Speechless Reader Model: A neurocognitive model for human reading reveals cognitive underpinnings of baboon lexical decision behavior.

Benjamin Gagl (benjamin.gagl@univie.ac.at)

Ivonne Weyers (ivonne.weyers@univie.ac.at)

Lukas Wurth (lukaswurth@hotmail.com)

Jutta Mueller (jutta.mueller@univie.ac.at)

Department of Linguistics, University of Vienna,
Sensengasse 3a, 1090 Vienna

Abstract

Animal reading studies have shown that word/non-word decision behavior can be performed by baboons and pigeons, despite their inability to access phonological and semantic representations. Previous modeling work used different learning models (e.g., deep-learning architectures) to successfully reproduce baboon lexical decisions. More transparent investigations of the implemented representations underlying baboons' behavior are currently missing, however. Here we apply the highly transparent Speechless Reader Model, which is motivated by human reading and its underlying neurocognitive processes, to existing baboon data. We implemented four variants that comprise different sets of representations—all four models implemented visual-orthographic prediction errors. In addition, one model included prediction errors derived from positional letter frequencies, one prediction errors constrained by specific letter sequences, and finally, a combinatory model combined all three prediction errors. We compared the models' behavior to that of the baboons and thereby identified one model which most adequately mirrored the animals' learning success. This model combined the image-based prediction error and the letter-based prediction error that also accounts for the transitional probabilities within the letter sequence. Thus, we can conclude that animals, similarly to humans, use prediction error representations that capture orthographic codes to implement efficient reading-like behavior.

Keywords: Animal reading; Baboons; Lexical categorization; cognitive modeling; transparent orthographic representations

Introduction

The invention of writing systems is a critical cultural achievement underlying the success of human societies. Reading involves letter recognition, their sequential combination and conversion to vocal word forms and finally, the extraction of meaning from the written text (e.g., Coltheart et al., 2001). Recent comparative cognitive research studied reading-like behavior in baboons (Grainger et al., 2012; Rajalingham et al., 2020) and pigeons (Scarf et al., 2016), investigating the evolutionary basis of reading. These studies examined if animals can differentiate written words from non-words, when rewarded for correct responses. Both studies showed that this classic lexical decision task, which is also often used to investigate human reading (e.g., Balota & Chumbley, 1984), can be solved by non-human animals with high accuracy (~74%), yet after considerable training. These datasets sparked a lively debate on how animals address these tasks (Hannagan et al., 2014; Linke et al., 2017) and how comparable animals' lexical categorizations are to human reading (Katz et al., 2012).

The apparent difference between animal word/non-word categorization and human reading is that the former lacks speech and meaning processing (Katz et al., 2012). In other words, human readers likely decipher the orthographic code via linguistic processing routes, mapping the written words onto their corresponding speech representations (e.g., Coltheart et al., 2001). In contrast, animals solving the lexical categorization task for a reward. So that they have not learned to perform grapheme-to-phoneme conversion and do not have this linguistic route at their disposal, therefore, animals and humans likely differ notably in how they process written input. As an objection, it has been argued that humans might still exploit the characteristics of orthographic codes similarly to baboons, namely on the lower, non-linguistic level, to optimize their reading performance (Grainger et al., 2012). Supporting evidence for this notion comes from recent studies (e.g., Gagl, Richlan, et al., 2020) using models that describe lexical categorization in a way that is naive to phonology and semantics. The research shows that lexical categorization, which baboons and pigeons also master, is possibly implemented in a left-ventral occipito-temporal region in the human cortex known to be highly relevant for reading, namely the visual word form area (Dehaene & Cohen, 2011). Similarly, Rjalingham et al. (2020) found that neuronal activity collected from macaque monkeys exposed to several orthographic processing tasks not only closely mirrored baboon performance (Grainger et al., 2012), but highlighted the inferior temporal cortex as a possible precursor to human brain networks involved in reading. Thus, current evidence suggests that the cortical areas involved in human reading might be built upon visual processing networks present in primates, which already allow for low-level, non-linguistic computations to be performed on orthographic input. Assuming such a shared origin, animals and humans could similarly exploit the orthographic code of written stimuli to reach very different goals.

