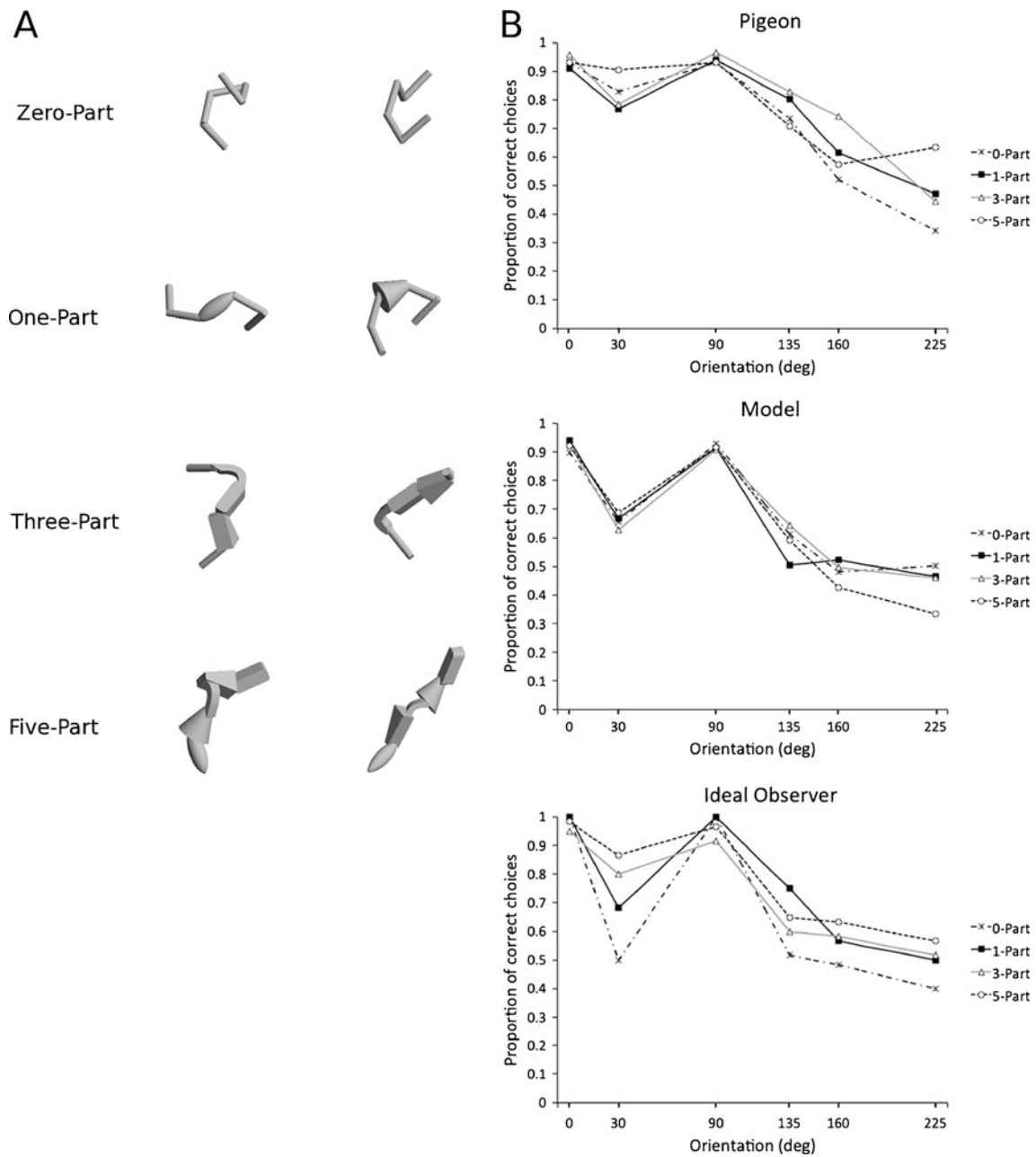


Fig. 4 (a) Examples of the stimuli used by Spetch et al. (2001) to test the effect of adding distinctive parts to an object on the object's recognition from novel viewpoints. (b) Pigeon data (top) and results simulated with the general-processes model (middle) and an ideal-

observer model (bottom) for the experiment carried out by Spetch et al., testing the effect of adding distinctive parts to an object on its recognition from novel viewpoints

particular way, as in all of the conditions except the zero-part condition.

The experimental results are shown in the top panel of Fig. 4b. It can be seen that pigeons' accuracy with novel views of the objects (30°, 135°, 160°, and 225° of rotation) progressively decreased as a function of distance from the training views in all four conditions. This result cannot be explained by either structural description models or image-based models, and it is different from what was found with

human participants, who experienced the same experimental procedures. Is it possible that a model based on mechanisms known to function in primates could explain the pigeons' results, despite their being different from humans' results?

The answer is "yes," as indicated by the results of a simulation with the general-processes model that is shown in the middle panel of Fig. 4b. Just as was observed for the birds, the model yielded essentially the same pattern of results in all of the conditions: high levels of performance at

the training views, and a decrement in accuracy as a function of distance from the training views. The model did not yield generalization levels quite as high as those observed for the pigeons. Also, the model did not reproduce the order of performances across conditions within each of the test orientations. For example, performance in the five-part condition was higher than in the other conditions at 30° and 225° of rotation. The authors did not report statistical tests for these effects, so it is unknown whether any of them are reliable. The most important effect reported in Spetch et al. (2001)—a decrement in performance as a function of rotational distance, regardless of condition—is reproduced by the general-processes model. Thus, a model that is based on mechanisms that are known to function in primates behaved more similarly to pigeons than to people in this experiment.

The results of a simulation with the ideal observer, shown in the bottom panel of Fig. 4b, suggest that most aspects of the experimental results can arise from the pattern of low-level similarities in the stimuli. Other aspects of the experimental results, however, argue against this explanation. For example, the ideal observer cannot explain why pigeons showed above-chance performance with the 30° and 135° testing views in the zero-part condition. Thus, good performance with these testing views cannot be due to physical similarity between the training and testing stimuli.

More recent results have also argued against an explanation of object recognition in birds in terms of an image-based model involving either mental rotation (Tarr & Pinker, 1989) or interpolation between viewpoints in memory (Bülthoff & Edelman, 1992). Peissig et al. (2002) trained pigeons to discriminate grayscale images of four geons. They found that training with multiple views of an

object along one axis of rotation enhanced generalization to novel views of the object along both the *training* axis and an *orthogonal* axis. The results from their most important test, involving novel views of the objects rotated along an orthogonal axis, are plotted in the left panel of Fig. 5. The radial coordinate in the polar plot represents the proportion of correct responses, whereas the angular coordinate represents the degree of rotation from the training image (assigned to 0° of rotation). The chance level for this test was .25 for correct responses, which is indicated by the small circular lines in the radial charts displayed in Fig. 5. It can be seen that training with multiple views of the objects enhanced recognition at *all* of the novel viewpoints. The pigeons could not have used direct interpolation between the training views to aid performance with novel views along the orthogonal axis. Also, the amount of mental rotation that was required to recognize novel views along the *orthogonal axis* from the nearest training view was the same for pigeons that were given either one view or five views of the objects along the training axis. Thus, this pattern of results cannot be captured by several traditional cognitive models of object recognition. Although data from adult humans are not available for this experimental design, human infants have exhibited the same pattern of results shown by the pigeons in this study (Mash et al., 2007).

The results of a simulation with the general-processes model are depicted in the middle panel of Fig. 5. Although the model again does not achieve the high levels of generalization that were shown by the pigeons, it does successfully reproduce the consistent increase in generalization performance after training with multiple views. This increase occurs because it is likely that those properties of objects that are invariant across training views are also

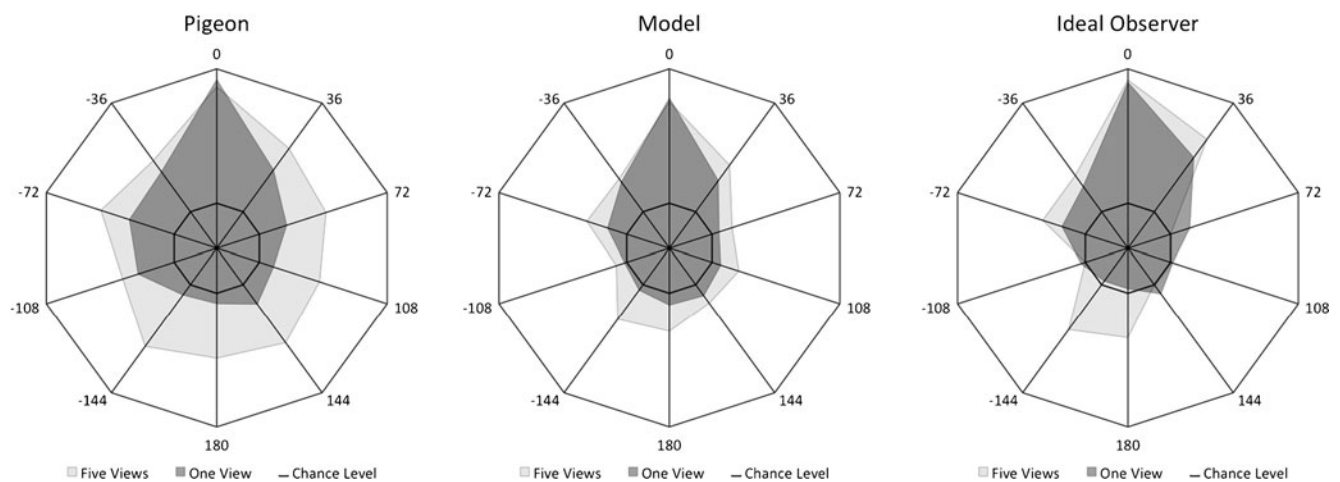


Fig. 5 Pigeon data (left) and results simulated with the general-processes model (middle) and an ideal-observer model (right) for the experiment performed by Peissig et al. (2002), testing the effect of training with views of an object rotated in depth along one axis on its recognition when rotated along an orthogonal axis. The radial

coordinate in each polar plot represents the proportion of correct responses, whereas the polar plot represents the degrees of rotation from the training image, the latter corresponding to 0° of rotation. The chance level for this test was .25 for correct responses and is indicated by the small circular line in each of the radial charts

present in many other novel testing views, regardless of the axis chosen to rotate the object and produce the testing views. Increasing the number of training views increases the likelihood of common properties between the testing and training objects; it also increases behavioral control by those properties that are common to several training images, because these properties are presented and associated with the correct responses more often than those properties that are idiosyncratic to each image.

