Non-Human Recognition of Orthography: How is it implemented and how does it differ from Human orthographic processing

Benjamin Gagl[a,b,1], Ivonne Weyers[b], Susanne Eisenhauer[c], Christian J. Fiebach[d], Michael Colombo[e], Damian Scarf[e], Johannes C. Ziegler[f], Jonathan Grainger[f], Onur Güntürkün[g], and Jutta L. Mueller[b,h]

The ability to robustly recognize strings of letters, a cornerstone of reading, was observed in Baboons and Pigeons despite their lack of phonological and semantic knowledge. Here, we apply a comparative modeling approach to investigate the neuro-cognitive basis of Human, Baboon, and Pigeon orthographic decision behavior, addressing whether phylogenetic relatedness entails similar underlying neuro-cognitive phenotypes. We use the highly transparent Speechless Reader Model (SLR), which assumes letter string recognition based on widely accepted computational principles of predictive coding so that orthographic decisions rely on a prediction error signal emerging from multiple, hierarchically ordered representational levels, i.e., low-level visual, letter, or letter sequence representations. We investigate which representations species use during successful orthographic decision-making. We introduce multiple SLR variants, each including one or multiple prediction error representations, and compare the simulations of each SLR variant to the orthographic decisions from individuals of three species after learning letter strings without meaning. Humans predominantly relied on letter-sequence-level representations, resulting in the highest task performance in behavior and model simulations. Baboons also relied on sequence-based representations but in combination with pixel- and letter-level representations. In contrast, all Pigeons relied on pixel-level representations, partly in combination with letter- and letter-sequence-level representations. These findings suggest that orthographic representations utilized in orthographic decisions reflect the phylogenetic distance between species: Humans and Baboons use more similar representations compared to Pigeons. Overall, the description of orthographic decisions based on a small set of representations and computations was highly successful in describing behavior, even for Humans who mastered reading in its entirety.

Predictive coding | Computational modelling | Individual differences | Humans | Baboons | Pigeons

Significance Statement

Imagine being able to read without ever learning the alphabet. Research has shown that baboons and pigeons can exhibit reading-like behavior, suggesting shared processes across the species involved. To increase our understanding of the similarities and differences between humans and animals in reading-like behavior, we use a computational model to uncover the underlying processes that enable humans, baboons, and pigeons to perform these tasks. We found that humans and baboons rely on similar processes, focusing on information related to letters and letter sequences. In contrast, pigeons rely more heavily on visual cues. This discovery sheds light on the evolution of processes underlying reading and reading-like behavior, indicating that the lower the evolutionary distance between species, the more similar processes are involved.

Script is a hallmark of Human cultural evolution and is unmatched in the animal kingdom (1). Reading of script is considered exclusive to Humans, requiring explicit instruction and specialized cognitive skills to retrieve word representations and meaning (2–5). Biological foundations of reading ability have been described, for instance, in the finding of a visual word form area in the Human left-ventral occipito-temporal cortex (1) or of genetic risk for dyslexia (6). The description of the biological basis of reading and the description of reading-like behavior for Baboons (7) and Pigeons (8) raises the question of the evolutionary roots of reading ability. Both species learned to distinguish remembered orthographic stimuli (i.e., similar to written words) from novel letter combinations in orthographic decision tasks. While this suggests similar skills in non-human animals and Humans at the surface, how the task is solved across species, i.e., which cognitive processes were involved, remains an open question. Note that computer vision models, with drastically different architectures, can also solve similar tasks (9–12), indicating that there are many ways to solve orthographic processing tasks. In the current study, we investigate the cognitive basis of reading(-like) behavior (i.e., orthographic decisions) in Humans, Baboons, and Pigeons and test if phylogenetic proximity leads to more human-like cognitive processes in primates than in birds utilizing a transparent computational model (13) that allows one to differentiate the individual neurocognitive representations underlying orthographic decisions.

Humans typically learn to read through grapheme-phoneme conversion only once robust phonological and semantic representations are already in place (14–16), thereby aligning their reading rate to that of speech in fluent readers

Author affiliations: [a]Self-learning Systems Lab, University of Cologne, Cologne, Germany ; [b]Department of Linguistics, University of Vienna, Vienna, Austria; [c]Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain; [d]Department of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany; [e]Department of Psychology, University of Otago, Dunedin, New Zealand; [f]Aix-Marseille University, CNRS, Aix-en-Provence, France; [g]Department of Biopsychology, Ruhr University Bochum, Bochum, Germany; [h]Vienna Cognitive Science Research Hub, Vienna, Austria

Author contributions: B.G., I.W., and J.M. designed research; B.G. performed research; B.G., S.E., C.F., M.C., D.S., J.Z., J.G., and O.G. contributed data; K.M.C., B.G. analyzed data and implemented the model; and B.G., I.W., O.G., and J.M. wrote the paper. All authors commented on the final draft.

No competing interests here.

[1] Correspondence to benjamin.gagl@uni-koeln.de

(17). On the neuronal level, efficient reading is characterized by the implementation of a lexical categorization process (in the left-ventral occipito-temporal cortex, Ref. (18, 19); in proximity to the so-called visual word form area; Ref. (1)) that differentiates between word and non-word letter strings based on orthographic word characteristics (20). Behavioral tasks based on the distinction between words and non-words, like the lexical decision task, are easily solved by Humans as they can rely on a large body of orthographic knowledge and a lexicon of rich phonological and lexico-semantic representations (2–5). Lexical decision tasks are standard repertoire to study reading and visual word recognition (e.g., Ref. (21, 22)). Hence, similar orthographic decision tasks (i.e., without the involvement of explicit semantic processing) have been implemented in Baboons (7), Pigeons (8), Humans (23) and Computer vision models (9, 10, 12).

Here, we investigate the representations underlying orthographic decisions in Humans, Baboons, and Pigeons. This phenotyping approach (i) describes which neuro-cognitive operations are involved in making orthographic decisions, (ii) allows a comparison of the types of representation implemented across species, and (iii) across individuals (see Ref. (24–26) for examples of using computational models for cognitive or neuro-congitive phenotyping). We implement this based on a transparent computational model, the Speechless Reader Model (13), that allows us to infer the representations used in orthographic decisions on the level of vision and orthography as opposed to rather nontransparent computer vision models (9, 10), in the which the high model complexity only allows initial, but highly valuable, post-hoc investigations of the internal workings of the model (e.g., see Ref. (10)). Also, the speechless reader outperforms computer vision models when modeling the learning curve of Baboons (13). Hence, comparisons between species that vary in phylogenetical proximity allow us to investigate the evolutionary roots of reading-like behavior based on neuro-cognitive phenotypes.

The Speechless Reader Model (SLR)

The SLR is based on neuro-cognitive assumptions of visual and orthographic processing in Human visual word recognition (13, 18, 27) and the principles of predictive coding (28–30). The implementation without phonological and semantic representations serves the purpose of modeling animal orthographic decisions, as animals cannot access this information. For Humans, the SLR allows us to estimate how well one can describe orthographic decisions based on visual-orthographic information. Therefore, the SLR enables us to assess the influence of phonological and semantic knowledge even when not needed for the task.