Yet, an explicit model capable of describing how animals (and humans) use the characteristics of the orthographic code to achieve efficient reading or lexical categorization is still amiss. Several model-based investigations used learning approaches with different architectures to successfully reproduce the baboon data (Hannagan et al., 2014; Linke et al., 2017). Learning models are great for investigating complex system behavior and differences in computational architectures (Cichy & Kaiser, 2019; Ma & Peters, 2020). Comparing the deep-model structure implemented by Hannagan et al., and the "flat"-model structure by Linke et al., is intriguing. Both models simulate the baboon data well, but the "flat" model is much simpler, suggesting that lexical categorization can be achieved based on non-hierarchical reinforcement learning of visual features. Precisely, the "flat" model consists of only one layer that learns the association of 14,476 visual features to either represent words or non-words based on the Rescorla-Wagner learning rule (Rescorla & Wagner, 1972). Thus, even with a low number of layers, learning models typically implement a high number of free parameters, i.e., parameters set during model training. This model characteristic is beneficial for achieving highly accurate predictions in various contexts (Cichy & Kaiser, 2019; Ma & Peters, 2020). For example, researchers have successfully used such models to investigate occipitotemporal brain activation in humans during object recognition (Güçlü & Gerven, 2015). Still, the use of free parameter setting comes at the cost of transparency of the implemented representations (Cichy & Kaiser, 2019; Ma & Peters, 2020) and naturally reduces these models' explanatory value. Thus, if one wants to understand how animals achieve accurate lexical categorization performance, "flat" or "deep" learning models, able to implement numerous different task, might not be the best option.

Here we implement a computational model that was developed specifically for reading like tasks, i.e., lexical categorization. The Speechless Reader Model combines previously evaluated assumptions describing human neurocognitive processes implemented in the occipitotemporal cortex during reading (Gagl, Davis, et al., 2020; Gagl, Richlan, et al., 2020; Gagl, Sassenhagen, et al., 2020). Like animals and previously used models, the Speechless Reader Model is naive to phonological and semantic processing. The specificity to one group of tasks allows that the model implements highly transparent representations and algorithms with the goal of correct lexical categorization while dispensing with any free parameters. The model presented here for the first time is based on previously developed model components relevant to human reading (Gagl, Davis, et al., 2020; Gagl, Richlan, et al., 2020; Gagl, Sassenhagen, et al., 2020). Thus, these general model characteristics sets it apart from previous investigations using domain-general learning models. From this study, we expect an increased understanding of (i) the underlying representations and algorithms implemented in baboons' lexical categorization and of (ii) the relation between humans' and baboons' cognitive processes in response to written words.

Previously, e.g., Linke et al. questioned if baboons implement any orthographic representations at all, as their model successfully mimicked baboon behavior while only processing visual features. To further address this question, we compare four models that include orthographic representations of different complexity. The most straightforward one is based on only word images, i.e., representations primarily including visual features similar to Linke et al.'s model on a conceptual level (see Gagl, Sassenhagen, et al., 2020 for more details). The second model further encodes positional letter frequencies. The third, then, includes sequential letter information. The fourth and final model ultimately combines the learning parameters of the second and third models. As a control model, we will include simulation data from the domain-general reinforcement model by Linke et al.. This characterizing of the orthographic cues used by baboons allows the investigation of similarities and differences between baboons and human reading.

The Speechless Reader Model

The Speechless Reader Model combines multiple component models implemented to describe neuro-cognitive processes in adult readers. For the first time in this form, prediction error representations based on the visual appearances of words (Gagl, Sassenhagen, et al., 2020) are combined with prediction error representations based on letter frequencies with and without accounting for the letter sequence (Gagl, Davis, et al., 2020). The prediction errors are accumulated before a word/non-word categorization process is initiated. This process functions as a gatekeeper, i.e., it prevents further semantic processing for unknown letter sequences in humans (Gagl, Richlan, et al., 2020). Previously, visual-orthographic prediction errors have been shown to correlate with posterior occipitotemporal brain activation (Gagl, Sassenhagen, et al., 2020) and letter-based prediction errors with parietal brain activation (Gagl, Davis, et al., 2020). Finally, lexical categorization difficulty was associated with the activation in the so-called visual word form area during word recognition (Gagl, Richlan, et al., 2020). Additionally, both the prediction errors and the word/non-word categorization process have been shown to successfully explain the variance in behavioral data from humans performing a lexical decision task similar to the baboon experiment (e.g., Gagl, Sassenhagen, et al., 2020). As such, the prediction errors, as well as the categorization process assumed in the Speechless Reader Model have been successfully correlated with behavioral and neurophysiological data. We therefore argue that the Speechless Reader Model offers a suitable alternative to domain-general learning models, as it rests upon domain-specific assumptions and evidence.