The results of a simulation with the ideal observer, shown in the right panel of Fig. 5, suggest that low-level stimulus similarity may have boosted generalization performance in the multiple-views group to some of the novel views, but not to all of them. Specifically, performance levels were actually *lower* in the five-views condition than in the one-view condition at 72°, 108°, 144°, and -108° of rotation.

In summary, although the general-processes model does not promote the same high levels of generalization to novel object views that were exhibited by the birds, it does reproduce the qualitative patterns of results that were observed in studies of rotational invariance.

Bias to rely on nonaccidental properties of images for object recognition Structural description theories of object recognition (e.g., Biederman, 1987; Hummel & Stankiewicz, 1998) propose that humans extract the three-dimensional structure of objects from nonaccidental edge properties in retinal images, such as parallelism and cotermination. Indeed, empirical results indicate that nonaccidental properties are particularly important features for object recognition by primates and birds (Biederman & Bar, 1999; Gibson, Lazareva, Gosselin, Schyns, & Wasserman, 2007; Kayaert, Biederman, & Vogels, 2003; Lazareva, Wasserman, & Biederman, 2008; Vogels, Biederman, Bar, & Lorincz, 2001).

Gibson et al. (2007) trained pigeons and people to discriminate four geons that were shown at a single viewpoint. Once high performance levels were reached, the researchers used the Bubbles procedure (Gosselin & Schyns, 2001) to determine which properties of the images were used for recognition. With this procedure, the information in images is partially revealed to an observer through a number of randomly located Gaussian apertures, or “bubbles,” and the observer’s response is used to determine which areas in an image are used to correctly recognize an object. Gibson et al. defined regions of interest in their images containing information about edges, shading, and edge cotermination. Figure 6a shows some examples of these different types of regions in one of the geons. This manipulation made it possible to determine the extent to which the information used by the subjects, as revealed through the Bubbles procedure, was contained in each region of interest. The left section of Fig. 6b shows the

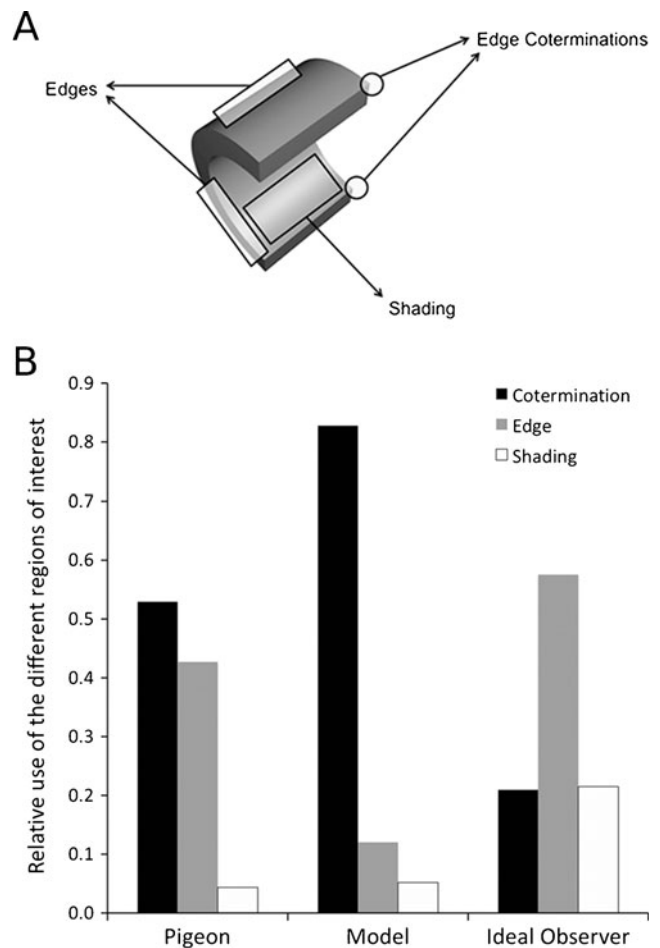


Fig. 6 (a) Examples of regions of a geon that were classified as containing edges, edge coterminations, and shading information in the Gibson et al. (2007) study. (b) Relative uses of cotermination, edge, and shading regions by pigeons (left), the general-processes model (middle), and an ideal-observer model (right) when trained with the procedures and stimuli from the Gibson et al. study

reanalyzed pigeon data from the Gibson et al. experiment. Pigeons relied on cotermination information more strongly than on other edge properties, and on those two edge properties much more strongly than on shading. The human data (not included in Fig. 6b) showed the same qualitative pattern of results.

As is illustrated in the center section of Fig. 6b, the general-processes model correctly reproduces the bias that was observed in pigeons and people to rely on cotermination information for the discrimination of simple volumes. The results from the ideal observer, shown in the right section of Fig. 6b, confirmed the conclusion reached in the original study that cotermination information was not the most informative nonaccidental property in this particular task. According to the ideal observer, in this task, the most diagnostic property for recognition was the objects’ edges. These results are particularly important, as they suggest that the bias that was shown by pigeons and people to rely on

nonaccidental properties does not require the explicit representation of three-dimensional volumes by the visual system, as proposed by structural description theories, but emerge from quite simple principles of biological visual computing.

The general-processes model combines the inputs from units that are sensitive to oriented edges, which are located early in the hierarchy of feature detectors, to produce the receptive fields of units in later stages. Because the preferred stimuli for units in later stages have been learned through experience with natural images, it is likely that they include both coterminations and elongated edges, which are frequently found in natural objects.

It is less clear why cotermination information was preferred over other edge properties to solve the task, especially because, in natural images, elongated edges are the more common of those two properties (Geisler, Perry, Super, & Gallogly, 2001). One possibility is that tolerance to image translation and scaling in the model renders the elongated contours of an object less reliable for identification, but the same process does not affect object coterminations to the same extent. Whereas two parallel edges can be perfectly aligned after translation, two corners are less likely to have the same local configuration after translation. In other words, two close, approximately parallel edges are

very likely to be confused as the same feature by an invariant edge detector. Two coterminations belonging to the same object, however, are more likely to have a different configuration (e.g., a corner pointing to the left vs. a corner pointing to the right), which would resolve the ambiguity added by invariant feature detection. Thus, translation-invariant detectors are more likely to confuse two edges than two coterminations if they are shown through slightly translated “bubbles.”

Recognition of complementary and scrambled object contours The results of several experiments led Cerella (1986) to conclude that pigeons recognize objects using local features, while ignoring their spatial relations and other global properties. This proposal stimulated several studies showing that pigeons do in fact represent spatial relations among features (Kirkpatrick-Steger & Wasserman, 1996; Kirkpatrick-Steger, Wasserman, & Biederman, 1998; Van Hamme, Wasserman, & Biederman, 1992; Wasserman, Kirkpatrick-Steger, Van Hamme, & Biederman, 1993).

In one of these experiments (Van Hamme et al., 1992), pigeons were trained to recognize line drawings of four objects, such as those in the left and center of Fig. 7a. The stimuli were created by deleting half of an object’s contour,

