Orthographic processing in pigeons (Columbia livia)

Damian Scarfa,b,1, Karoline Boyb, Anelisie Uber Reinertb, Jack Devinec, Onur Güntürkünb, and Michael Colomboa,b,1

aDepartment of Psychology, University of Otago, Dunedin 9054, New Zealand; bDepartment of Psychology, Institute of Cognitive Neuroscience, Biopsychology, Ruhr-University Bochum, D-44780 Bochum, Germany; and cDepartment of Physics, University of Otago, Dunedin 9054, New Zealand

Edited by Dale Purves, Duke University, Durham, NC, and approved July 29, 2016 (received for review May 17, 2016)

Learning to read involves the acquisition of letter–sound relationships (i.e., decoding skills) and the ability to visually recognize words (i.e., orthographic knowledge). Although decoding skills are clearly human-unique, given they are seated in language, recent research and theory suggest that orthographic processing may derive from the exaptation or recycling of visual circuits that evolved to recognize everyday objects and shapes in our natural environment. An open question is whether orthographic processing is limited to visual circuits that are similar to our own or a product of plasticity common to many vertebrate visual systems.

Here we show that pigeons, organisms that separated from humans more than 300 million y ago, process words orthographically. Specifically, we demonstrate that pigeons trained to discriminate words from nonwords picked up on the orthographic properties that define words and used this knowledge to identify words they had never seen before. In addition, the pigeons were sensitive to the bigram frequencies of words (i.e., the common co-occurrence of certain letter pairs), the edit distance between nonwords and words, and the internal structure of words. Our findings demonstrate that visual systems organizationally distinct from the primate visual system can also be exapted or recycled to process the visual word form.

On the surface, the human brain seems to have evolved for reading (1). Across individuals and cultures (2), reading activates an identical area in the left lateral occipitotemporal sulcus known as the visual word form area (VWFA) (3). This activation occurs no matter the case or font of the script (4), it increases in the transition from being illiterate to literate (5), and it increases with improvements in reading fluency (6). However, the presence of a VWFA is difficult to assimilate with the fact that writing was invented merely ~5,400 y ago, and only became widespread very recently in human history, making it impossible that an area of the human brain evolved specifically for reading (7).

Without the time to evolve, how can we explain the presence of the VWFA? One intriguing possibility is that the VWFA is the product of neuronal recycling, with its neurons learning to code visual stimuli (i.e., words) that greatly differ from the visual objects it initially evolved to code (8, 9).

Anatomically, the VWFA lies just downstream of the ventral visual (i.e., what) pathway, a hierarchy of areas critical to visual object and face recognition in human and nonhuman primates. Dehaene et al. (10) argue that this hierarchy of areas can be tuned to the visual word form, with areas lower in the hierarchy simply encoding letter identities, and areas in the upper echelons of the hierarchy, specifically the VWFA, tuned to the co-occurrence of letter pairs (i.e., bigrams) and letter strings. At the present time, there are no neurophysiological studies that have investigated whether neurons in the nonhuman primate temporal lobe can learn to code anything beyond individual alphabetic characters and symbols (11). However, a recent behavioral study demonstrated for the first time that nonhuman primates are sensitive to the statistical properties of words and can use these properties to distinguish them from strings of letters that are not words (12), an ability previously thought to be unique to humans, and one for which the VWFA is purportedly critical.

An open question is whether animals with brain architectures and visual systems dissimilar to primates also display this sensitivity to the statistical properties of words. Indeed, the recycling hypothesis has been built with our brain in mind and, more specifically, the hierarchical organization of the ventral visual system. In the absence of data from nonprimate species, however, we have no idea whether a primate brain is a prerequisite for orthographic processing. To answer this question, in the current study we assess orthographic processing in pigeons, an organism whose brain architecture (13) and visual system (14) are very different from humans. Astonishingly, we find that pigeons display every hallmark of orthographic processing displayed by Grainger et al.’s baboons (12, 15).