Based on the mentioned components, we assume the following hierarchical structure of the model:

1. Orthographic prediction error estimation based on the visual image of words.

2. Letter-based and sequence-sensitive prediction error estimation.
3. Probability estimation for categorizing a letter string as “word” by accumulating the prediction errors computed for all elements before.
4. Lexical categorization, i.e., deciding if the letter combination is a word.

In Figure 1, we present the structure of the Speechless Reader Model.

The visual-level prediction error is estimated based on all words that have been entered into the lexicon, which are transformed into grayscale images. After that, the model calculates a pixel-by-pixel mean across all known word images. When a specific pixel is black in a high number of previously encountered words, for instance, the model (or a human reader) will expect this pixel to be black with a high probability in any upcoming word. Gagl, Sassenhagen, et al. (2020) provides a more detailed description of the estimation of the visual-only image-based prediction. We furthermore implemented two prediction errors on the level of letters: (i) one based on the frequency of a letter in one of the four possible positions (e.g., the probability of encountering an "i" in the first position of a word), and (ii) one based on the sequential appearance of letters (the transitional probabilities between adjacent letters, i.e. the probability of encountering a "g" in second position after an "i" was the initial letter in the sequence).

All words already included in the lexicon are considered, and from these, the model estimates how often an individual letter is present in one of the four positions. For instance, the letters "e" and "s" are likely to occur at the end of English words. They would therefore receive relatively high values in the prediction for the final position if the model was trained on an entire English lexicon. The second letter-based prediction is sensitive to both letter frequencies and letter sequence. Suppose, for instance, the first letter of a word is an "s" and the model is now tasked with predicting the following, second letter. It will now not merely base its prediction on how frequently each possible candidate appears in the second position (as in i), but rather checks all known words for how often each candidate letter follows the initial "s". In other words, for letter-based prediction error that accounts for the letter sequence, the model will consider only words in the lexicon with the same preceding letter (or sequence of letters) to estimate the frequency of the upcoming letter(s). Thus, this letter-based prediction also includes the transitional probabilities between letters, i.e., the reader's knowledge about letter sequences (see Gagl, Davis, et al., 2020 for more details).

The estimations for both image- and letter-level prediction errors similarly follow the computational principles described by the predictive coding theory (i.e., top-down predictions to achieve efficient bottom-up prediction error processing, see Clark, 2013 for more details). Accordingly, the model combines all word knowledge to predict the upcoming letter sequence before a stimulus is presented. These predictions are based on all words that have been included in the lexicon so far. There are no words in the lexicon initially, but the model will learn more and more words with training. A word does not become part of the lexicon after the first presentation, but only after it has been correctly categorized as a word by the baboon at least five times out of seven presentations (71% accuracy). Accuracy rates for each word are continuously updated throughout training and thereby also allow for the possibility that the monkey might forget an already learned word. The criterion implemented for forgetting is that if the baboon repeatedly categorizes a word as a non-word, and accuracy drops below 71% for this word, the model will remove the word from the lexicon. Consequently, the size of the lexicon is continuously changing and prediction errors are recalculated for each new stimulus.