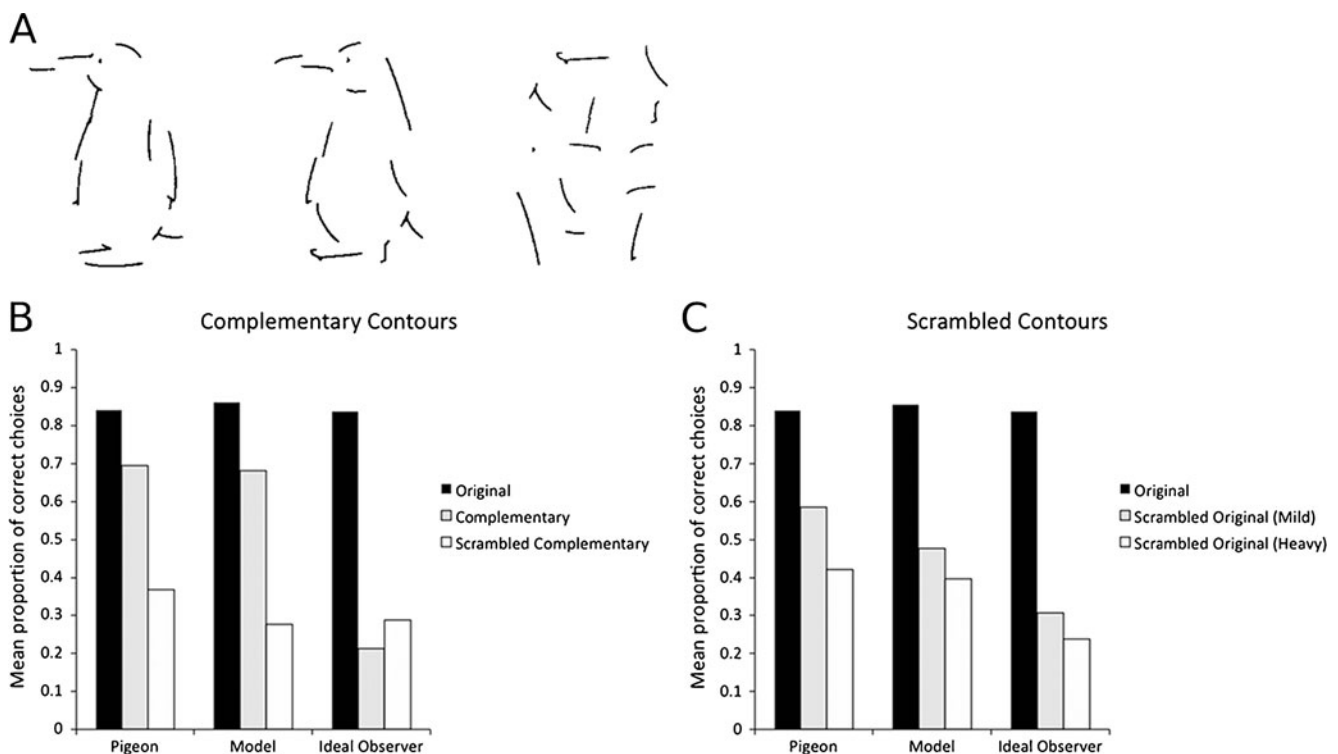


Fig. 7 (a) Some of the stimuli used by Van Hamme et al. (1992) to study the transfer of performance across complementary contours of an image (left, middle), and the scrambled version of the contours (right). (b) Pigeon data (left) and results simulated with the general-processes model (middle) and an ideal-observer model (right) for Van Hamme et al.’s Experiment 2, in which they evaluated transfer of

recognition performance across complementary contours of the training images and their scrambled versions. (c) Pigeon data (left) and results simulated with the general-processes model (middle) and an ideal-observer model (right) for Van Hamme et al.’s Experiment 3, testing the effect of different degrees of spatial scrambling on the recognition of object line drawings

which allowed the investigators to train the pigeons to recognize one partial object contour and to test with the complementary contour. If the pigeons represented these objects using only the local lines and vertices that were available in the training images, then they should fail to accurately recognize the complementary contours, and their performance should be close to the chance level of .25 correct choices. However, as shown in Fig. 7b, pigeons recognized these complementary contours at a high level of accuracy, suggesting that their visual system could infer object structure from the partial contours that were seen during training. To control for the possibility that this result was due to similarity between local features in the original and complementary contours, the pigeons were also tested with stimuli in which the complementary contours were spatially scrambled, as in the image shown in the right part of Fig. 7a. Pigeons' performance dropped precipitously to the scrambled stimuli (see Fig. 7b), although it was still a bit above chance, suggesting that both local features and spatial structure are important in pigeons' recognition of object contours.

A follow-up experiment, in which pigeons were tested with scrambled versions of the original training stimuli, strengthened this conclusion and further suggested that the intact spatial organization of an object contour might be more important for recognition than local features. As is shown in Fig. 7c, accuracy with scrambled versions of the original stimuli—which retained all of the local features and disrupted the drawings' spatial organization—was lower than with their complementary contours (see Fig. 7b), which retained all of the spatial organization and disrupted the drawings' local features. The test results also indicated that the effect of scrambling image features depended on whether the spatial translation of features was mild or heavy. Mild scrambling involved translating the original features within their original quadrants, whereas heavy scrambling involved translating half of the features to the adjacent (clockwise) quadrant. Heavier spatial translation of features led to larger drops in performance.

These and other experiments in which pigeons were tested with scrambled versions of object contours have led to the rejection of an explanation of avian object recognition via a "bag-of-features" representation, as proposed by Cerella (1986). Although the general-processes model used here also uses a bag-of-features representation, it retains information about the spatial structure of objects by using a large bank of overlapping features. The question is, however, whether using a bank of overlapping features is enough to support the generalization of performance across complementary contours, which do not share any local features.

As shown in Fig. 7b and c, the answer is "yes." The general-processes model can reproduce all of the key aspects of the results reported by Van Hamme et al.

(1992). In contrast, the ideal-observer model yielded performance that was at or near chance to all of the testing stimuli in both experiments. The high levels of generalization to complementary contours that were shown by the general-processes model could be due to two mechanisms. First, units at the end of the processing hierarchy are both selective to complex shapes and relatively invariant to spatial translation and scaling. Some line contours in the two complementary images were quite similar, but they appeared in slightly different positions in the image. In the example shown in Fig. 7a, the complementary contours in the chest and belly of the penguin are quite similar and are located in only slightly different spatial positions. Line contours such as these could have activated the same feature detectors in the two complementary images. Second, two contours in the complementary images could have moderately activated the same unit if they both overlapped with different sections of its preferred stimulus. In the example shown in Fig. 7a, a unit that is selective for a long vertical curve could have been activated by the two halves of the penguin's back that were present in the two complementary images.

The effects of scrambling are reproduced because of the large number of overlapping feature detectors that were included in the general-processes model, which retain considerable information about the spatial organization of each object. Importantly, this model can also reproduce the above-chance performance that was observed with the scrambled stimuli, which results from the fact that some diagnostic visual features are not affected by scrambling.

The results of these simulations suggest a way to reconcile the contradictory results found by Cerella (1980), on the one hand, and by more recent studies (e.g., Kirkpatrick-Steger et al., 1998; Wasserman et al., 1993), on the other. One important aspect of Cerella's (1980) experiment is that he divided cartoon objects into only three parts (legs, torso, and head) that preserved substantial local spatial information; furthermore, these parts were horizontally aligned when they were scrambled, which retained some global properties, such as the object's aspect ratio. Thus, Cerella's (1980) manipulation might have left untouched much of the spatial structure to which the pigeons were sensitive, not affecting their performance to a substantial degree. On the other hand, in more recent studies, objects were divided into many parts and their scramblings were misaligned. These stimulus manipulations might have disrupted spatial structure information at several different scales. If, as proposed here, pigeons represent objects through a large bank of overlapping feature detectors, the effects of the spatial rearrangement of an object's parts would depend on exactly how the rearrangement was accomplished, because spatial information is implicitly represented and distributed across all of

the feature detectors. Some of these detectors implicitly code for the arrangement of local features, whereas others code for the more global arrangement of object parts. Because all of these aspects of the stimulus can gain control over behavior during learning, the results of a scrambling test critically depend on what type of spatial information is affected by the scrambling procedure.

Effect of changes in surface features Pigeons are very sensitive to transformations of an object that leave its outline intact, while changing only its internal surface features. Several studies (Cabe, 1976; Cook, Wright, & Kendrick, 1990; Young, Peissig, Wasserman, & Biederman, 2001) have provided evidence that the drop in object recognition performance that is caused by the deletion of surface features is larger when pigeons are tested with line drawings than when they are tested with silhouettes.

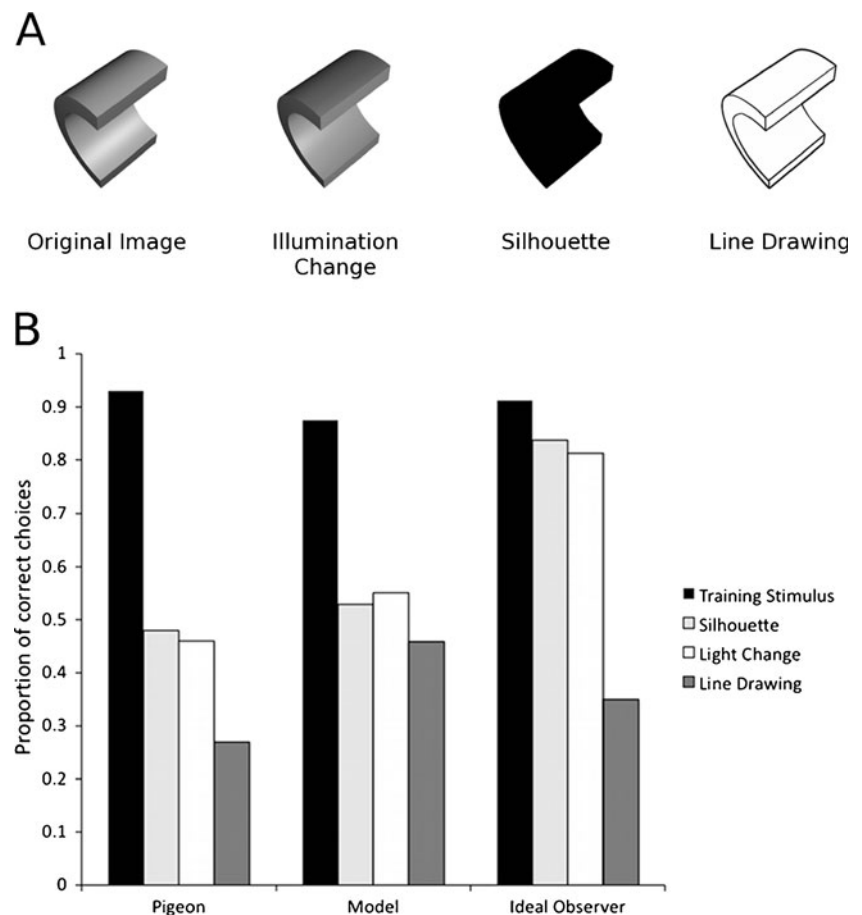
Young et al. (2001, Exp. 1) trained pigeons to discriminate four geons that were rendered from a single viewpoint. After they achieved high levels of performance in the task, the birds were tested with the same geons transformed in three ways (see Fig. 8a): (1) under a change in the direction of illumination; (2) with all of the internal shading removed and replaced by the average luminance of the object, producing a

silhouette; or (3) with the object's internal shading removed, but with all of the edge information retained, producing a line drawing. The results are presented in the left part of Fig. 8b and show that all of the transformations produced large drops in accuracy. Performance with the line drawings was the poorest, being close to the chance level of .25, whereas performance levels with the silhouette and light change stimuli were somewhat better and similar. This pattern of results is rather surprising, because both the line drawing and the light change stimuli contained much more information about the three-dimensional structure of the objects than did the silhouette stimuli, while conserving the same contour information.