The core of the model are three types of prediction error representations, which integrate the sensory input (i.e., the visually presented word or non-word) with the knowledge stored in the lexicon of each participant. For example, the orthographic prediction error (oPE; Ref. (12, 27)) integrates word knowledge and sensory input on the pixel level. Interestingly, the oPE, although implemented based on purely visual pixel-level information, is similar to orthographic word characteristics (i.e., that describe the letter combinatorics of a word like the Orthographic Levenshtein distance; Ref. (20); r

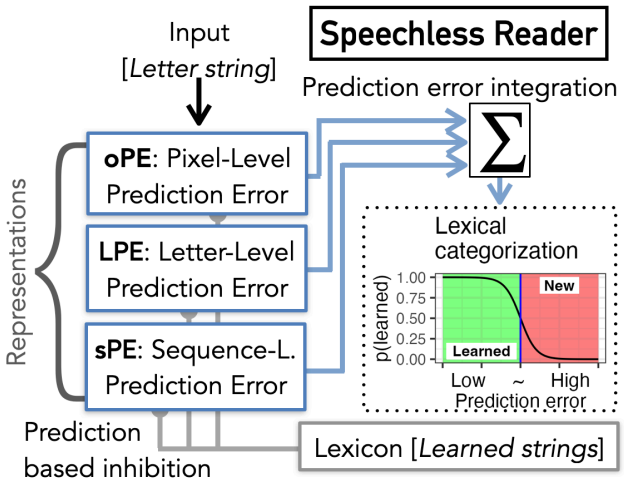

Fig. 1. *Speechless Reader Model*, including model input and three types of prediction error representations based on: visual-orthographic representations (oPE; Ref. (12, 27)), letter-position representations (LPE), and letter-sequence representations (sPE; Ref. (13)). Prediction errors are calculated based on the learned letter strings stored in the lexicon, the resulting representations are summed into a word-likeness estimate (18). The output of the categorization process is binary, indicating if the input is considered learned or new.

across languages from .25 to .4, Ref. (27)). Thus, this simple pixel-level operation produces more than a pure vision-related word characteristic as more abstract information emerges.

In the current version of the SLR, the prediction error representations operate on three distinct levels (see Fig. 1): pixel (orthographic prediction error: oPE; Ref. (12, 27)), letter (LPE), and letter-sequence (sPE; Ref. (13)). In other words, for an input letter string such as *WIND*, predictions from the lexicon convey the probability of a pixel, letter, or letter sequence at a specific position based on the previously learned letter strings in the lexicon (e.g., for LPE: how probable is the letter "W" at the initial position given the stored items in the lexicon, similar to a conditional probability of the letter at a specific position given a lexicon). In the full model, prediction error values are calculated for all three levels, which are summed up to generate a word-likeness estimate for each encountered letter string. This estimate then forms the basis for the orthographic decision that determines the model output: is a letter string learned or unknown (see Eq. 1, 2, 3, 4, 5, and 6).

Here, we use the SLR to model three datasets: (i) Human data from a pseudoword learning task, in which participants learned letter strings without explicit semantic associations (23, 31), (ii) Baboon data (7) and (iii) Pigeon data (8). We created an individualized lexicon comprising the letter strings learned successfully for each Human (N = 37), Baboon (N = 6), and Pigeon (N = 4). We then generated model simulations for each participant using each prediction error representation and possible combination of these representations (resulting in $N = 7$ model variants; e.g., combining the oPE and the LPE only in one variant or using only the sPE in another variant). In addition, we used individual categorization thresholds (cf. blue line in Fig. 1 and 6). Then, we compared the model and participant performance for each participant (see Appendix Fig. S1-3), and the simplest model with the lowest mean squared error was considered the best-fitting model. For the

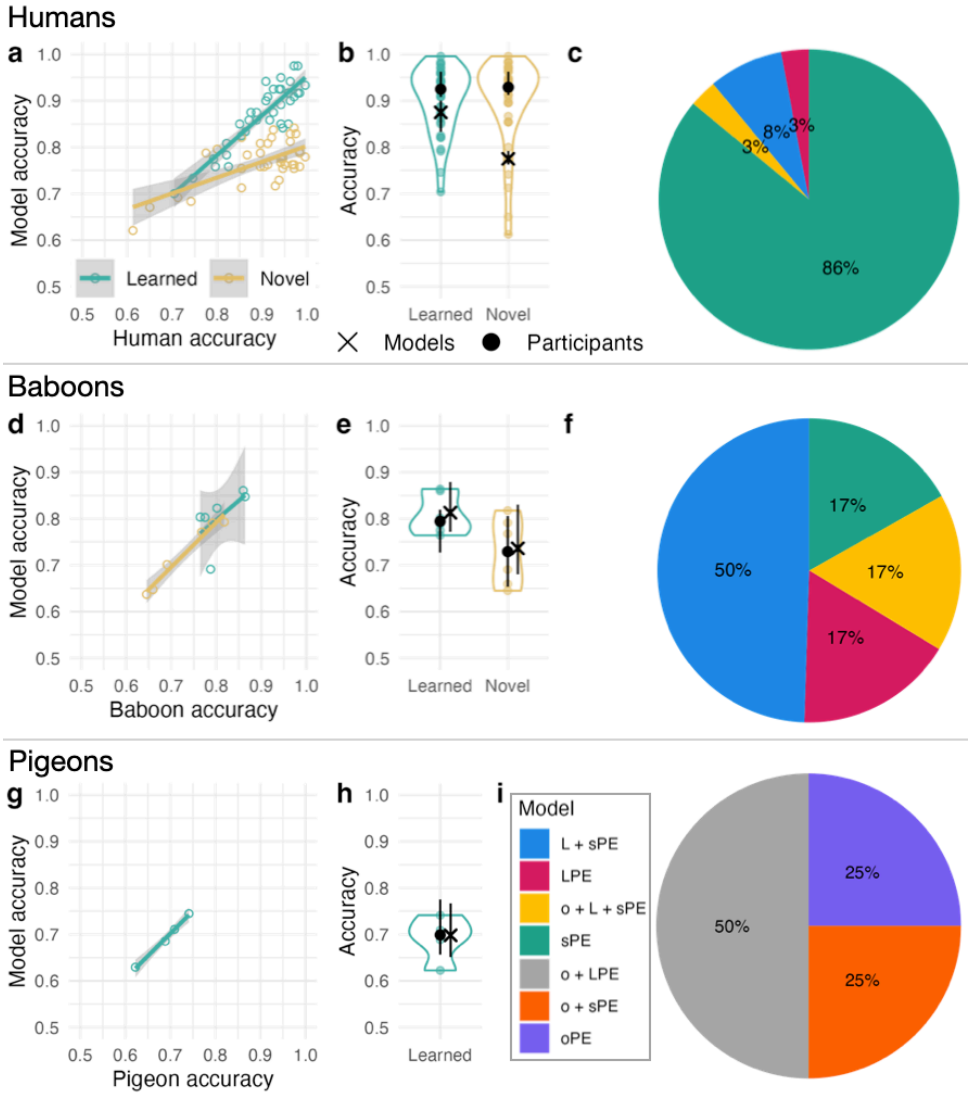

Fig. 2. Model fit and neuro-cognitive phenotypes for decision behavior for each participant by species. (a,d,g) Correlation of behavioral accuracy (x) and model accuracy (y) of each participant for learned (green) and novel (yellow) stimuli. Mean model accuracy was extracted from the models that most adequately described the performance on an individual level. Lines represent linear regression lines with confidence intervals. (b,e,h) Behavioral results (mean accuracy per individual; dots and violin, separated for learned and novel words), including group mean (black dot) and the mean model simulation results of the individual winning models (black X) with bootstrapped 95% confidence intervals. Note that only the data on learned letter strings was available for Pigeons. (c,f,i) Proportion of model variants that best described Human, Baboon, and Pigeon orthographic decision behavior.

interpretation, we focus on the inner workings of the models that show the highest similarity to the behavior of each participant, in other words, the neuro-cognitive phenotype for orthographic decisions.