Results

The pigeons were able to discriminate between an increasing number of four-letter words and the 7,832 four-letter combinations that only resembled words (i.e., nonwords). Specifically, over the course of training, the pigeons learned to discriminate between 26 and 58 words (Q32: 26 words, Q35: 58 words, Q41: 32 words, and Q43: 57 words). Although possessing a smaller vocabulary than the baboons (mean number of words learnt: baboons, 139 vs. pigeons, 43), the pigeons behaved identically to the baboons when presented with novel words, in that they made significantly fewer nonword responses to the novel words than to the nonwords [t(3) = 4.91; P < 0.05] and classified the novel words as words at a level significantly above chance (50%) [t(3) = 3.15; P < 0.05 (one-tail)]. At a minimum, this transfer suggests that orthographic processing in pigeons, an organism whose brain architecture (13) and visual system (14) are significantly different from that of primates, code words orthographically.

Significance

A novel theory suggests that orthographic processing is the product of neuronal recycling, with visual circuits that evolved to code visual objects now co-opted to code words. Here, we provide a litmus test of this theory by assessing whether pigeons, an organism with a visual system organizationally distinct from that of primates, code words orthographically. Pigeons not only correctly identified novel words but also display the hallmarks of orthographic processing, in that they are sensitive to the bigram frequencies of words, the orthographic similarity between words and nonwords, and the transposition of letters. These findings demonstrate that visual systems neither genetically nor organizationally similar to humans can be recycled to represent the orthographic code that defines words.

Author contributions: D.S., K.B., A.U.R., O.G., and M.C. designed research; K.B., A.U.R., and M.C. performed research; D.S., J.D., and M.C. analyzed data; and D.S. and M.C. wrote the paper.

The authors declare no conflict of interest.

1To whom correspondence may be addressed. Email: damian@psy.otago.ac.nz or colombo@psy.otago.ac.nz.

Published under a CC BY 4.0 license.

11272–11276 | PNAS | October 4, 2016 | vol. 113 | no. 40

www.pnas.org/cgi/doi/10.1073/pnas.1607870113
Values < 0.05). That is, the more frequent a bigram was in a pigeon's vocabulary, the better the pigeon performed on words that contained that bigram. In humans, sensitivity to bigrams is one of the bases on which reading is built (16).

With respect to nonwords, pigeons' accuracy was directly related to a nonword's orthographic similarity to known words (Fig. 1A) (Q32: $r^2 = 0.68$, Q35: $r^2 = 0.64$, Q41: $r^2 = 0.93$, and Q43: $r^2 = 0.92$; all $P$ values < 0.05). Orthographic similarity was determined both by calculating edit distance from the Levenshtein (17) distance (i.e., the number of letter insertions, deletions, and substitutions required to transform a nonword into a known word) and by averaging the 20 lowest edit distance values to derive a single OLD20 (orthographic Levenshtein distance 20) (18) value for each nonword. In essence, the higher the OLD20 value, the more changes that would have to be made to transform the nonword into a known word. The pigeons' performance parallels that displayed by humans (19) (Fig. 1B) and baboons (12) (Fig. 1C).

Finally, we assessed whether the pigeons showed the transposed-letter effect, another hallmark of orthographic processing (20). The transposed-letter effect refers to the finding that nonwords created by transposing adjacent letters in a word (e.g., “very” transposed to “vrey”) are often misclassified as words (21). Indeed, transposed-letter effects are generally only observed in children who have acquired some literacy skills (22). That is, the more frequent a bigram was in a word's vocabulary, the better the pigeon performed on words differently from chance [$t(3) = 1.78; P = 0.17$], and to substituted words significantly below chance [$t(3) = 8.02; P < 0.05$]. On this measure, the pigeons' performance is actually more comparable to that of literate humans (23) (Fig. 2B) than the baboons' performance (15, 24) (Fig. 2C). Indeed, pigeons' differential performance on known words and transposed words suggests they were highly sensitive to the relative position of the letters within words, consistent with work demonstrating their ability to extract ordinal knowledge from visual sequences (25–29). In addition, there was also a strong correlation between the strength of the transposed-letter effect and the pigeons' performance on words ($r^2 = 0.96; P < 0.05$), but not between the transposed-letter effect and the total number of words learnt ($r^2 = 0.19; P = 0.57$) or performance on nonwords ($r^2 = 0.64; P = 0.20$). Again, this finding parallels the baboons' performance (15) and is consistent with human data demonstrating that the stronger words are encoded, the stronger the transposed-letter effect (30).