The middle part of Fig. 8b shows that this qualitative pattern of results is reproduced by the general-processes model, which produces large drops in accuracy to all of the testing stimuli, with the largest drop to the line drawings. The right part of Fig. 8b shows that large drops in performance were not invariably predicted by the ideal-observer model; indeed, this model predicted a large drop in performance only for the line drawings.

Why does the general-processes model produce such large drops in performance with these test stimuli? The answer lies in the fact that, at the beginning of the hierarchy

Fig. 8 (a) Examples of the types of testing stimuli used by Young et al. (2001) to test the recognition of geons by pigeons across variations in surface features: illumination change, silhouette, and line drawing versions of the original training stimuli. (b) Pigeon data (left) and results simulated with the general-processes model (middle) and an ideal-observer model (right) for Young et al.'s Experiment 1, which tested recognition of geons across variations in surface features: silhouette, illumination change, and line drawing versions of the original training stimuli



of processing, the image is filtered using a battery of Gabor functions at several different scales. In this wavelet-like decomposition of the image (Field, 1999; Stevens, 2004), which is carried out by V1 neurons in the primate brain, filters at the largest scales can detect changes in luminance such as those produced by a gray object over a white background; filters at smaller scales can detect changes in luminance such as those produced by the reflection of light on the smooth surface of the object; and filters at the smallest scales can detect local changes in luminance such as those produced by the object's edges. During learning, the model relies on all of this information for recognition. The more information about luminance changes that is removed from the original object, regardless of the spatial scale of these changes, the more performance will suffer.

One aspect of the results that was not reproduced by the general-processes model was that the pigeons' performance with line drawings was at chance. Because pigeons can readily discriminate line drawings of complex objects (Kirkpatrick, 2001), this result cannot be due to the birds' inability to detect the fine-edge information in the images. Instead, this recognition failure might result from a bias, not reproduced by the model, toward relying more heavily on luminance changes at larger scales whenever this information is available. That is, pigeons may rely much more heavily than the model on the coarse disparities in luminance that exist between the inside and the outside of each object, so that when this information is removed, performance drops precipitously.

Effects of exponential and linear variations in size Many studies have tested pigeons' recognition of objects varying in size (Larsen & Bundesen, 1978; Peissig, Kirkpatrick, Young, Wasserman, & Biederman, 2006; Pisacreta, Potter, & Lefave, 1984). In general, pigeons show generalization of performance to known objects at novel sizes. Nevertheless, they also show decrements in accuracy to novel sizes, with larger decrements to more disparate sizes.

Peissig et al. (2006) also explored which types of size transformations yield symmetrical generalization gradients. Four geons were linearly varied in size, by increasing size in constant steps, or exponentially varied in size, by increasing size in steps approximately proportional to the current size. Examples of both transformations are shown in Fig. 9a. Pigeons were trained with the size in the middle of both scales (geons labeled "0" in Fig. 9a) in a four-alternative forced choice task until they reached high levels of recognition accuracy. Then they were tested with the linear and exponential size variations, producing the generalization gradients shown in the left part of Fig. 9b. The exponential scale produced a more symmetrical generalization gradient, whereas the linear scale yielded smaller drops in performance for stimuli that increased in

size than for stimuli that decreased in size by the same amount. These results suggest that visual size conforms to Fechner's law: namely, a logarithmic relation between physical and perceived size.

The gradients resulting from a simulation of this experiment with the general-processes model are shown in the middle part of Fig. 9b, and the ideal-observer results are shown in the right part of this figure. The general-processes model reproduced all of the important aspects of the behavioral results: above-chance performance for all sizes, which decreased monotonically as a function of the difference from the original size; a symmetrical gradient for stimuli that changed along an exponential scale; and an asymmetrical gradient for stimuli that changed along a linear scale. None of these results was predicted by the ideal observer. The general-processes model did not reproduce the small generalization decrement that was observed with sizes -1 and 1 in the linear transformation condition. However, the qualitative pattern of results was nicely reproduced, including the specific relation between the physical image changes and the corresponding changes in performance.

The complexity of the model makes it difficult to specify why the experimental results were reproduced. One possibility is that, in a linear size transformation, more feature detectors that respond to the training image reduce their response with an increase in size than with a decrease in size. Imagine a simplified, unidimensional model, in which a set of edge detectors, all of the same size and scale, are collinearly arranged. The model is presented with a line, and a number of units respond vigorously and are associated with a response. If we linearly increase the size of this line, the same edge detectors will still respond, and high generalization will be observed. However, if we linearly decrease the size of this line, a number of edge detectors that had previously responded will stop doing so, and a response decrement will be observed. This process could be transpiring, at a larger scale, with the linear size transformation. In the exponential size transformation, though, decrements are smaller than increments. In our example, the size increments might be so dramatic as to "misalign" the edges in the object with the original detectors, producing a stronger generalization decrement because the object's edges are now outside the detectors' receptive fields.

Discussion

This article integrates neurobiological, behavioral, and computational evidence for the involvement of "general processes" in object recognition by people and pigeons. Our main goal here was to present arguments and evidence

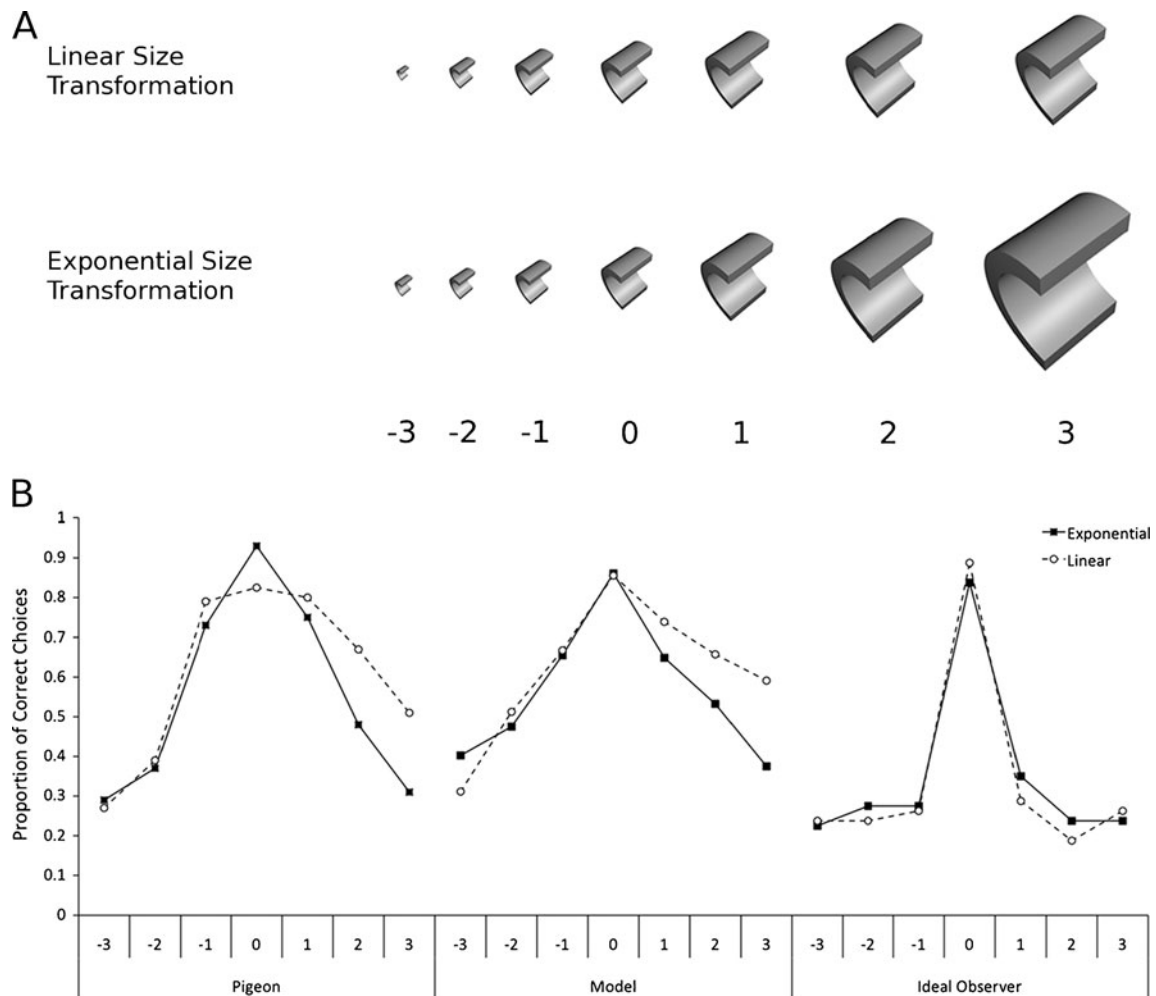


Fig. 9 (a) Examples of the types of stimuli used by Peissig et al. (2006) to test the recognition of geons by pigeons across linear and exponential size variations. Numbers represent scaling steps from the original training images (denoted by 0). (b) Pigeon data (left) and

results simulated with the general-processes model (middle) and an ideal-observer model (right) for Experiment 2 in Peissig et al. (2006), which tested the recognition of geons across linear and exponential size variations

suggesting (1) that differences between birds and primates should not be taken as evidence of completely specialized visual systems, (2) that it is likely that these two types of animals share common processes of object recognition, and (3) that birds can and should be used by the cognitive neuroscience community as a model for the study of *some* of the processes that are involved in object recognition by people.