Results

Mean orthographic decision accuracy (see Figure 2) showed that Humans performed best, with an average accuracy higher than 90%, whereas Baboon accuracy at 75% correct and Pigeon accuracy at 70% correct were significantly lower (i.e., no overlap in confidence intervals). Although the model simulations mirror this general pattern of results, we see that confidence intervals of the means of behavior (dot) and simulations (x) overlap for Baboons and Pigeons but not for Humans (see Figure 2). Thus, on average, the best-fitting speechless reader models performed at a similar accuracy as Baboons and Pigeons but are less accurate than Humans. This finding indicates, that the description of human orthographic decision performance is potentially influenced by knowledge that is currently not implemented in the model.

Table 1. Linear regression models investigating the association of model simulations and participant behavior.

Species (R^2)	Parameter	Estimate (Std. Error)	t	p
Humans (0.50)	Model	0.69 (0.08)	8.38	> 0.001
	Learned	-0.04 (0.02)	-2.39	0.020
Baboons (0.93)	Model	0.96 (0.11)	8.72	> 0.001
	Learned	-0.02 (0.01)	-1.75	0.114
Pigeons (0.99)	Model	1.02 (0.06)	17.63	0.003

Note. We tested the interaction of stimulus conditions and model simulation and found no significant interaction in Humans and Baboons.

Nevertheless, our model simulations were highly correlated with behavior across all three species (see Table 1 and Figure 2a,d,g), indicating that the selected models adequately represent the behavior of the individual participants. The pie charts in Figure 2 illustrate which model variants (including different prediction error representations) showed the highest fit to the individual Human, Baboon, and Pigeon data, respectively. We found that for Humans, the best-fitting model, in all except one participant, included letter-sequence-level representations: For 86% of participants, the best-

fitting model exclusively implemented letter-sequence-level representation (sPE). For 8%, it was combined with letter-sequence-level and letter-level representations (L + sPE), and for 3%, the whole model with all three representations was most adequate (o + L + sPE). The behavior of one participant (3%) without the letter-sequence-level representation was most adequately represented by a model with only the letter-level representation (LPE). Interestingly, the same model variants (i.e., including the same set of combinations of representations) exhibited the best fit for the Baboon data, albeit with different portions. Only one of the six Baboons' behavior was best described with the sPE model, which was most common in Humans. Most Baboons (50%; N = 3) required the implementation of a combination of letter-sequence-level and letter-level representations (L + sPE), the second most common type of representation found in Humans. For Pigeons, the set of the best-fitting models was different without overlap in the combinations of representations to the other species. All adequate models included pixel-level representations (oPE) in combination with letter-level representations (o + LPE) and letter-sequence-level representations (o + sPE) in one Pigeon each.

To measure the stability of the phenotyping, we estimated the split-half reliability (Table 2 upper section). We based the analysis on the agreement of the representations from the individual models fitted to one-half of the data with models fitted to the other half. We implemented random splitting and repeated the procedure one hundred times. Overall, reliability was high (agreement > 60% for 7 out of 9 representations), with the results for the letter and letter-sequence representations in Pigeons showing the lowest reliability. To assess if the estimation of the phenotypes relies heavily on the best-fitting model, we reevaluated the three representations based on the mean squared error from all implemented models (see Table 2 lower section). The lowest mean squared error emerged when models for Humans included the letter-sequence-level representations. For Baboons, we find letter-sequence-level and letter-level representations showing the lowest errors. For Pigeons, the error is lowest for the models including the pixel-level prediction error. These analyses substantiate the main findings from the best-fitting models.

Table 2. Investigation of the reliability and stability of the individual phenotype models. Upper section: Mean split-half reliability based on representation agreement in percent, including 95% confidence intervals in parentheses. Splitting was implemented randomly and repeated 100 times. Lower section: Mean squared errors, including 95% confidence intervals, from all implemented models (i.e., for all seven representation combinations and all thresholds) separated by species and representation (i.e., combining models that included the representation; for the oPE measure, this would include o + L + sPE, o + LPE, o + sPE, and oPE models). The latter part is included to investigate the independence of the phenotype estimation from the individual best-fitting model.

	Humans	Baboons	Pigeons
	Parameter stability (Split-half Reliability)		
Overall	87.6% [87.2 88.1]	90.0% [88.8 91.2]	68.3% [65.5 71.2]
oPE	96.3% [95.9 96.7]	85.8% [84.5 87.2]	89.3% [86.5 92.0]
LPE	69.9% [68.9 71.0]	99.7% [99.3 100.0]	56.8% [52.8 60.8]
sPE	95.4% [95.0 95.9]	84.5% [82.2 86.6]	59.0% [54.5 64.0]
	Model independent mean squared error		
oPE	0.26 [0.26 0.26]	0.22 [0.22 0.23]	0.23 [0.22 0.24]
LPE	0.24 [0.24 0.24]	0.19 [0.18 0.19]	0.24 [0.23 0.25]
sPE	0.23 [0.23 0.23]	0.19 [0.18 0.19]	0.26 [0.26 0.27]

Discussion

In an orthographic decision task (i.e., categorization of letter strings as learned or novel), we examined the neurocognitive representations in Humans, Baboons, and Pigeons. We could infer and compare potential representations implemented in orthographic decisions across species based on a transparent Speechless Reader computational model (13). Although previous investigations comparing orthographic decision behavior focused on the high similarity between the behavioral performance across species (7, 8), here we used the opportunity to investigate the differences between species and individual participants using neuro-cognitive phenotypes. Since we integrate results from Humans, Baboons, and Pigeons, we can compare the orthographic representations of species separated by 324 million years of independent evolution (32). Overall, we found evidence that all species use, at times, pixel-, letter- and letter-sequence-level representations. Yet, the contributions of those representational levels to the task performance varied between species. Humans predominantly relied on the implementation of letter-sequence-level representations suggesting that they extract relevant information from letter strings based on combinations of at least two letters. To some extent, this is also true for Baboons. However, they used letter-level representations to a larger extent than Humans. We found only one Baboon that implemented a letter-sequence-level representation only, which was predominant in Humans (i.e., found in 86%). In contrast, pixel-level representations were dominant for Pigeons (i.e., used by all Pigeons). These findings indicate that all three species solved the task but employed partly different cognitive strategies to reach orthographic decisions. Moving from letter-sequence via letter to pixel-level representations, Humans, Baboons, and Pigeons could be ordered along a continuum to integrate visual-orthographic information successfully.

The perceptual demands of the analyzed orthographic decision studies, which mirror the tasks of local and global visual processing, are of significant importance. Comparative studies involving Humans, non-human Primates, and Pigeons have revealed that perceptual specializations on global or local stimulus features impact various cognitive tasks (33–36). The striking similarity of our findings to those on global/local processing suggests a potential overlap in the visuo-cognitive processes required for visual orthographic integration. Humans have a strong propensity to group small elements into global configurations. Thus, they preferably process a large letter instead of its constituent small letters (37), easily see the global shape from partially deleted line drawings (38), integrate distant patterns into illusory shapes (39), and exploit perceptual principles like similarity and proximity to see a global form (40). When using the same tasks to compare Baboons or other Monkeys with Humans, Baboons respond more slowly to global than to local targets (41), while Macaques selectively attend to fine details of compound stimuli (42). Detailed analyses of these and similar results make it likely that these differences occur at the

attentional level, where perceptual grouping operations seem more attention-demanding for Monkeys. At the same time, they come easy in Humans (41). In addition to attentional differences, early visual cortical areas in Humans may be especially adept at encoding and integrating global visual information (43).