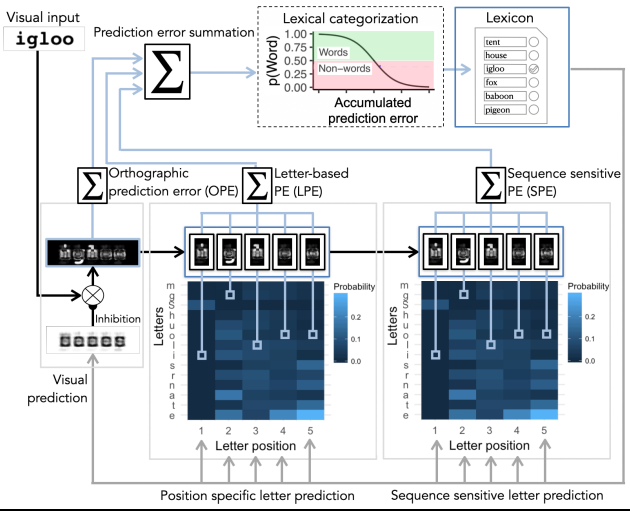

Figure 1: Schematic depiction of the Speechless Reader Model, including all three prediction error representations (visual level orthographic prediction error, OPE; letter-based prediction error, LPE; letter-based sequence sensitive prediction error, SPE). Gray arrows depict the predictions that originate from all words included in the lexicon. Black arrows represent the input to the model parts, and blue arrows show prediction error representation flow. Sum signs indicate prediction error accumulation, e.g., accumulating the individual pixel values from image-based prediction errors (OPE). The scaled sum of all available prediction errors enters the lexical categorization stage that differentiates between words and non-words.

Until now, we have only described the predictions implemented in the model (depicted by the gray arrows in Figure 1), i.e., the model's state before a word is presented. When a word is shown to the model, the predictions inhibit the expected sensory input before processing. Thus, only the residuals, i.e., the unpredicted information, have to be processed. This process increases neuronal processing efficiency (e.g., Price & Devlin, 2011). The orthographic prediction error is implemented by a pixel-by-pixel subtraction of the sensory input image from the prediction computed from on all stored images. For the letter-based

predictions, we subtract the letter frequency at a specific position (which is a value between 0 and 1) from 1, which would describe the situation that a letter cannot be predicted (i.e., highest prediction error possible). The prediction errors are then summated within each level. That is, for image-based prediction errors, the pixel-by-pixel values are added; for letter-based prediction errors, the values of all four letters are summed. In the models that include more than one prediction error, all prediction errors are transformed to the same scale and again summated (see sum signs and blue arrows in Figure 1).

The final step implements lexical categorization based on the accumulated prediction errors (see Summerfield & de Lange, 2014 for a similar idea). Here we estimate, for all models with different prediction error combinations individually, the probability that a presented four-letter combination is a word based on the distribution of prediction errors from all previously encountered words and a subset of non-words of equal size. The expectation here is that accumulated prediction errors for words should be smaller than for non-words, since they are based on learned words and training should therefore improve prediction accuracy (Gagl, Sassenhagen, et al., 2020; Price & Devlin, 2011). In contrast, the orthographic appearance, letter combinations and/or letter sequences of non-words should be unexpected and produce larger (accumulated) prediction errors. As a result, the distributions of prediction errors for words and non-words should not or only partially overlap (Balota & Chumbley, 1984). Thus, one can estimate the probability of a given input representing a word, p(Word), based on its prediction error value (see Figure 1; see Gagl, Richlan, et al., 2020 for more details this probability estimation based on classical orthographic similarity measures). To evaluate model performance on lexical decision accuracy, we fitted a decision boundary. This boundary allowed the model to decide if a presented letter string is a word or not.

Methods

Dataset. The present investigation used the baboon lexical decision data by Grainger et al. (2012). The dataset comprises accuracy data from more than 300,000 individual word and non-word trials of six baboons performing a lexical categorization task.

Simulations. We compared simulations from four model implementations with different sets of prediction errors:

1. Orthographic prediction error only (O PE Simulation).
2. OPE plus the letter-based prediction error (OL PE Simulation) .
3. OPE plus the letter-based sequence sensitive prediction error (OS PE Simulation).
4. One model combining all three prediction errors (OLS PE Simulation).

Comparing these four implementations allows us to draw conclusions about which of these prediction errors (or combination thereof) are likely involved in baboon lexical decision. We consider the underlying representation necessary for the computation of the orthographic prediction error (O PE) to be the one with the lowest complexity since it is only based on the words' visual features. The letter-based prediction error (OL PE) is of higher complexity because it includes the ability to store and recognize individual letters. The sequence-sensitive prediction error (OS PE) is then even more elaborate, as all previously encountered letter combinations are stored and considered for PE computation.