Specifically, we reviewed prior work suggesting a role for error-driven associative learning in visual object categorization and we presented new computational evidence that some of the principles of visual processing discovered in the primate ventral stream can account for a wide range of behavioral results from studies of avian object perception. The successful series of simulations provides a proof of concept that general mechanisms of visual shape processing underlie key aspects of object recognition across even distantly related vertebrate species. The computational evidence adds to a wealth of

empirical results that point to the existence and significance of such general processes.

We do caution that these results should not be taken as more than a proof of concept, unless future research on the neurophysiology of the avian tectofugal visual pathway reveals that the mechanisms embedded in feedforward hierarchical object recognition architectures are indeed at work in birds, as they are in primates. As we noted earlier, there is some empirical support for this assertion (Li et al., 2007), but evidence from neural recordings across all stages of the tectofugal pathway and its pallial targets will be needed in order to reach a stronger conclusion.

Relatedly, we emphasize that we have used the model of Serre and colleagues (Serre, 2006; Serre et al., 2005; Serre et al., 2007) simply as a tool to provide such proof of concept. It is likely that this model does not capture the full complexity of visual processing in the ventral stream. However, we have used this model because it implements

feedforward and hierarchical processing of shape information, a mechanism that seems to be backed up by an important amount of evidence from primate neurophysiology (reviewed earlier). Our conclusions will remain valid unless new data disprove the central role that is currently assigned to feedforward and hierarchical processing in the construction of receptive-field selectivity and invariance for neurons in the ventral stream.

Regardless of these considerations, it is noteworthy that a model rooted in primate neurobiology does such a good job explaining avian object recognition, given the behavioral and neurobiological differences in object recognition by birds and primates reviewed in the introduction. We hope that the present computational work, together with our previous empirical work, will encourage neuroscientists to further test the hypothesis of common neurocomputational mechanisms and to consider birds as potentially fruitful models for the study of biological object recognition.

If, as we have suggested, feedforward and hierarchical visual shape processing is common to such distantly related taxa as birds and primates, then two important questions immediately arise. The first and most obvious question, from a comparative standpoint, is whether these common processes are the results of conservation or of independent evolution. Some might propose that the results of studies of avian object recognition can only inform the study of human object recognition if the underlying structures are homologous. We, on the contrary, believe that even if the neural structures underlying object recognition in birds and primates are not homologous, but do implement similar mechanisms of visual processing, then the study of such mechanisms in birds at both behavioral and neurobiological levels can give us important insights into biological vision, in general, and human vision, in particular.

The fact that visual environments seem to select one particular form of visual computation and not others prompts the second question, which is even more important for our understanding the evolution of cognition: Namely, what gives this form of computation its “generality” for solving visual problems across different environments? If feedforward hierarchical processing is truly a mechanism that is shared by many vertebrate species, then it will be critical for future work to provide insight into which aspects of the varied visual tasks that are faced by these species have shaped the way in which this form of visual processing works.

A key point of discussion is that the model that we have used yielded results that more closely approximated pigeon than human behavior in several experimental designs. The model was sensitive to a number of stimulus transformations to which pigeons, but not people, are also sensitive. People exhibit viewpoint-invariant object recognition when they are tested with the appropriate stimuli (Biederman & Gerhardstein, 1993) and with interpolated novel views of

an object (Spetch & Friedman, 2003; Spetch et al., 2001); they exhibit scale-invariant recognition (Biederman & Cooper, 1992) and they do not exhibit dramatic drops in performance after the deletion of internal surface features (Biederman & Ju, 1988; Hayward, 1998). If our interpretation of the simulation results presented here in terms of general processes is correct, then the behavioral disparities that have been found between pigeons and people are likely to be the consequence of differences in processes other than the extraction of visual shape information and error-driven categorization learning. That is, whereas these two mechanisms may be common to both birds and primates, both types of animals may also show more specialized processes that have an impact on object recognition. As was noted earlier, this pattern of results is exactly what we would expect by studying most forms of cognition from an evolutionary perspective.

For example, better recognition with interpolated views of an object, reported in people but not in pigeons (Spetch & Friedman, 2003; Spetch et al., 2001), might require the ability to manipulate representations of objects through either “mental rotation” (Tarr & Pinker, 1989) or view interpolation (Bülthoff & Edelman, 1992); these processes might be absent in birds.

Regarding specialized adaptations in categorization learning, Ashby and colleagues (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Valentin, 2005) have hypothesized two categorization learning systems in people. The *implicit* system is based on procedural learning and is believed to be implemented by the circuitry of the basal ganglia, similar to the categorization system that we have proposed. The *explicit* system is based on logical reasoning, working memory, and executive attention, and is believed to be implemented by the frontal cortex. Recent evidence (Smith et al., 2011) has suggested that pigeons, unlike people and other primates, might only possess the implicit system, which they use to learn any categorization task. If this conclusion is correct, it is an open question how the explicit system might influence visual categorization tasks that are usually encountered in natural environments, such as object recognition across changes in viewpoint. The explicit categorization system, or any of its components, could underlie many of the disparities that have been observed between people and pigeons in object recognition.

Note, however, that we have found evidence suggesting that error-driven learning is involved in human categorization of natural objects, even when an explicit verbal rule can be used to learn a particular task (see Soto & Wasserman, 2010b). Thus, error-driven learning may underlie not only the strengthening of stimulus–response associations, but also the competition between the outputs of these different systems for control over performance.

The strictly feedforward model that we have used does not take into account the fact that connections across the ventral pathway are bidirectional (Salin & Bullier, 1995). Whereas feedforward connections are thought to be responsible for driving the properties of neuronal responses to external stimuli, feedback connections—which are more diffuse—are thought to have a modulatory effect on the activity of neurons (Kveraga, Ghuman, & Bar, 2007).

One possibility is that feedback projections implement attentional processes that facilitate the processing of the information that is relevant for a particular task (see Gilbert & Sigman, 2007; Kastner & Ungerleider, 2000). Attention and other processes that are carried out through the feedback visual system might be very important for some aspects of human object recognition, such as the manipulation of object representations and the selective filtering of object information that is irrelevant for a particular task (such as surface features or viewpoint-dependent shape properties). This role of the feedback system might explain why the model that we used cannot explain some aspects of human behavior. It also suggests the interesting possibility that the model does so well in explaining pigeons' behavior because the feedback system in these animals does not carry out the same functions that it does in primates.

Regardless of the actual underlying cause(s) of the behavioral disparities that were seen between primates and birds, our modeling results underscore the importance of adopting an atomistic and “bottom-up” approach (de Waal & Ferrari, 2010; Shettleworth, 2010a) to the comparative study of cognition. Even in the face of behavioral and neuroanatomical disparities between species, one cannot dismiss the possibility that some of the subprocesses underlying a complex behavior are shared. In the case of object recognition, some discrepancies suggest that the mechanisms of object recognition might be different in primates and birds. Our simulations suggest that this possibility is not necessarily the case for visual shape processing mechanisms: Most aspects of object recognition that seem to be unique to birds can emerge from computational principles that are known to exist in the primate visual system.

The proposal that complex cognition can arise from the interaction of simple, basic processes is widespread among theories of human cognition. Comparative cognition can be instrumental in the discovery of the basic computational processes that bring about complex cognition in biological systems. We believe that this search should focus on general processes that are common to many species. Such processes, selected by diverse environments throughout evolution, are a solution to what are likely to be the most vital aspects of the computational problems that are faced by biological systems; these processes should be deemed the cornerstone on which biological cognition rests.

Author note This research was supported by National Institute of Mental Health Grant MH47313 and by National Eye Institute Grant EY019781. We thank Marcia Spetch for providing some of the stimuli that were used in our simulations.

References

- Antzoulatos, E. G., & Miller, E. K. (2011). Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron*, *71*, 243–249. doi:10.1016/j.neuron.2011.05.040.
- Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, *10*, 1313–1321. doi:10.1038/nn1975.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481. doi:10.1037/0033-295X.105.3.442.
- Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 1–36). New York: Academic Press.
- Ashby, F. G., & Valentin, V. V. (2005). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Categorization in cognitive science* (pp. 548–572). New York: Elsevier.
- Atkinson, A. P., & Wheeler, M. (2004). The grain of domains: The evolutionary–psychological case against domain-general cognition. *Mind and Language*, *19*, 147–176. doi:10.1111/j.1468-0017.2004.00252.x.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147. doi:10.1037/0033-295X.94.2.115.
- Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, *39*, 2885–2899.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 121–133. doi:10.1037/0096-1523.18.1.121.
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1162–1182. doi:10.1037/0096-1523.19.6.1162.
- Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 38–64.
- Bitterman, M. E. (2000). Cognitive evolution: A psychological perspective. In C. M. Heyes & L. Huber (Eds.), *The evolution of cognition* (pp. 61–79). Cambridge, MA: MIT Press.
- Bolam, J. P., Hanley, J. J., Booth, P. A. C., & Bevan, M. D. (2000). Synaptic organisation of the basal ganglia. *Journal of Anatomy*, *196*, 527–542.
- Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, *11*, 747–759. doi:10.1038/nrn2931.
- Boynton, G. M., & Hegde, J. (2004). Visual cortex: The continuing puzzle of area V2. *Current Biology*, *14*, R523–R524.
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, *89*, 60–64.
- Cabe, P. A. (1976). Transfer of discrimination from solid objects to pictures by pigeons: A test of theoretical models of pictorial perception. *Perception & Psychophysics*, *19*, 545–550.
- Cadiou, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and