Pigeons are well-known for showing a prominent local precedence in various tasks (44–46). These findings are very likely due to their ecology as granivorous birds that have to discriminate small grains by minute featural differences against a cluttered background with their frontal visual field. The frontal field of Pigeons is represented in the tectofugal pathway corresponding to the mammalian extrageniculocortical system (47, 48). This system is far less prone to utilize global visual information than areas of the primate ventral cortical stream (36, 49). Thus, the neural adaptation to a granivorous lifestyle drives local precedence in Pigeons and could be responsible for their preference for pixel-level orthographic processing.

The differential specialization within a global-to-local processing continuum is not fixed in any of the three species we discuss. All of them have the remarkable ability to alter their search strategy in task- or training-dependent manners (50–52). As we have observed in our analysis, this adaptability is a fascinating aspect of their cognitive strategies. While, on average, Humans, Baboons, and Pigeons are positioned on a continuum of letter-sequence-, letter-, and pixel-level representation, each species shows individual differences and is, in principle, capable of utilizing all three strategies.

The finding of a substantial reliance on letter sequence representations in Human orthographic decisions aligns with the general notion that efficient readers implement large orthographic elements for word recognition and reading (i.e., letter combinations or words; Ref. (2, 3, 53)). In literacy acquisition, increased reading competence strongly decreases the influence of the number of letters on reading time (i.e., longer reading times for longer words; e.g., Ref.(15, 54, 55)) and increases the perceptual span (5, 56, 57). Still, when recognizing unknown letter strings (e.g., pseudowords), they rely again on letter units (58, 59). The findings of this study show that after training, Human readers implement orthographic decisions with non-words based on larger units combining multiple letters. However, the implemented model variants fall short of simulating Human behavior as accurately as the Baboon and Pigeon behavior.

A potential explanation is that Humans may rely on additional or different representations that we still need to implement in the model. Likely, candidate representations could be easily added to the model (when respecting the nature of the representations as prediction errors) and implemented on orthographic but also lexico-semantic and phonological processing levels (e.g., phoneme-level prediction error representations). Here, Human participants were efficient readers with a working reading system in place. However, we only modeled data from a pseudoword learning task (i.e., unknown letter strings; Ref. (23)), so it is likely that phonology played a role, given compelling evidence for its influence on visual word recognition (e.g., when learning to read; Ref. (15, 60); in adult readers (18, 61–63); when reading problems occur; Ref. (14, 64)). Another aspect is that in contrast to more classical measures of orthography, the sequence-based prediction error implemented here respects the exact sequence, always starting with the first letter of the string. Here, we chose this representation as the first hypothesis on the sequence level motivated by finding a word length effect for pseudowords in adult reading (58, 59). However, the letter-sequence level implementation is only one of many potential representations that can characterize the orthographic structure of letter strings (e.g., based on bi- or tri-grams could have higher flexibility to model transposed letter effects; e.g., see Ref. (20, 53)). Future work on integrating further representations in the model will allow a systematic comparison of representations used by Humans and, simultaneously, offer new possibilities to describe differences between individuals and species in orthographic processing and beyond.

Conclusion. Our analyses reveal that the ability of the three species to differentiate between words and nonwords orthographically is based on partly different visuo-cognitive strategies. While Human readers mainly relied on letter-sequence representations, Baboons utilized letter and letter-sequence level strategies, while Pigeons heavily relied on pixel-level representations. These findings imply that pure "success testing" in comparative cognitive studies obstructs the more profound differences in species-specific cognitive strategies. Instead, a "signature testing" approach is needed that reconstructs how a task is solved (65). Such an approach can uncover species-specific visuo-cognitive strategies and reveal the potential neuroevolutionary adaptations that drive these cognitive strategies.

ACKNOWLEDGMENTS. We thank Ulrike Basten for helping with the reliability estimation. This research was funded by the German Research Foundation (Grant no. 523332674 awarded to B.G.), the European Research Foundation (Project AVIAN MIND ERC-2020-ADG, LS5, GA No. 101021354 awarded to O.G.), and Royal Society of New Zealand (Marsden Fund grant 19-UOO-162 awarded to M.C.).

Materials and Methods

Behavioral data. The data used for model simulations stems from Eisenhauer et al. (23) (Humans), Grainger et al. (7) (Baboons), and Scarf et al. (8) (Pigeons). Eisenhauer et al. (23) tested 37 Human participants on a pseudoword lexical decision task of 880 trials, including eight presentations of 60 to-be-learned pseudowords (i.e., two presentations in each of the four sessions). Grainger et al. (7) used operant conditioning setups to train six Baboons (*Papio papio*) in a word/non-word categorization task. Scarf et al. (8) trained four Pigeons (*Columba livia*) on a similar task. The limiting factor in this comparison is the Pigeon data, of which we could only obtain the responses to the learned letter strings of the last part of the experiment, while earlier responses and the responses to non-words were missing. To make the data comparable across species, we limited our primary analysis to the final 10,000 trials of each Baboon dataset and the two last sessions in the Human dataset, amounting to 440 trials per participant. Furthermore, the limitation to the learned items led to an adaptation of the lexical process part of the model (18), allowing the model to be adapted to individual learner thresholds for each participant (see below). For further information on participants and stimulus materials, please refer to the original publications.

Model implementation. Instead of automatically estimating the categorization boundary (blue line in Fig. 1) based on the overlap of word-likeness estimates from words and non-words, we implemented the categorization process based only on the word-likeness distribution of the learned letter strings (also described

$$Prediction_{i,j} = \frac{\sum_1^n StoredItems_{i,j}}{n} \quad [1]$$

$$PE_s = \sum SensoryInput_{i,j} - Prediction_{i,j} \quad [2]$$

$$oPE_s = \sum \begin{cases} 1 & \text{if } (SensoryInput_{i,j} - Prediction_{i,j}) > .5 \\ 0 & \text{otherwise} \end{cases} \quad [3]$$

in Ref. (18)). To find an adequate categorization threshold, we calculated the model responses for all possible model variants and multiple possible threshold values (range .5-1; see Eq. 6) for each participant (see Supporting Information: Table S1 and Fig. S1, S2, S3). To fit the optimal boundary, we compared the model simulations to the behavior of each participant.

For all of our model simulations, we investigate the behavior after initial training (i.e., when the performance of animals is relatively stable). For all datasets, the items in the lexicon are determined based on the performance in trials used in the analysis (all trials after the first 33,041 trials, i.e., determined by the available data for both Baboons and Pigeons and data from sessions 3 and 4 from the Human dataset). We use the criterion for learned items, which was previously reported (Accuracy > .71), to determine the participant-specific implementation of the items in the individual lexicons of the speechless reader. When the lexicon items are known, we can estimate the top-down predictions (gray lines in Figure 1) on the level of pixels, letters, and letter sequences (see Eq. 1, described in Ref.(12, 13, 27)). When estimated on the pixel level, the *StoredItems* matrix consists of gray values defining the image of each item stored in the lexicon (i,j = 140, 40 each). For letters, the *StoredItems* matrix consists of all letters (for four-letter words: i,j = 4, 26 each); and for the letter sequence level, of all possible letter sequences (e.g., for four-letter words: Letters 1 and 2; Letters 1, 2, and 3; i,j = 2, x where x is the number of all possible letter combinations for the two and three letter sequences). For each pixel, letter, and letter sequence, a mean over the number of items in the lexicon (n) is calculated to represent the prediction. Note that letter-based representations were informative as for the animal studies, the letter frequencies ranged from 2.6 to 5.3% at any of the five positions. Also, explorations with inverted sequence prediction errors (i.e., starting the sequence with the last letter of a word) showed much less accurate results, indicating that animals used sequence representations in the reading direction.