In addition, we used the simulations by Linke et al., (2017) as a control model to compare with our four Speechless Reader Model simulations. R-based programs estimated model performance by calculating trial-specific predictions based on the presented words up to a particular trial. In preparation for each trial, the model dynamically updates the lexicon that updates all predictions. When the model gets new input from the next stimulus, it calculates prediction errors based on the current status of predictions set before. After that, the prediction errors are combined as the accumulated prediction error. This parameter then is the basis for calculating the probability of the input being a word (p(Word)). Based on this parameter, the model implements the lexical categorization. Finally, one parameter, the decision boundary for lexical categorization, is set to .5. Hence, whenever the probability p(Word) exceeds .5, the lexical categorization results in a "word" response; whenever it is lower, the model response is "non-word".

Analysis. We implemented statistical model comparisons and the comparison of simulation and baboon data based on generalized linear mixed models (using the glmer() function of the lme4 package in R). First, we estimated a baseline model that included stimulus category (word vs. non-word) and the normalized log.-transformed trial number as fixed effects. We normalized the trial number to obtain numeric values in a smaller range and we log-transformed the value to account for the log.-shaped learning effect shown in the original baboon study (e.g., Grainger et al., 2012). As random effects on the intercept, we included stimulus and baboon. Since the data was binary (0 for incorrect and 1 for correct), we assumed a binomial distribution for our statistical analysis. We added the simulated model performances from our four Speechless reader implementations as fixed effects to the statistical models. We compared the Akaike information criterion (AIC) from the statistical models against the baseline model to learn if model simulations in general increased model fit and to compare them against each other, to discover which model assumptions represented the data best.

Results

When comparing the performance data of all four Speechless Reader simulations (Figure 2, plotted in blue), they all follow a similar, classical learning pattern: a steep initial increase that levels out. Compared amongst each other, the four

alternative prediction error-based models result in different behaviors, and model simulations differ individually for all baboons (due to different individual stimulus sequences, cf. Grainger et al. 2012). With increasing experience, the models' performance implementing the simpler prediction errors, i.e., O PE only and OL PE, dropped below the baboon's performance (cp. Red vs. dark blue lines in Figure 2). The average accuracy achieved by these two models was at 59% and 65%, respectively. In contrast, the two models that implemented the sequence-specific prediction error (OS PE and OLS PE, light blue lines in Figure 2) performed better than the models without and were closer to the baboons' accuracy rate (74%) with average accuracies of 74% and 77%, respectively. Interestingly, these latter two models also fitted the baboon performance better than the model by Linke et al. (2017; yellow graph in Figure 2). Their domain-general model exhibits a much higher performance than both the baboons and all of our simulations, particularly in early trials (average accuracy: 88%). Thus, the learning curve of their model is much steeper than that of both the baboons and our Speechless Reader Model implementations.