- invariance. *Journal of Neurophysiology*, 98, 1733–1750. doi:10.1152/jn.01265.2006.
- Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science*, 264, 1333–1336.
- Centonze, D., Picconi, B., Gubellini, P., Bernardi, G., & Calabresi, P. (2001). Dopaminergic control of synaptic plasticity in the dorsal striatum. *European Journal of Neuroscience*, 13, 1071–1077.
- Cerella, J. (1977). Absence of perspective processing in the pigeon. *Pattern Recognition*, 9, 65–68.
- Cerella, J. (1980). The pigeon's analysis of pictures. *Pattern Recognition*, 12, 1–6.
- Cerella, J. (1986). Pigeons and perceptrons. *Pattern Recognition*, 19, 431–438.
- Chakravarthy, V. S., Joseph, D., & Bapi, R. S. (2010). What do the basal ganglia do? A modeling perspective. *Biological Cybernetics*, 103, 237–253.
- Cook, R. G., Wright, A. A., & Kendrick, D. F. (1990). Visual categorization by pigeons. In M. L. Commons, R. J. Herrnstein, S. Kosslyn, & D. Mumford (Eds.), *Behavioral approaches to pattern recognition and concept formation* (pp. 187–214). Hillsdale, NJ: Erlbaum.
- Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 85–116). New York: Cambridge University Press.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America. A*, 2, 1160–1169.
- de Waal, F. B. M., & Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences*, 14, 201–207. doi:10.1016/j.tics.2010.03.003.
- Durstewitz, D., Kröner, S., & Güntürkün, O. (1999). The dopaminergic innervation of the avian telencephalon. *Progress in Neurobiology*, 59, 161–195.
- Engelage, J., & Bischof, H.-J. (1996). Single-cell responses in the ectostriatum of the zebra finch. *Journal of Comparative Physiology. A*, 179, 785–795.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1–47.
- Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Review of Neuroscience*, 23, 441–471.
- Field, D. J. (1999). Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions of the Royal Academy A*, 357, 2527–2542.
- Fredes, F., Tapia, S., Letelier, J. C., Marín, G., & Mpodozis, J. (2010). Topographic arrangement of the rotundo-entopallial projection in the pigeon (*Columba livia*). *The Journal of Comparative Neurology*, 518, 4342–4361.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, 23, 5235–5246.
- Frost, B. J., Scilley, P. L., & Wong, S. C. P. (1981). Moving background patterns reveal double- opponency of directionally specific pigeon tectal neurons. *Experimental Brain Research*, 43, 173–185. doi:10.1007/BF00237761.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, 193–202.
- Gallistel, C. R. (1999). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1179–1191). Cambridge, MA: MIT Press.
- Gawne, T. J., & Martin, J. M. (2002). Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *Journal of Neurophysiology*, 88, 1128–1135.
- Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41, 711–724. doi:10.1016/S0042-6989(00)00277-7.
- 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. *Current Biology*, 17, 336–340.
- Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, 54, 677–696. doi:10.1016/j.neuron.2007.05.019.
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, 41, 2261–2271.
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, 27, 649–677. doi:10.1146/annurev.neuro.27.070203.144220.
- Gu, Y., Wang, Y., & Wang, S. R. (2000). Regional variation in receptive field properties of tectal neurons in pigeons. *Brain, Behavior and Evolution*, 55, 221–228.
- Hayward, W. G. (1998). Effects of outline shape in object recognition. *Journal of Experimental Psychology. Human Perception and Performance*, 24, 427–440. doi:10.1037/0096-1523.24.2.427.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160, 106–154.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195, 215–243.
- Hummel, J. E., & Stankiewicz, B. J. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition*, 5, 49–79. doi:10.1080/713756775.
- Izawa, E.-I., Aoki, N., & Matsushima, T. (2005). Neural correlates of the proximity and quantity of anticipated food rewards in the ventral striatum of domestic chicks. *European Journal of Neuroscience*, 22, 1502–1512. doi:10.1111/j.1460-9568.2005.04311.x.
- Izawa, E., Zachar, G., Aoki, N., Koga, K., & Matsushima, T. (2002). Lesions of the ventro- medial basal ganglia impair the reinforcement but not the recall of memorized color discrimination in domestic chicks. *Behavioural Brain Research*, 136, 405–414.
- Jassik-Gerschenfeld, D., & Guichard, J. (1972). Visual receptive fields of single cells in the pigeon's optic tectum. *Brain Research*, 40, 303–317.
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, 53, 891–903. doi:10.1016/j.neuron.2007.02.015.
- Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58, 1233–1258.
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence*, 4, 237–285.
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315–341. doi:10.1146/annurev.neuro.23.1.315.
- Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience*, 22, 212–224. doi:10.1111/j.1460-9568.2005.04202.x.
- Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience*, 23, 3016–3027.

- Kirkpatrick, K. (2001). Object recognition. In R. G. Cook (Ed.), *Avian visual cognition*. Available at www.pigeon.psy.tufts.edu/avc/kirkpatrick/.
- Kirkpatrick-Steger, K., & Wasserman, E. A. (1996). The what and the where of the pigeon's processing of complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 60–67. doi:10.1037/0097-7403.22.1.60.
- Kirkpatrick-Steger, K., Wasserman, E. A., & Biederman, I. (1998). Effects of geon deletion, scrambling, and movement on picture recognition in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 24, 34–46. doi:10.1037/0097-7403.24.1.34.
- Kouh, M., & Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Computation*, 20, 1427–1451.
- Kriegeskorte, N. (2009). Relating population-code representations between man, monkey, and computational models. *Frontiers in Neuroscience*, 3, 363–373.
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, 65, 145–168.
- Lampl, I., Ferster, D., Poggio, T., & Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *Journal of Neurophysiology*, 92, 2704–2713. doi:10.1152/jn.00060.2004.
- Larsen, A., & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 1–20. doi:10.1037/0096-1523.4.1.1.
- Laverghetta, A. V., & Shimizu, T. (2003). Organization of the ectostriatum based on afferent connections in the zebra finch (*Taeniopygia guttata*). *Brain Research*, 963, 101–112.
- Lazareva, O. F., Wasserman, E. A., & Biederman, I. (2008). Pigeons and humans are more sensitive to nonaccidental than to metric changes in visual objects. *Behavioural Processes*, 77, 199–209.
- Li, D., Xiao, Q., & Wang, S. (2007). Feedforward construction of the receptive field and orientation selectivity of visual neurons in the pigeon. *Cerebral Cortex*, 17, 885–893.
- Lumsden, E. A. (1977). Generalization of an operant response to photographs and drawings/silhouettes of a three-dimensional object at various orientations. *Bulletin of the Psychonomic Society*, 10, 405–407.
- Macphail, E. M., & Bolhuis, J. J. (2001). The evolution of intelligence: Adaptive specializations versus general process. *Biological Reviews*, 76, 341–364.
- Martinez, L. M., & Alonso, J. M. (2003). Complex receptive fields in primary visual cortex. *The Neuroscientist*, 9, 317–331.
- Mash, C., Arterberry, M. E., & Bornstein, M. H. (2007). Mechanisms of visual object recognition in infancy: Five-month-olds generalize beyond the interpolation of familiar views. *Infancy*, 12(1), 31–43.
- Minamimoto, T., Saunders, R. C., & Richmond, B. J. (2010). Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron*, 66, 501–507.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, 6, 414–417. doi:10.1016/0166-2236(83)90190-X.
- Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16, 1936–1947.
- Morris, G., Nevet, A., Arkadir, D., Vaadia, E., & Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nature Neuroscience*, 9, 1057–1063.
- Nguyen, A. P., Spetch, M. L., Crowder, N. A., Winship, I. R., Hurd, P. L., & Wylie, D. R. W. (2004). A dissociation of motion and spatial-pattern vision in the avian telencephalon: Implications for the evolution of “visual streams. *Journal of Neuroscience*, 24, 4962–4970.
- Nicola, S. M., Surmeier, D. J., & Malenka, R. C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annual Review of Neuroscience*, 23, 185–215.
- Papini, M. R. (2002). Pattern and process in the evolution of learning. *Psychological Review*, 109, 186–201. doi:10.1037/0033-295X.109.1.186.
- Peissig, J. J., Kirkpatrick, K., Young, M. E., Wasserman, E. A., & Biederman, I. (2006). Effects of varying stimulus size on object recognition in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 32, 419–430. doi:10.1037/0097-7403.32.4.419.
- Peissig, J. J., Wasserman, E. A., Young, M. E., & Biederman, I. (2002). Learning an object from multiple views enhances its recognition in an orthogonal rotational axis in pigeons. *Vision Research*, 42, 2051–2062. doi:10.1016/S0042-6989(02)00128-1.
- 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. *Journal of Experimental Psychology: Animal Behavior Processes*, 26, 115–132. doi:10.1037/0097-7403.26.2.115.
- Perrett, D., & Oram, M. (1993). Neurophysiology of shape processing. *Image and Vision Computing*, 11, 317–333.
- Pisacreta, R., Potter, C., & Lefave, P. (1984). Matching of varying-size form stimuli in the pigeon. *Bulletin of the Psychonomic Society*, 22, 591–593.
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3-dimensional objects. *Nature*, 343, 263–266.
- Reid, R. C., & Usrey, W. M. (2004). Functional connectivity in the pathway from retina to striate cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 673–679). Cambridge, MA: MIT Press.
- Reiner, A. (2002). Functional circuitry of the avian basal ganglia: Implications for basal ganglia organization in stem amniotes. *Brain Research Bulletin*, 57, 513–528.
- Reiner, A., Yamamoto, K., & Karten, H. J. (2005). Organization and evolution of the avian forebrain. *The Anatomical Record. Part A, Discoveries in Molecular, Cellular, and Evolutionary Biology*, 287, 1080–1102. doi:10.1002/ar.a.20253.
- Reynolds, J. N. J., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, 15, 507–521. doi:10.1016/S0893-6080(02)00045-X.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025. doi:10.1038/14819.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3, 1199–1204. doi:10.1038/81479.
- Riley, D. A., & Langley, C. M. (1993). The logic of species comparisons. *Psychological Science*, 4, 185–189.
- Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, 10, 1615–1624. doi:10.1038/nn2013.
- Rolls, E. T., & Milward, T. (2000). A model of invariant object recognition in the visual system: Learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation*, 12, 2547–2572.
- Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: Two fundamental challenges for visual processing. *Current Opinion in Neurobiology*, 20, 382–388.
- Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Reviews*, 75, 107–154.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80, 1–27.
- Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 36, 241–263.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate for prediction and reward. *Science*, 275, 1593–1599.

- Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, *32*, 265–278.
- Serre, T. (2006). *Learning a dictionary of shape-components in visual cortex: Comparison with neurons, humans, and machines (CBCL Paper 260/MIT-CSAIL-TR 2006-028)*. Cambridge, MA: Massachusetts Institute of Technology.
- Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., & Poggio, T. (2005). *A theory of object recognition: Computations and circuits in the feedforward path of the ventral stream in primate visual cortex (MIT AI Memo 2005-036/CBCL)*. Cambridge, MA: Massachusetts Institute of Technology.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*, 6424–6429.
- Shettleworth, S. J. (1993). Where is the comparison in comparative cognition? Alternative research programs. *Psychological Science*, *4*, 179–184.
- Shettleworth, S. J. (2000). Modularity and the evolution of cognition. In C. M. Heyes & L. Huber (Eds.), *the evolution of cognition* (pp. 43–60). Cambridge, MA: MIT Press.
- Shettleworth, S. J. (2010a). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, *14*, 477–481. doi:10.1016/j.tics.2010.07.002.
- Shettleworth, S. J. (2010b). *Cognition, evolution, and behavior*. New York: Oxford University Press.
- Shimizu, T., & Bowers, A. N. (1999). Visual circuits of the avian telencephalon: Evolutionary implications. *Behavioural Brain Research*, *98*, 183–191.
- Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience and Biobehavioral Reviews*, *32*, 219–236.
- Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla–Wagner model. *Psychonomic Bulletin & Review*, *3*, 314–321.
- Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., et al. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review*, *18*, 414–421. doi:10.3758/s13423-010-0047-8.
- Soto, F. A., & Wasserman, E. A. (2010a). Error-driven learning in visual categorization and object recognition: A common-elements model. *Psychological Review*, *117*, 349–381. doi:10.1037/a0018695.
- Soto, F. A., & Wasserman, E. A. (2010b). Missing the forest for the trees: Object-discrimination learning blocks categorization learning. *Psychological Science*, *21*, 1510–1517. doi:10.1177/0956797610382125.
- Spetch, M. L., & Friedman, A. (2003). Recognizing rotated views of objects: Interpolation versus generalization by humans and pigeons. *Psychonomic Bulletin & Review*, *10*, 135–140.
- Spetch, M. L., & Friedman, A. (2006). Comparative cognition of object recognition. *Comparative Cognition & Behavior Reviews*, *1*, 12–35. doi:10.3819/ccbr.2008.10002.
- 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. *Journal of Experimental Psychology: General*, *130*, 238–255. doi:10.1037/0096-3445.130.2.238.
- Stevens, C. F. (2004). Preserving properties of object shape by computations in primary visual cortex. *Proceedings of the National Academy of Sciences*, *101*, 15524–15529.
- Tarr, M. J., Bühlhoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, *8*, 282–289. doi:10.1111/j.1467-9280.1997.tb00439.x.
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*, 233–282. doi:10.1016/0010-0285(89)90009-1.
- Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, *35*, 3053–3069.
- Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, *32*, 193–254.
- Ungerleider, L. G., & Haxby, J. V. (1994). “What” and “where” in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165.
- Van Hamme, L. J., Wasserman, E. A., & Biederman, I. (1992). Discrimination of contour-deleted images by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *18*, 387–399. doi:10.1037/0097-7403.18.4.387.
- Veenman, C. L., Wild, J. M., & Reiner, A. (1995). Organization of the avian “corticostriatal” projection system: A retrograde and anterograde pathway tracing study in pigeons. *The Journal of Comparative Neurology*, *354*, 87–126.
- Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *Journal of Cognitive Neuroscience*, *13*, 444–453.
- Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*, 43–48.
- Wang, Y. C., Jiang, S., & Frost, B. J. (1993). Visual processing in pigeon nucleus rotundus: Luminance, color, motion, and looming subdivisions. *Visual Neuroscience*, *10*, 21–30.
- Wasserman, E. A. (1993). Comparative cognition: Toward a general understanding of cognition in behavior. *Psychological Science*, *4*, 156–161.
- 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. *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 205–221.
- Wasserman, E. A., Kirkpatrick-Steger, K., Van Hamme, L. J., & Biederman, I. (1993). Pigeons are sensitive to the spatial organization of complex visual stimuli. *Psychological Science*, *4*, 336–341. doi:10.1111/j.1467-9280.1993.tb00575.x.
- Wersing, H., & Körner, E. (2003). Learning optimized features for hierarchical models of invariant object recognition. *Neural Computation*, *15*, 1559–1588.
- Wylie, D. R. W., Gutierrez-Ibanez, C., Pakan, J. M. P., & Iwaniuk, A. N. (2009). The optic tectum of birds: Mapping our way to understanding visual processing. *Canadian Journal of Experimental Psychology*, *63*, 328–338. doi:10.1037/a0016826.
- Yamane, Y., Tsunoda, K., Matsumoto, M., Phillips, A. N., & Tanifuji, M. (2006). Representation of the spatial relationship among object parts by neurons in macaque inferotemporal cortex. *Journal of Neurophysiology*, *96*, 3147–3156.
- Yanagihara, S., Izawa, E., Koga, K., & Matsushima, T. (2001). Reward-related neuronal activities in basal ganglia of domestic chicks. *Neuroreport*, *12*, 1431–1435.
- Young, M. E., Peissig, J. J., Wasserman, E. A., & Biederman, I. (2001). Discrimination of geons by pigeons: The effects of variations in surface depiction. *Animal Learning & Behavior*, *29*, 97–106.
- Yu, A. J., Giese, M. A., & Poggio, T. (2002). Biophysiological plausible implementations of the maximum operation. *Neural Computation*, *14*, 2857–2881.

Supplementary Material

The Hierarchical Model of Serre and Colleagues

The model of visual processing that we used for our simulations was directly taken from Serre, Oliva, and Poggio (2007). Those authors have provided MATLAB code which we used in our simulations and can be found at the following website:

<http://cbcl.mit.edu/software-datasets/index.html>

The original model consists of 9 hierarchically ordered layers of units: Layers S1, C1, S2, C2, S2b, S3, C2b, C3, and S4. Our simulations used only the first 8 layers. The final layer in the original model was part of a classification learning procedure, which we have replaced in our simulations with a reinforcement learning procedure. Each one of the layers in the model is organized in a number of feature maps, in which all of the units have the same selectivity, but their receptive fields are at different scales and positions. The selectivity of all of the units in the same feature map is represented by a vector of weights \mathbf{w} .

As in all hierarchical models of visual processing, this model extends the original functional architecture that was proposed by Hubel and Wiesel (1968), comprising two layers of units which are interspersed across the hierarchy. Units in the S layers carry out a feature selection operation, responding to features that result from a combination of inputs from units in the prior layer. Each simple unit receives inputs from n_{S_k} units from the prior layer and it has a preferred pattern of stimulation that is represented by the vector of weights \mathbf{w} . Given a vector \mathbf{x} representing the input received by the unit, its response y is a Gaussian-like tuning operation:

$$y = \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{n_{S_k}} (w_j - x_j)^2\right) \quad (1)$$

, where σ represents the sharpness of the tuning function and j indexes the synaptic inputs to the unit.

Units in the C layers carry out a pooling operation, combining inputs from n_{C_k} units in the prior layer tuned to the same feature over a range of positions and scales. Given a vector \mathbf{x} representing the input received by the unit, its response y is equal to the maximum value among its inputs:

$$y = \max_{j=1 \dots n_{C_k}} x_j \quad (2)$$

Layer S1. All images were resized to 256×256 pixels before being processed by the filters in Layer S1. The weight vector \mathbf{w} for S_1 units takes the form of a Gabor function, described by the following equation:

$$F(u_1, u_2) = \exp\left(-\frac{(u_1'^2 + \gamma^2 u_2'^2)}{2\sigma^2}\right) \cos\left(2\pi \frac{u_1'}{\lambda} + \psi\right) \quad (3)$$

, where (u_1, u_2) are 2D coordinates within the receptive field of the S_1 unit, and

$$u_1' = u_1 \cos \theta + u_2 \sin \theta$$

, and

$$u_2' = -u_1 \sin \theta + u_2 \cos \theta$$

The phase offset parameter, ψ , was set to 0 for all of the units and the aspect ratio γ was fixed to 0.3. There were 68 feature maps in Layer S1, obtained from the combination of 4 orientations θ (0° , 45° , 90° , and 135°) and 17 sizes of receptive fields (in pixels: 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37 and 39). Each feature map consisted of the repetition of the same feature detector in different positions on the image. Each pixel in the image was thus processed by a full set of 68 feature detectors. For each receptive field size, the values of the parameters λ and σ were obtained from the following *ad hoc* equations:

$$\sigma = 0.0036size^2 + 0.35size + 0.18 \quad (4)$$

$$\lambda = \frac{\sigma}{0.8} \quad (5)$$

, which were created by the authors to reproduce properties of V1 simple cells (see Serre et al., 2005).

The response of S1 units was given by Equation 1, where $w_j = F(u_1, u_2)$ was computed according to Equation 3 and x_j was the intensity value of pixel j in a small patch of the input image that fell within the unit's receptive field.