Next, we integrate the predictions with the sensory input by subtraction (see Formula 2). The *SensoryInput* matrix is similar to the prediction on each level, so the subtraction can be implemented for each letter or letter sequence. Note that we represent the sensory input as 1 for the letter and letter-sequence level, indicating the highest possible error when the input is unpredictable. This value is reduced by the prediction value for the letter at the given position (e.g., in English, there is a high probability for the *s* being the last letter, so we expect a low prediction error at this position) or for the letter sequence relevant (i.e., *th* is likely for English for the initial letter sequence). We further developed the formulation for the pixel level (oPE). We conjectured that a binary prediction error would be most appropriate when modeling word/non-word categorization behavior (see Eq. 3 and (12)). We found that a threshold at 0.5 is well suited to implement the binary prediction error representation. After integrating prediction and sensory input to generate the prediction error, all matrix values are summed up to get one value summarizing the overall prediction error on each prediction error level for each letter string.

We calculated the prediction errors on all three levels for each learned and novel letter string available in the datasets. In preparation for the integration of multiple prediction errors to estimate word-likeness based on prediction errors, we normalized each set of prediction errors to a value between 0 and 1 (PE_{norm}; see Eq. 4). After that, we can accumulate multiple prediction error values by summation and division (see Eq. 5). The resulting prediction error estimate is normalized in all model variants that integrate more than one prediction error representation.

In the final step, we implement the lexical categorization process. Initially, we implemented the boundary at the highest lexical categorization difficulty (i.e., when distributions of prediction error-based word likeness of learned and novel letter strings overlapped the most, see (18)). Since only the learned letter strings were available from the Pigeon data set, we conceived of an alternative approach that would allow us to implement lexical categorization without including novel letter strings. We achieved this by considering multiple thresholds applied to the prediction errors of the learned letter strings (see Eq. 6). Since we could not automatically calculate the optimal threshold, which would have been the case in the original formulation (18), multiple thresholds in the range from $0.5 - 0.95$ were applied to all model variants to achieve a binary value, which indicates whether a letter string is learned or not (see the blue line in Fig. 1). Interestingly, this alternative approach resulted in better model fits than the original implementation used for Baboons in the past (i.e., cp. Ref. (13)). Thus, for the present analysis, we focused on the approach that can be implemented without assuming that non-words must be stored.

With this model, we simulated the behavior of each participant with each model variant (i.e., all prediction error combinations with all possible thresholds, N = 7). To determine the most appropriate model, we compared the model and participant behavior for all trials based on the mean squared error (see Fig. S1-3a; see also mean model accuracy in Fig. S1-3b). We selected the threshold value that resulted in the lowest absolute difference between the data and the simulations from all model variants. (Note that a different lexicon is assumed for each participant, which is why simulations differed between participants). The model variant with the lowest mean squared error was considered to include the prediction error representations most likely implemented to achieve the orthographic decision performance of the individual participant (see Figure 2c,f,i and Table S1). The model was implemented in Python, scripts, and data will be made available when accepted for publication.

Data analysis. We tested simulation quality with linear regression models implemented in GNU R, investigating the similarity of the individual differences between model simulation and participant behavior. In addition, to investigate the group-level similarity, we provide descriptive statistics, including boot-strapped 95%

$$PE_{norm} = \frac{PE_s - \min(PE_s)}{\max(PE_s) - \min(PE_s)} \quad [4]$$

$$PE_{accumulated} = \frac{\sum_1^n PE_{norm,n}}{n} \quad [5]$$

$$Learned(PE) = \begin{cases} 1 & \text{if } PE_{accumulated} > Threshold \\ 0 & otherwise \end{cases} \quad [6]$$

$$Threshold = \left[.5 - .95\right]$$

confidence intervals. To examine the stability of the implemented prediction error representations, we ran one hundred split-half reliability estimations (e.g., see (26) for the importance of reliability estimates in the context of computational phenotyping), estimating the overlap between the assumed representations between the models fitted to either of the two halves. This was done for each species and each model variant.

References.