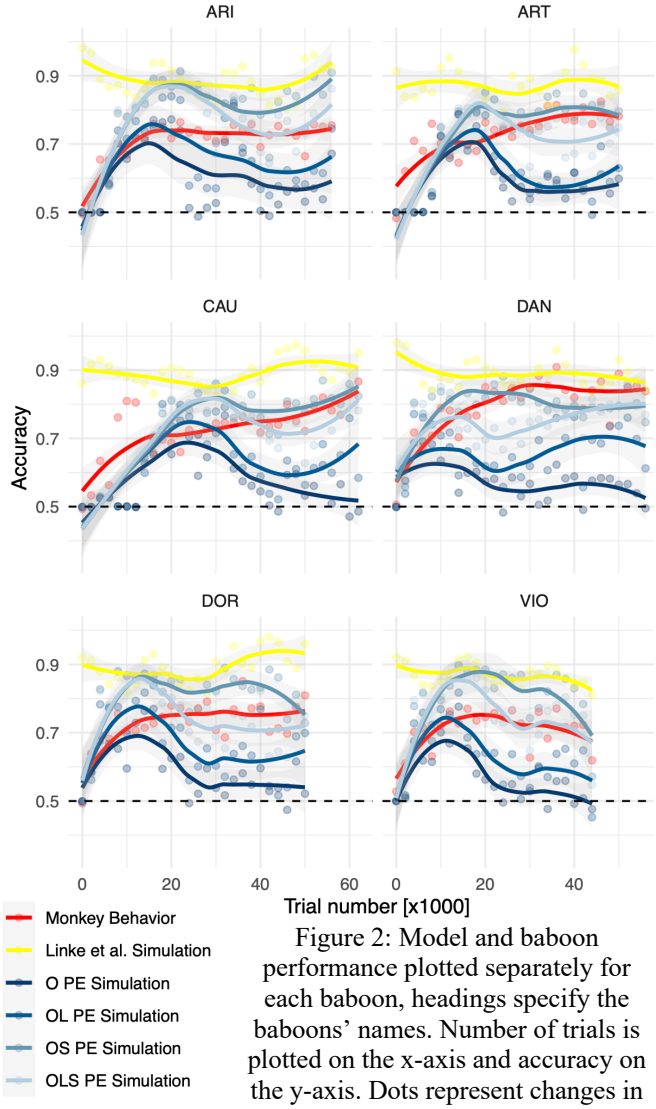

Figure 2: Model and baboon performance plotted separately for each baboon, headings specify the baboons' names. Number of trials is plotted on the x-axis and accuracy on the y-axis. Dots represent changes in mean accuracy across 2,000 trials. Red lines show the individual baboon's performance, respectively; blue lines show the performance of the four Speechless Reader Model simulations; and the yellow lines indicate simulation results by Linke et al. (2017).

In the generalized linear mixed model analysis of the baboon data, we found significant effects of the log. trial number (Estimate (Est.) = 0.03; Standard error (SE) = 0.01; t =4.1; p <.001) and stimulus category (Est. = 0.65; SE = 0.06; t =10.6; p <.001), reproducing previously published analyses (Grainger et al., 2012). Entering the Speechless Reader and domain-general models' performances as parameters into the models resulted only in small numerical changes of these two effects. For the model parameters, we found significant positive effects on baboon performance for all models (O PE model: Est. = 0.43; SE = 0.02; t = 23.6; p < .001; OL PE model: Est. = 0.86; SE = 0.03; t = 34.0; p < .001; OS PE model: Est. = 1.59; SE = 0.02; t = 78.1; p < .001; OLS PE model: Est. = 1.48; SE = 0.02; t = 60.5; p < .001; Linke model: Est. = 0.18; SE = 0.02; t = 9.2; p < .001). In model comparisons (Figure 3), we found that all models explained more variance than the baseline model (i.e., all Chi-Squared > 221; all p's < .001). The OS PE model explained more variance than all other contrasted models (i.e., all Chi-Squared > 115; all p's < .001). This relative difference was present for all phases of the experiment (i.e., when tested in steps of 10,000 trials) and for each but one baboon individually. When running the model comparisons for each baboon separately, we discovered that for five baboons (ART, ARI, DAN, DOR, VIO), the relative differences were the same as in the overall sample (see Fig. 3). For CAU, the OLS PE model including all prediction errors was slightly better than the OS PE model.

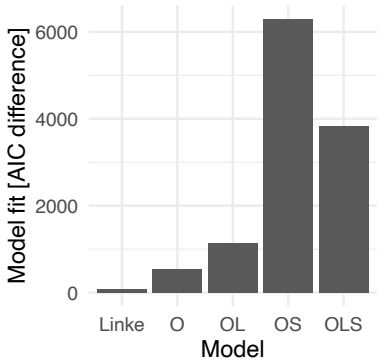

Figure 3: Model fits of all four models (OPE only model: O; OPE plus LPE: OL; OPE plus SPE: OS; OPE plus LPE plus: OLS), in AIC differences in relation to a baseline model without the model parameters, using the entire dataset.