Layer C1. Units in C1 received their inputs from a grid of topologically organized S1 units with the same orientation selectivity; their response was computed according to Equation 2. To determine the inputs to C1 units, S1 units were organized in 8 bands according to their receptive field size. For each band, a grid of size $\Delta N_{C1}^S \times \Delta N_{C1}^S$ was built and a max was taken over the resulting grid and across all scales in the band. To improve processing time, C1 units were downsampled and computed every ϵ_{C1} S1 units. The bands and parameter values were:

Band 1: sizes 7 and 9, $\Delta N_{C1}^S=8$, $\epsilon_{C1}=3$.

Band 2: sizes 11 and 13, $\Delta N_{C1}^S=10$, $\epsilon_{C1}=5$.

Band 3: sizes 15 and 17, $\Delta N_{C1}^S=12$, $\epsilon_{C1}=7$.

Band 4: sizes 19 and 21, $\Delta N_{C1}^S=14$, $\epsilon_{C1}=8$.

Band 5: sizes 23 and 25, $\Delta N_{C1}^S=16$, $\epsilon_{C1}=10$.

Band 6: sizes 27 and 29, $\Delta N_{C1}^S=18$, $\epsilon_{C1}=12$.

Band 7: sizes 31 and 33, $\Delta N_{C1}^S=20$, $\epsilon_{C1}=13$.

Band 8: sizes 35, 37 and 39, $\Delta N_{C1}^S=22$, $\epsilon_{C1}=15$.

Layer S2. Each S2 unit received 10 inputs selected at random from a 3×3 neighborhood of C1 units at 4 possible orientations. The response of the unit was given by Equation 1, where \mathbf{w} is a vector of weights learned from natural images using a very simple procedure. During learning, the model was presented with a randomly selected image patch and the pattern of activation in the C1 inputs was set as its preferred pattern \mathbf{w} (see Serre, 2006). About 2,000 selectivities were learned this way and were replicated across the visual field and all 8 scale bands.

Layer S2b. The responses of S2b units were computed as those of S2 units, but from 100 inputs selected at random from an $\Delta N_{S2b} \times \Delta N_{S2b}^S$ neighborhood of C1 units at 4 possible orientations. There were 4 different neighborhood sizes ΔN_{S2b} : 6, 9, 12, and 15. There were about 500 learned selectivities at each neighborhood size, for a total of 2,000.

Layer C2. Units in C2 received their inputs from a grid of topologically organized S2 units with the same selectivity; their response was computed according to Equation 2. To determine the inputs to C2 units, S2 units were organized in 4 bands according to receptive field size, each created by the union of 2 C1 bands. For each band, a grid of size $\Delta N_{C2}^S \times \Delta N_{C2}^S$ was built and a max was taken over the resulting grid and across the two scales. To improve processing time, C2 units were downsampled and computed every ϵ_{C2} S2 units. The bands and parameter values were the following:

Band 1: C1 bands 1 and 2, $\Delta N_{C2}^S=8$, $\epsilon_{C1}=3$.

Band 2: C1 bands 3 and 4, $\Delta N_{C2}^S=12$, $\epsilon_{C1}=7$.

Band 3: C1 bands 5 and 6, $\Delta N_{C2}^S=16$, $\epsilon_{C1}=10$.

Band 4: C1 bands 7 and 8, $\Delta N_{C2}^S=20$, $\epsilon_{C1}=13$.

Layer C2b. The responses of C3 units were computed as those of C2 units, but with a receptive field of the size of the whole stimulus, obtained by pooling all C1 bands together. The grid size $\Delta N_{C_3}^S$ was equal to 40.

Layer S3. The responses of S3 units were computed as those of S2 units, randomly selecting 100 inputs from a 3×3 neighborhood of C1 units at all possible S2 selectivities. Again, there were a total of 2,000 learned selectivities.

Layer C3. The responses of C3 units were computed as those of C2 units, but with a receptive field of the size of the whole stimulus, obtained by pooling all C1 bands together. The grid size $\Delta N_{C_3}^S$ was equal to 40.

The Ideal Observer Model.

The ideal observer model used in our simulations was developed by Tjan et al. (1995) and it is often used as a benchmark in visual categorization studies. The model allows one to determine which aspects of an experimental outcome can be explained as resulting directly from task and stimulus demands and not from any visual processing carried out by the experimental subjects. This model assumes an observer that knows all possible training stimuli and their assigned responses *a priori*. The task of the observer is to choose the object O_i with the highest *a posteriori* probability given the presentation of an image R ; that is, it must choose a response that maximizes $p(O_i|R)$. There are a number of templates (e.g., views) of O_i that are known to the observer, represented by T_{ij} , where j indexes the template. The probability of object O_i being presented is the sum of the probabilities of each of the individual templates being presented; according to Bayes' rule this probability is:

$$p(O_i|R) = \sum_j p(T_{ij}|R) = \sum_j \frac{p(R|T_{ij})p(T_{ij})}{p(R)} \quad (6)$$

If the images T_{ij} and R are vectors of pixels, each with a luminance value, and under the assumption that image R is generated from the template of a single object, perturbed by the addition of Gaussian noise with zero mean and standard deviation σ , then Tjan et al. (1995) have shown that maximizing Equation 6 is equivalent to maximizing the following:

$$L'(i) = \sum_j \exp\left(-\frac{1}{2\sigma^2}\|R - T_{ij}\|^2\right)p(T_{ij}) \quad (7)$$

That is, the optimal strategy is to compute the sum of a similarity measure between R and each template T_{ij} , weighted by the probability of template T_{ij} .

Note that this ideal observer assumes a visual task in which there is inherent noise in the signal that serves as input to the pigeons' visual system. Uncertainty, in the form of normally distributed noise, is assumed to be an important part of the visual environment. Our simulations thus required adding Gaussian noise with zero mean to the images that were presented to the ideal observer. The value of σ was adjusted to bring the performance of the ideal observer to the level that was exhibited by the subjects in each experiment. This process allowed us to determine what performance we would expect from the birds if they were basing their responding on pixel similarities among the visual stimuli and their suboptimal performance with the training stimuli was due to noise in the signal. This adjustment of σ was performed using QUEST (Watson & Pelli, 1983), an adaptive psychometric procedure that is useful to find the value of a stimulus variable that produces a particular performance value (e.g., 85% correct). The noise statistics were available to the ideal observer, which compared the resulting image against all of the training images included in the task, giving as output the category label with the highest a

posteriori probability computed according to Equation 7. In trials with novel test stimuli, the category label was computed based on knowledge of the training images alone.

Supplementary Table 1

Standard error for all mean values obtained from simulations with the general processes model and reported in the main article.

Experiment; Figure	Condition	Standard Error
Spetch et al. (2001); Figure 4B	Group 0-part	
	Rotation 0°	.0102
	Rotation 30°	.0311
	Rotation 90°	.0114
	Rotation 135°	.0226
	Rotation 160°	.0185
	Rotation 225°	.0326
	Group 1-part	
	Rotation 0°	.0099
	Rotation 30°	.0218
	Rotation 90°	.0092
	Rotation 135°	.0207
	Rotation 160°	.0184
	Rotation 225°	.0282
	Group 3-part	
	Rotation 0°	.0115
	Rotation 30°	.0218
	Rotation 90°	.0068
	Rotation 135°	.0235
	Rotation 160°	.0213
	Rotation 225°	.0267
Group 5-part		
Rotation 0°	.0101	
Rotation 30°	.0176	
Rotation 90°	.0077	
Rotation 135°	.0211	
Rotation 160°	.0208	
Rotation 225°	.0251	
Peissig et al. (2002); Figure 5	Group Five Views	
	Rotation 0°	.0131

	Rotation 36°	.0155
	Rotation 72°	.0154
	Rotation 108°	.0242
	Rotation 144°	.0137
	Rotation 180°	.0080
	Rotation -144°	.0203
	Rotation -108°	.0171
	Rotation -72°	.0185
	Rotation -36°	.0142
	Group One View	
	Rotation 0°	.0099
	Rotation 36°	.0164
	Rotation 72°	.0183
	Rotation 108°	.0208
	Rotation 144°	.0120
	Rotation 180°	.0237
	Rotation -144°	.0175
	Rotation -108°	.0139
	Rotation -72°	.0205
	Rotation -36°	.0106
<hr/>		
Gibson et al. (2007); Figure 6B	Cotermination	.0254
	Edge	.0093
	Shading	.0098
<hr/>		
Van Hamme et al. (1992, Exp. 2); Figure 7B	Original	.0063
	Complementary	.0539
	Scrambled Complementary	.0349
<hr/>		
Van Hamme et al. (1992, Exp. 3); Figure 7C	Original	.0064
	Scrambled Mild	.0460*
	Scrambled Heavy	.0264*
<hr/>		
Young et al. (2001); Figure 8B	Training Stimulus	.0018
	Silhouette	.0143
	Light Change	.0132
	Line Drawing	.0156
<hr/>		
Peissig et al. (2006); Figure 9B	Exponential Size Transformation	

Size -3	.0279
Size -2	.0163
Size -1	.0231
Size 0	.0159
Size 1	.0201
Size 2	.0105
Size 3	.0371

Linear Size Transformation

Size -3	.0117
Size -2	.0156
Size -1	.0110
Size 0	.0197
Size 1	.0183
Size 2	.0154
Size 3	.0127

* These standard errors are large relative to the small difference in the mean proportion of correct responses between the Scrambled Mild and Scrambled Heavy conditions. A larger set of 30 simulations confirmed the mean difference between conditions ($M_{\text{Mild}} = .44$; $M_{\text{Heavy}} = .39$) and resulted in smaller standard errors ($SE_{\text{Mild}} = .0198$; $SE_{\text{Heavy}} = .0164$).