1. S Dehaene, L Cohen, The unique role of the visual word form area in reading. *Trends Cogn. Sci.* **15**, 254–262 (2011).
2. M Coltheart, K Rastle, C Perry, R Langdon, J Ziegler, DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* **108**, 204–256 (2001).
3. C Perry, JC Ziegler, M Zorzi, Nested incremental modeling in the development of computational theories: The cdp model of reading aloud. *Psychol. Rev.* **114**, 273–315 (2007).
4. R Engbert, A Nuthmann, EM Richter, R Kliegl, SWIFT: A dynamical model of saccade generation during reading. *Psychol. Rev.* **112**, 777–813 (2005).
5. K Rayner, The 35th sir frederick bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* **62**, 1457–1506 (2009).
6. RK Olson, Dyslexia: nature and nurture. *Dyslexia* **8**, 143–159 (2002).
7. J Grainger, S Dufau, M Montant, JC Ziegler, J Fagot, Orthographic processing in baboons (papio papio). *Science* **336**, 245–248 (2012).
8. D Scarf, et al., Orthographic processing in pigeons (columba livia). *Proc. Natl. Acad. Sci.* **113**, 11272–11276 (2016).
9. M Linke, F Bröker, M Ramscar, H Baayen, Are baboons learning "orthographic" representations? probably not. *PLOS ONE* **12**, e0183876 (2017).
10. T Hannagan, JC Ziegler, S Dufau, J Fagot, J Grainger, Deep learning of orthographic representations in baboons. *PLoS ONE* **9**, e84843 (2014).
11. T Hannagan, A Agrawal, L Cohen, S Dehaene, Emergence of a compositional neural code for written words: Recycling of a convolutional neural network for reading. *Proc. Natl. Acad. Sci.* **118** (2021).
12. W Fu, B Gagl, Specifying the orthographic prediction error for a better understanding of efficient visual word recognition in humans and machines. *bioRxiv* (2024).
13. B Gagl, I Weyers, JL Mueller, Speechless reader model: A neurocognitive model for human reading reveals cognitive underpinnings of baboon lexical decision behavior. in *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 43, (2021).
14. A Castles, K Rastle, K Nation, Ending the reading wars: Reading acquisition from novice to expert. *Psychol. Sci. Public Interest* **19**, 5–51 (2018).
15. B Gagl, S Hawelka, H Wimmer, On sources of the word length effect in young readers. *Sci. Stud. Read.* **19**, 289–306 (2015).
16. JC Ziegler, C Perry, M Zorzi, Modelling reading development through phonological decoding and self-teaching: implications for dyslexia. *Philos. Transactions Royal Soc. B: Biol. Sci.* **369**, 20120397 (2014).
17. B Gagl, et al., Eye movements during text reading align with the rate of speech production. *Nat. Hum. Behav.* **6**, 429–442 (2022).
18. B Gagl, et al., The lexical categorization model: A computational model of left ventral occipito-temporal cortex activation in visual word recognition. *PLOS Comput. Biol.* **18**, e1009995 (2022).
19. B Gagl, K Gregorová, Investigating lexical categorization in reading based on joint diagnostic and training approaches for language learners. *npj Sci. Learn.* **9** (2024).
20. T Yarkoni, D Balota, M Yap, Moving beyond coltheart's n: A new measure of orthographic similarity. *Psychon. Bull. Rev.* **15**, 971–979 (2008).
21. DA Balota, JI Chumbley, Are lexical decisions a good measure of lexical access? the role of word frequency in the neglected decision stage. *J. Exp. Psychol. Hum. perception performance* **10**, 340 (1984).
22. K Gregorová, J Turini, B Gagl, MLH Võ, Access to meaning from visual input: Object and word frequency effects in categorization behavior. *J. Exp. Psychol. Gen.* (2023).
23. S Eisenhauer, CJ Fiebach, B Gagl, Context-based facilitation in visual word recognition: Evidence for visual and lexical but not pre-lexical contributions. *eneuro* **6**, ENEURO.0321–18.2019 (2019).
24. EH Patzelt, CA Hartley, SJ Gershman, Computational phenotyping: Using models to understand individual differences in personality, development, and mental illness. *Pers. Neurosci.* **1** (2018).
25. P Schwartenbeck, K Friston, Computational phenotyping in psychiatry: A worked example. *eneuro* **3**, ENEURO.0049–16.2016 (2016).
26. R Schurr, D Reznik, H Hillman, R Bhui, SJ Gershman, Dynamic computational phenotyping of human cognition. *Nat. Hum. Behav.* (2024).
27. B Gagl, et al., An orthographic prediction error as the basis for efficient visual word recognition. *NeuroImage* **214**, 116727 (2020).
28. RPN Rao, DH Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
29. M Spratling, A review of predictive coding algorithms. *Brain Cogn.* **112**, 92–97 (2017).
30. A Clark, Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
31. S Eisenhauer, B Gagl, CJ Fiebach, Predictive pre-activation of orthographic and lexical-semantic representations facilitates visual word recognition. *Psychophysiology* **59** (2021).
32. M dos Reis, et al., Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25**, 2939–2950 (2015).
33. B Weinstein, Matching-from-sample by rhesus monkeys and by children. *J. Comp. Psychol.* **31**, 195–213 (1941).
34. G Spinozzi, C De Lillo, V Truppa, G Castorina, The relative use of proximity, shape similarity, and orientation as visual perceptual grouping cues in tufted capuchin monkeys (cebus apella) and humans (homo sapiens). *J. Comp. Psychol.* **123**, 56–68 (2009).
35. V Truppa, DA De Simone, C De Lillo, Short-term memory effects on visual global/local processing in tufted capuchin monkeys (sapajus spp.). *J. Comp. Psychol.* **130**, 162–173 (2016).
36. W Clark, M Colombo, Seeing the forest for the trees, and the ground below my beak: Global and local processing in the pigeon's visual system. *Front. Psychol.* **13** (2022).
37. D Navon, Forest before trees: The precedence of global features in visual perception. *Cogn. Psychol.* **9**, 353–383 (1977).
38. J Fagot, M Tomonaga, Global and local processing in humans (homo sapiens) and chimpanzees (pan troglodytes): Use of a visual search task with compound stimuli. *J. Comp. Psychol.* **113**, 3–12 (1999).
39. J Fagot, M Tomonaga, Effects of element separation on perceptual grouping by humans (homo sapiens) and chimpanzees (pan troglodytes): perception of kanizsa illusory figures. *Animal Cogn.* **4**, 171–177 (2001).
40. D Burke, P Everingham, T Rogers, M Hinton, S Hall-Aspland, Perceptual grouping in two visually reliant species: Humans (homo sapiens) and australian sea lions (neophoca cinerea). *Perception* **30**, 1093–1106 (2001).
41. C Deruelle, J Fagot, Visual search for global/local stimulus features in humans and baboons. *Psychon. Bull. Rev.* **5** (1998).
42. DD Kurylo, J Van Nest, B Knepper, Characteristics of perceptual grouping in rats. *J. Comp. Psychol.* **111**, 126–134 (1997).
43. M Furlan, AT Smith, Global motion processing in human visual cortical areas v2 and v3. *J. Neurosci.* **36**, 7314–7324 (2016).

44. KK Cavoto, RG Cook, Cognitive precedence for local information in hierarchical stimulus processing by pigeons. *J. Exp. Psychol. Animal Behav. Process.* **27**, 3–16 (2001).
45. TB Patton, G Szafranski, T Shimizu, Male pigeons react differentially to altered facial features of female pigeons. *Behaviour* **147**, 757–773 (2010).
46. J Emmerton, JC Renner, Local rather than global processing of visual arrays in numerosity discrimination by pigeons (columba livia). *Animal Cogn.* **12**, 511–526 (2009).
47. M Remy, O Güntürkün, Retinal afferents to the tectum opticum and the nucleus opticus principalis thalami in the pigeon. *J. Comp. Neurol.* **305**, 57–70 (1991).
48. O Güntürkün, U Hahmann, Functional subdivisions of the ascending visual pathways in the pigeon. *Behav. Brain Res.* **98**, 193–201 (1999).
49. WJ Clark, M Colombo, The functional architecture, receptive field characteristics, and representation of objects in the visual network of the pigeon brain. *Prog. Neurobiol.* **195**, 101781 (2020).
50. T Fremouw, WT Herbranson, CP Shimp, Priming of attention to local or global levels of visual analysis. *J. Exp. Psychol. Animal Behav. Process.* **24**, 278–290 (1998).
51. Y Yamazaki, U Aust, L Huber, M Hausmann, O Güntürkün, Lateralized cognition: Asymmetrical and complementary strategies of pigeons during discrimination of the "human concept". *Cognition* **104**, 315–344 (2007).
52. J Martin-Malivel, Discrimination of contour-deleted images in baboons (papio papio) and chimpanzees (pan troglodytes). *Animal Cogn.* **14**, 415–426 (2011).
53. S Dehaene, L Cohen, M Sigman, F Vinckier, The neural code for written words: a proposal. *Trends Cogn. Sci.* **9**, 335–341 (2005).
54. E Marinus, PF de Jong, Variability in the word-reading performance of dyslexic readers: Effects of letter length, phoneme length and digraph presence. *Cortex* **46**, 1259–1271 (2010).
55. P Zoccolotti, et al., Word length effect in early reading and in developmental dyslexia. *Brain Lang.* **93**, 369–373 (2005).
56. A Sperlich, J Meixner, J Laubrock, Development of the perceptual span in reading: A longitudinal study. *J. Exp. Child Psychol.* **146**, 181–201 (2016).
57. J Snell, M Meeter, J Grainger, Evidence for simultaneous syntactic processing of multiple words during reading. *PLOS ONE* **12**, e0173720 (2017).
58. BS Weekes, Differential effects of number of letters on word and nonword naming latency. *The Q. J. Exp. Psychol. Sect. A* **50**, 439–456 (1997).
59. MJ Yap, DE Sibley, DA Balota, R Ratcliff, J Rueckl, Responding to nonwords in the lexical decision task: Insights from the english lexicon project. *J. Exp. Psychol. Learn. Mem. Cogn.* **41**, 597–613 (2015).
60. R Bijeljac-Babic, V Millogo, F Farioli, J Grainger, A developmental investigation of word length effects in reading using a new on-line word identification paradigm. *Read. Writ.* **17**, 411–431 (2004).
61. S Hawelka, S Schuster, B Gagl, F Hutzler, Beyond single syllables: The effect of first syllable frequency and orthographic similarity on eye movements during silent reading. *Lang. Cogn. Process.* **28**, 1134–1153 (2013).
62. JC Ziegler, L Ferrand, AM Jacobs, A Rey, J Grainger, Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *The Q. J. Exp. Psychol. Sect. A* **53**, 671–692 (2000).
63. M Conrad, M Carreiras, S Tamm, AM Jacobs, Syllables and bigrams: Orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 461–479 (2009).
64. K Galuschka, E Ise, K Krick, G Schulte-Körne, Effectiveness of treatment approaches for children and adolescents with reading disabilities: A meta-analysis of randomized controlled trials. *PLoS ONE* **9**, e89900 (2014).
65. E Kakrada, M Colombo, Beyond the mark: Signatures of self-recognition in fish. *Learn. Behav.* **52**, 5–6 (2023).