Discussion

In the present study, we investigated the mental representations and algorithms on which baboons base their lexical decision behavior. We implemented a cognitive

model, the Speechless Reader Model, which combines previous component models of human visual word recognition computationally. The model allowed us to investigate if baboons implement mere image-based or higher-level, letter-based prediction error representations to achieve high performance in lexical decisions. We found that all model simulations correlated positively with the baboon performance. The model implementation combining the image-based prediction error and the letter sequence-based prediction error explained the largest amount of variance in baboon behavior overall. In other words, our simulations clearly show that baboons' representations most likely go beyond mere storage and processing of a large number of low-level visual features representing the written input, as suggested by Linke et al. (2017). It seems rather that in addition to such visual features (which one might already classify as orthographic, for a discussion, see Gagl, Sassenhagen, et al., 2020), the monkeys also rely on representing the transitional probabilities from letter-to-letter in a letter sequence.

The finding that monkeys are sensitive to transitional probabilities in linguistic material is not new. Several studies investigating different primate species show behaviors that can be explained by representations that implement transitional probabilities (for reviews, see Santolin & Saffran, 2017; Wilson, Marslen-Wilson & Petkov, 2017). Specifically, baboons can learn even non-adjacent dependency relationships between temporally separated units in the visual modality (Malassis, Rey & Fagot, 2018).

New here is that we used a model to explain the neurocognitive processes underlying human reading (Gagl, Davis, et al., 2020; Gagl, Richlan, et al., 2020; Gagl, Sassenhagen, et al., 2020) to model baboon lexical decision performance. This implementation showed that to solve lexical decisions, baboons likely use prediction error representations similarly to how humans use these representations during visual word recognition. Even if these processes are likely different from human "linguistically" motivated reading, they seem to share certain features with the reading process (Grainger et al., 2012). Even more so, in successfully modeling the baboon data with the Speechless Reader Model, we demonstrated that this model specific to the domain of reading outperformed a domain-general model such as the one by Linke et al. (2017). This finding is surprising because one would not expect reading-specific representations and algorithms to be implemented by animals lacking phonological and semantic, i.e., linguistic, information. These findings suggest that some aspects that have been considered and modeled as pertaining exclusively to the linguistic domain may originate from more domain-general capacities shared by humans and primates. Thus, human reading could be rooted in the cognitive capabilities shared with our closest phylogenetic relatives.

When comparing previous modeling studies with the present one, domain-general models (e.g., Hannagan et al., 2014; Linke et al., 2017), using procedures that make critical model representations intransparent, could hardly identify which cognitive processes underly baboon behavior. These models implement a high number of parameters fitted to a specific input so that it is almost impossible to infer which information is actually represented (e.g., see Güçlü & Gerven, 2015 for examples from object recognition). In contrast, the Speechless Reader Model does not include any free parameters, and only the predictions are adapted when a new word enters the lexicon. This way, one can trace each prediction or prediction error value down to the words included in the lexicon and eventually to the visual input of the model.

Even though these previous models were able to extract somewhat critical information from the visual features of the letter strings to differentiate between words and non-words, they ultimately performed much better in the categorization task than both the baboons and our Speechless Reader Model. In other words, these implementations are not necessarily representative of the processes implemented by baboons. However, the Speechless Reader Model (i) better represented the baboon behavior as its categorization accuracy was more similar to baboons, and (ii) the simulations could explain more of the variance in the baboon data. Notably, the transparent representations of the Speechless Reader Model further allowed specifying a likely combination of prediction error representations and categorization algorithms implemented by baboons.

This investigation of baboon behavior with the help of a transparent reading model will now motivate new studies, including further studies on human reading. A critical difference between baboons and efficient human readers is that human lexical decision accuracy can be much higher (i.e., around 90% correct; Balota et al., 2004). So that new studies could investigate if this increase in accuracy can be explained by additionally considering phonological and semantic information for the Speechless Reader implementation. Interesting could also be the investigations of beginning readers, dyslexic readers, and adults at different ages. Of significant interest will be how the different prediction error representations change during development or are dependent on individual differences such as reading experience or reading speed.

In sum, this Speechless Reader Model-based investigation showed that baboons implement image- and letter-based representations, including transitional probabilities, to achieve relatively high lexical decision performance. The combination of prediction error style representations and evidence accumulation has proven to be a valuable tool to model baboon lexical categorization behavior. Since the model components have been motivated by human behavior and neuronal processes, the current investigation suggests that both human and baboon behavior can in part rely on a similar set of representations and algorithms.