Discussion

Our research demonstrates that orthographic processing is not limited to primates. Although pigeons acquired fewer words than Grainger et al.'s (12) baboons, they were able to correctly classify novel words and were sensitive to the bigram frequency of words, the orthographic distance of nonwords, and the relative position of letters within a word. What mechanism underlies this remarkable performance? One possibility is that the pigeons and baboons simply memorized the words and used this as the basis for responding. Aspects of the current data, however, are inconsistent with this view. Indeed, the pigeons' and baboons' ability to correctly classify novel words suggests their performance was not based on the rote memorization of words. Of course, one might argue that, rather than being based on the memorization of words, their performance was based on the memorization of the nonwords. However, although baboons' and pigeons' long-term memory capacities are impressive, the number of nonwords in the current experiment (7,832) is well above their respective capacity limits (31). Further, even if the pigeons and baboons could memorize all 7,832 nonwords, it is difficult to

![Fig. 1. Performance on nonwords as a function of a nonword's orthographic similarity to known words for pigeons (A), humans (B), and baboons (C). Note that, as denoted by the y axis ranges, although the pigeons and baboons display a comparable range of scores, the human data cluster within a much tighter accuracy range. Error bars represent 95% confidence intervals. B and C reprinted with permission from ref. 12.](image-url)
see how the memorization account could explain the bigram frequency, OLD20, and transposed-letter data.

A promising alternative mechanism to memorization, one that could account for both the baboons’ and pigeons’ data, is conceptualization (32). The acquisition of a concept can be distilled down to the ability to generalize within classes of stimuli and discriminate between classes of stimuli (33). The pigeons’ and baboons’ ability to acquire words clearly demonstrates that they were able to discriminate between two classes of stimuli (i.e., words and nonwords), and their performance with novel words demonstrates their ability to generalize these concepts (34). Critically, the conceptualization account can also explain the bigram frequency, OLD20, and transposed-letter data, which, in essence, are all measures of how closely stimuli within one class or category resemble one another (e.g., the bigram frequencies of words) and how closely they approximate stimuli in a different class (e.g., the orthographic similarity of nonwords to words). In this context, our data support the view that pigeons’ conceptualization abilities make them an ideal animal model with which to investigate the early stages of human word learning (35).

Conclusions
The current findings demonstrate that the pigeon’s visual system can represent things beyond individual objects or symbols (36, 37) and code the statistical properties of letter strings. The fact that our pigeons’ performance was indistinguishable from that of the baboons across four markers of orthographic processing strongly suggests that the ability to process orthographic information is not limited to the primate brain. Our findings add to a growing body of work demonstrating that birds are ideal models with which to investigate the origins of language (38–40). At the level of the neuronal recycling hypothesis, our findings may represent the most powerful evidence of Dehaene’s (1) thesis: that neurons in a visual system neither genetically nor organizationally similar to humans can not only code words but also the statistical properties that define them.

Methods
Subjects. This research was approved by the University of Otago Animal Ethics Committee. Eighteen experimentally naive pigeons (Columba livia) were initially trained on the paradigm. After ~8 mo of training, the 18 birds were reduced to the four subjects that demonstrated the greatest aptitude for discriminating words from nonwords. By this time, the selected subjects had acquired a mean of 14 words (Q32: 8 words, Q35: 23 words, Q41: 9 words, and Q43: 15 words), whereas the excluded subjects had acquired a mean of just 4 words. This selection stage was conducted for two reasons. First, building a decent vocabulary is a prerequisite for investigating orthographic processing, and therefore we selected the subjects that had the greatest potential for acquiring a large vocabulary within a reasonable time. Second, on a practical level, with four testing chambers and other planned experiments, we simply did not have the resources required to run the entire sample of subjects daily beyond the 8-mo period.

Apparatus and Stimuli. Subjects were trained in one of four standard operant chambers. The front wall of each chamber housed a Perspex panel with five apertures. The center square aperture measured 3.3 \times 2 \text{ cm} and was encircled by four circular apertures, each 2.5 \text{ cm} in diameter. The center-to-center measurement was 5 \text{ cm} for the left and right circular apertures and 2.75 \text{ cm} for the upper and lower apertures. Only the center square and the upper and lower circle apertures were used in the current study. Sitting behind the Perspex panel was a 17-inch computer monitor used to display the stimuli. Positioned between the Perspex panel and computer monitor was a 17-inch touch frame used to record subject’s responses. Wheat was made available via a food hopper located at the front of the box, 21 \text{ cm} below the center square aperture. A ventilation fan was housed in the rear of each chamber and provided background noise of 80 dB to mask all extraneous noise.