Table S1. Number of in the lexicon for each participant. Also, mean squared error, representations, and threshold for the winning model for each participant.

	N	Error	oPE	LPE	sPE	Thres.
Human/1	56	0.10	0	0	1	80
Human/2	59	0.01	0	0	1	95
Human/3	49	0.10	0	1	1	83
Human/4	45	0.09	0	0	1	77
Human/5	55	0.12	0	0	1	89
Human/6	59	0.11	0	0	1	89
Human/7	51	0.02	0	0	1	83
Human/8	53	0.12	0	0	1	89
Human/9	58	0.03	0	0	1	92
Human/10	59	0.14	0	0	1	92
Human/11	55	0.08	0	0	1	89
Human/12	48	0.09	0	0	1	86
Human/13	57	0.07	0	0	1	89
Human/14	59	0.08	0	0	1	89
Human/15	58	0.11	0	0	1	86
Human/16	52	0.09	0	0	1	83
Human/17	55	0.11	0	0	1	86
Human/18	56	0.12	0	0	1	89
Human/19	55	0.11	0	0	1	89
Human/20	55	0.11	0	0	1	89
Human/21	57	0.03	0	0	1	92
Human/22	58	0.13	0	0	1	89
Human/23	42	0.02	1	1	1	83
Human/24	44	0.09	0	0	1	80
Human/25	50	0.05	0	0	1	77
Human/26	40	0.07	0	0	1	71
Human/27	56	0.13	0	0	1	86
Human/28	54	0.10	0	0	1	89
Human/29	54	0.09	0	0	1	89
Human/30	54	0.05	0	0	1	92
Human/31	50	0.08	0	1	1	68
Human/32	52	0.10	0	0	1	83
Human/33	34	0.01	0	1	0	68
Human/34	56	0.07	0	0	1	80
Human/35	52	0.10	0	0	1	89
Human/36	52	0.06	0	0	1	89
Human/37	48	0.10	0	1	1	77
Baboon/ARI	67	0.02	0	1	0	86
Baboon/ART	87	0.02	0	1	1	93
Baboon/CAU	73	0.02	0	1	1	93
Baboon/DAN	271	0.03	0	1	1	84
Baboon/DOR	96	0.02	0	0	1	82
Baboon/VIO	86	0.01	1	1	1	87
Pigeon/Q32	8	0.01	1	0	0	95
Pigeon/Q35	28	< 0.01	1	0	1	94
Pigeon/Q41	21	< 0.01	1	1	0	94
Pigeon/Q43	32	< 0.01	1	1	0	91

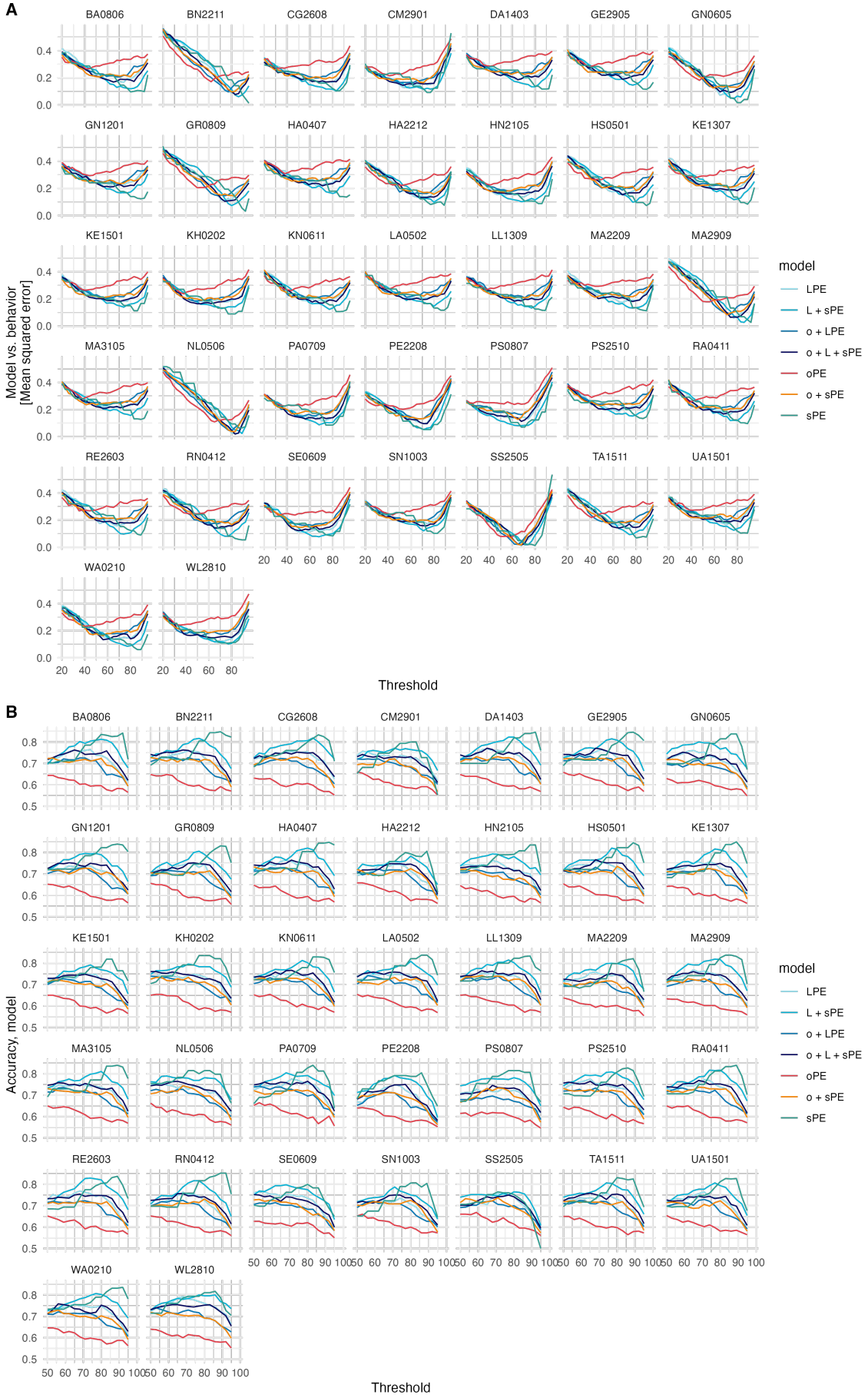

Fig. S1. Model comparisons and model accuracy for all models and all humans. (A) Shows the mean squared error of each model (i.e., all combinations of representations and all tested thresholds) and human behavior. (B) Model accuracy from all tested models.