Acknowledgements

We very much thank Maja Linke for providing code, data and simulations.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Cichy, R. M., & Kaiser, D. (2019). Deep Neural Networks as Scientific Models. *Trends in Cognitive Sciences*, *23*(4), 305–317. https://doi.org/10.1016/j.tics.2019.01.009

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Gagl, B., Davis, M. H., & Fiebach, C. J. (2020, October 21). *Visual word recognition involves an abstract letter-based prediction error in posterior vOT*. Meeting of the Society for the Neurobiology of Language. https://doi.org/10.13140/RG.2.2.33416.14086

Gagl, B., Richlan, F., Ludersdorfer, P., Sassenhagen, J., Eisenhauer, S., Gregorova, K., & Fiebach, C. J. (2020). The lexical categorization model: A computational model of left-ventral occipito-temporal cortex activation in visual word recognition. *BioRxiv*, 085332. https://doi.org/10.1101/085332

Gagl, B., Sassenhagen, J., Haan, S., Gregorova, K., Richlan, F., & Fiebach, C. J. (2020). An orthographic prediction error as the basis for efficient visual word recognition. *NeuroImage*, *214*, 116727. https://doi.org/10.1016/j.neuroimage.2020.116727

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic Processing in Baboons (Papio papio). *Science*, *336*(6078), 245–248. https://doi.org/10.1126/science.1218152

Güçlü, U., & Gerven, M. A. J. van. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Hannagan, T., Ziegler, J. C., Dufau, S., Fagot, J., & Grainger, J. (2014). Deep Learning of Orthographic Representations in Baboons. *PLOS ONE*, *9*(1), e84843. https://doi.org/10.1371/journal.pone.0084843

Katz, L., Perfetti, C. A., & Shankweiler, D. (2012). Reading Too Much Into Baboon Skills? *Science*, *336*(6085), 1100–1100. https://doi.org/10.1126/science.336.6085.1100-a

Linke, M., Bröker, F., Ramscar, M., & Baayen, H. (2017). Are baboons learning "orthographic" representations? Probably not. *PLOS ONE*, *12*(8), e0183876. https://doi.org/10.1371/journal.pone.0183876

Ma, W. J., & Peters, B. (2020). A neural network walks into a lab: Towards using deep nets as models for human behavior. *ArXiv:2005.02181 [Cs, q-Bio]*. http://arxiv.org/abs/2005.02181

Malassis, R., Rey, A., & Fagot, J. (2018). Non-adjacent dependencies processing in human and non-human primates. Cognitive Science, 42, 1677–1699.

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, *15*(6), 246–253. https://doi.org/10.1016/j.tics.2011.04.001

Rajalingham, R., Kar, K., Sanghavi, S., Dehaene, S., & DiCarlo, J. J. (2020). The inferior temporal cortex is a potential cortical precursor of orthographic processing in untrained monkeys. *Nature Communications*, *11*(1), 3886. https://doi.org/10.1038/s41467-020-17714-3

Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of rein- forcement and nonreinforcement. In: Black AH, Prokasy WF, editors. Classical conditioning II: Current research and theory. New York: Appleton Century Crofts; 1972. p. 64–99.

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (Papio Papio). *Cognition*, *123*(1), 180–184. https://doi.org/10.1016/j.cognition.2011.12.005

Santolin, C., & Saffran, J. R. (2017). Constraints on statistical learning across species. Trends in Cognitive Sciences, 22(1), 52–63.

Scarf, D., Boy, K., Reinert, A. U., Devine, J., Güntürkün, O., & Colombo, M. (2016). Orthographic processing in pigeons (Columba livia). *Proceedings of the National Academy of Sciences*, *113*(40), 11272–11276. https://doi.org/10.1073/pnas.1607870113

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*, *18*(3), 733–745. https://doi.org/10.1007/s10071-015-0840-x

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Wilson, B., Marslen-Wilson, W. D., & Petkov, C. I. (2017). Conserved sequence processing in primate frontal cortex. Trends in Neurosciences, 40,72–82.