The word and nonword stimuli consisted of letter strings in Arial 12-point font bold. Words were drawn from the pool of 308 words acquired by Grainger et al.’s (12) baboons. Similarly, the nonword stimuli consisted of the 7,832 stimuli used by Grainger et al. (12). A black eight-point star used for nonword responses was 1.5 \text{ cm} in diameter.

Procedure. All subjects were first trained to eat from the hopper. After hopper training, an autoshaping procedure was used until subjects were consistently pecking stimuli presented in any of the three apertures. After shaping, subjects were presented with their first word.
To reach criterion on a word, a subject had to perform at ≥66% on the old word and the nonwords across two consecutive sessions. In addition, on the second criterial day, a subject needed to perform ≥66% on the old word. To ensure that the performance on old words was maintained throughout the experiment, a retraining procedure was used after a subject’s performance on old words fell below 66% across 4 consecutive days. For retraining, a session consisted of 50 trials on the old word in which the subject had performed most poorly on across the 4 days and 50 nonword trials. Similar to the initial criterion, when a subject performed at ≥66% on the old word and the nonwords, they were transferred back to the standard training procedure.

**Novel Word Test.** The novel word test consisted of a single session in which subjects were presented with 50 novel words and 50 nonwords. At the time of the novel word test, Q32 was on word 25, Q35 on word 45, Q41 on word 45, and Q43 on word 45.

**Transposition Test.** The transposition test was conducted across several sessions and occurred several weeks after the novel word test. The session was identical to a standard training session, with the exception that either two (Q32 and Q41) or four (Q35 and Q43) probe trials were presented randomly within the session. Half of the probe trials consisted of a transposed word, and half consisted of substituted words. Transposed words were created by simply transposing the two middle characters in a known word. Words with identical middle characters were excluded. With respect to substituted words, any consonant and vowel in the second or third letter position was replaced with a letter of the same type (i.e., vowel or consonant) drawn at random from a list of all available vowels and consonants. At the time of the transposition test, Q32 was on word 26, Q35 on word 58, Q41 on word 32, and Q43 on word 57.

**Analysis.** The correlations between word bigram frequencies and word accuracy and between nonword OLD20 values and nonword accuracy are based on all trials after pigeons acquiring their 20th word. We used this approach, rather than Grainger et al.’s (12) approach of using the 20,000th trial onward, to ensure the pigeons had acquired the minimum number of words required for the calculation of a nonword OLD20 value and to ensure all subjects had a minimum vocabulary of 20 words. This approach provided us with an average of 7,728 word trials (Q32: 2,808 trials, Q35: 12,837 trials, Q41: 11,309 trials, and Q43: 12,039 nonword trials (Q32: 5,400 trials, Q35: 18,990 trials, Q41: 6,850 trials, and Q43: 16,916 trials) to analyze.

For bigram frequencies, following Grainger et al. (12), we calculated the frequency of adjacent letter pairs (i.e., letter 1 and letter 2, letter 2 and letter 3, and letter 3 and letter 4) in the pool of words a pigeon had learned and used the mean of these three values to assign a bigram frequency to a word. The frequency was then rounded down to two decimal places, and words were grouped into bins between 0 and 1 with a step size of 0.01. Finally, the accuracy at each step was then correlated with the mean bigram frequency of the step. The OLD20 values were treated in a similar manner with the exception that they were grouped into bins between 0 and 4 with a step size of 0.1. For both the bigram frequency and OLD20 correlations, if a step had fewer than 20 trials, it was excluded from the analysis.

**ACKNOWLEDGMENTS.** We thank C. Bell and C. Watson for assistance with data collection; and M. Beran, V.P. Bingman, H. Hayne, J. Hunter, K. Imuta, A. Knott, T. Ruffman, and M.L. Platt for reading earlier drafts of the manuscript. This work was supported by University of Otago Department of Psychology research funds (to M.C.) and Deutsche Forschungsgemeinschaft Grant SFB874 (to O.G.).