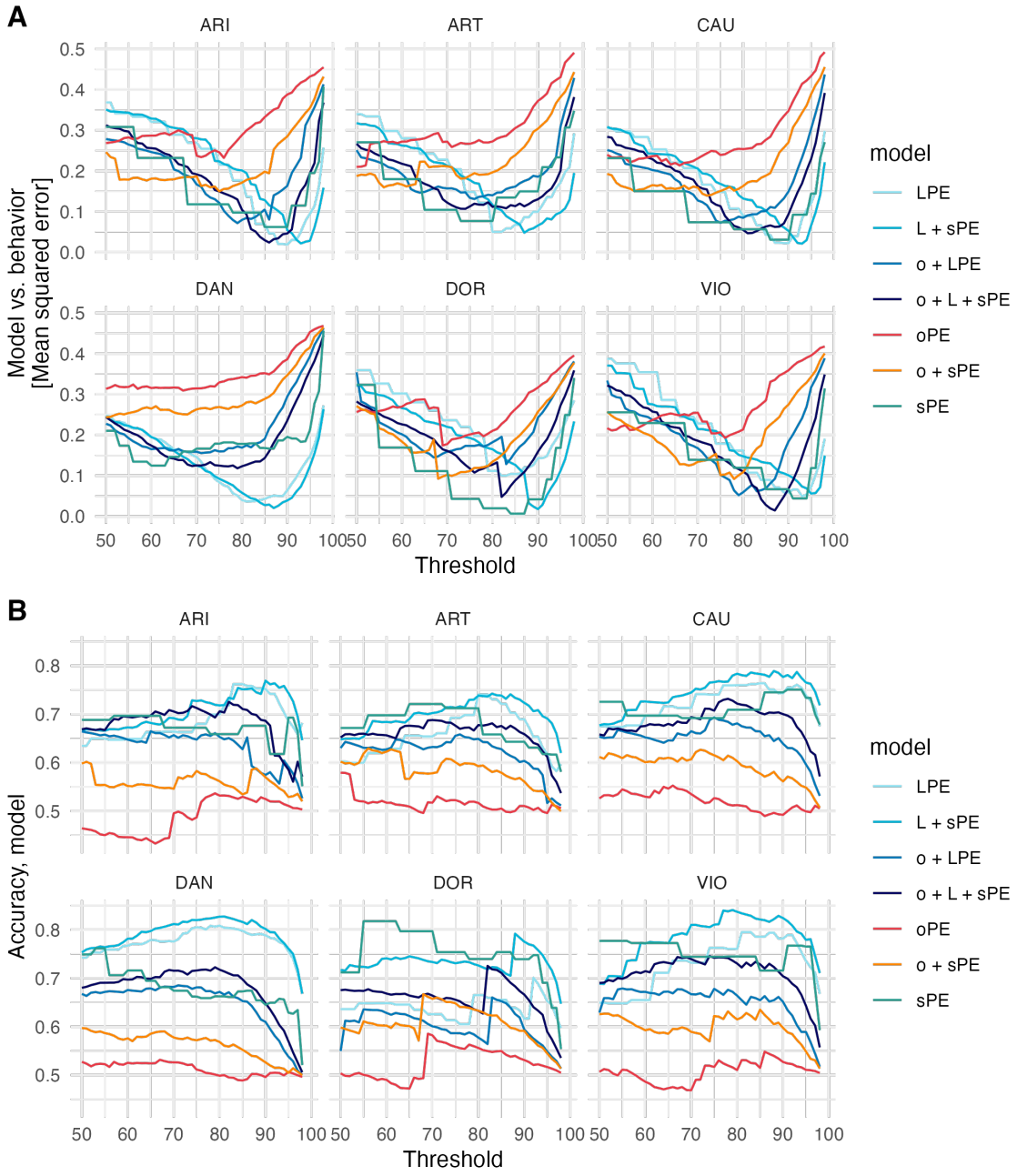

Fig. S2. Model comparisons and model accuracy for all models and all Humans. (A) Shows the mean squared error of each model (i.e., all combinations of representations and all tested thresholds) and Baboon behavior. (B) Model accuracy from all tested models.

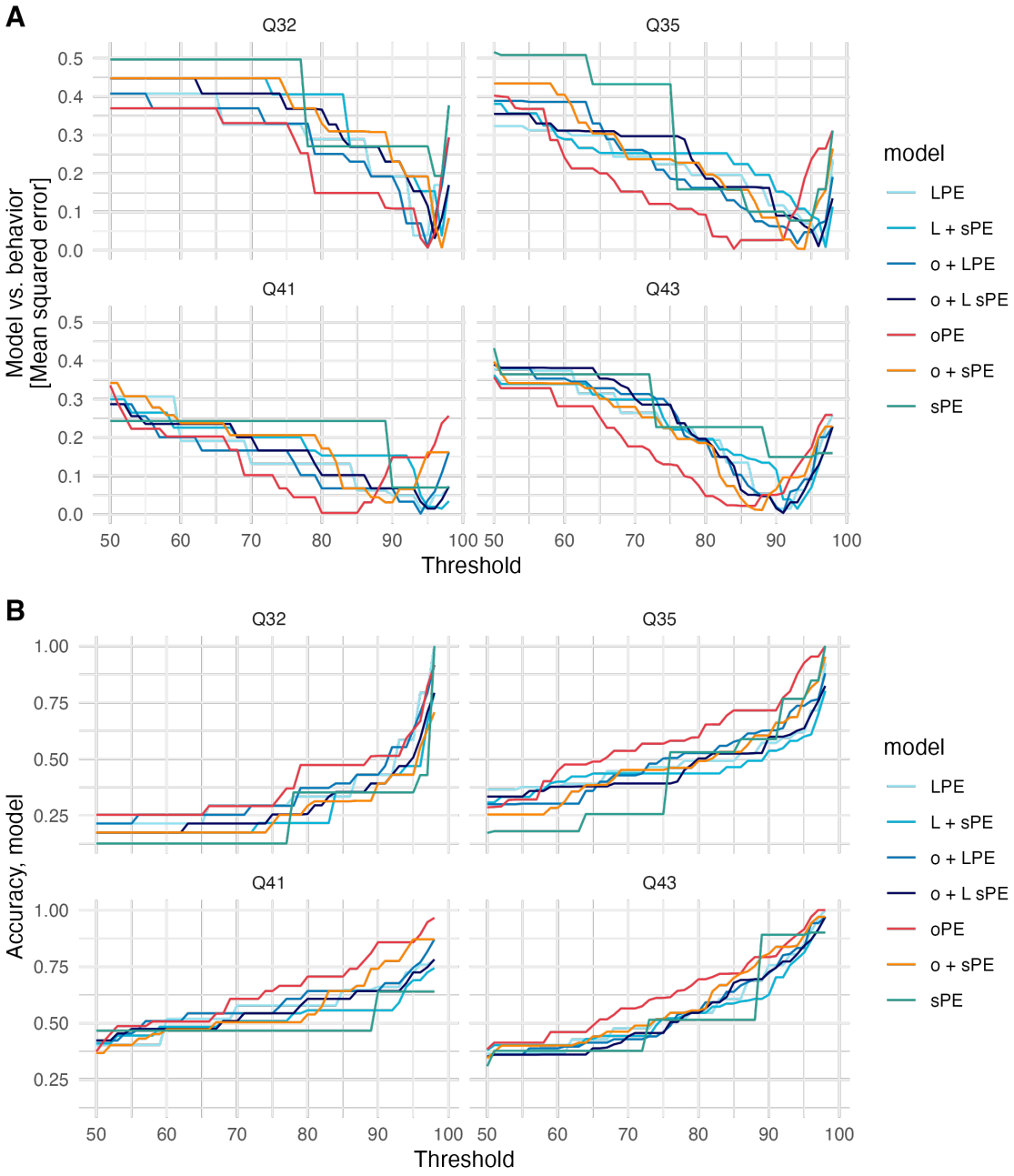

Fig. S3. Model comparisons and accuracy for all models and all Pigeons. (A) Shows the mean squared error of each model (i.e., all combinations of representations and all tested thresholds) and pigeon behavior. (B) Model accuracy from all tested models.