Under review at Psychological Science
August 1, 2013

1.2.0.0

Evidence for Orthographic Processing in Baboons *(Papio papio)*? A Visual Familiarity Account

John R. Vokey
University of Lethbridge

Randall K. Jamieson
University of Manitoba

Grainger, Dufau, Montant, Ziegler, and Fagot (2012a) taught six baboons to discriminate words from nonwords in an analogue of the lexical decision task. The baboons endorsed novel words more readily than novel nonwords, and had trouble rejecting nonwords that were orthographically similar to learned words. In a subsequent test (Ziegler, Hannagan, et al., 2013), the same animals demonstrated the transposed-letter effect. All three results are often taken as hallmarks of orthographic processing in skilled human readers. We show by simulation of the unique learning trajectory of each of the baboons that the results can be interpreted equally well as an example of simple, familiarity-based discrimination of pixel-maps, without orthographic processing.

Keywords: baboons, familiarity, orthographic processing

Grainger et al. (2012a) taught baboons to discriminate words from nonwords in an analogue of the lexical decision task. They argued that their baboons learned the discrimination by orthographic processing (i.e., computation of letter identities and their relative positions). They named two key pieces of evidence to support the claim: (a) their baboons endorsed novel words more readily than novel nonwords (i.e., learning generalized to unstudied items) and (b) their baboons had trouble rejecting nonwords that were orthographically similar to the words that they had already learned (i.e., a feature of orthographic processing in skilled human readers). Based on the evidence, Grainger et al. (2012a) concluded that orthographic processing precedes language and that the primate brain is better prepared to process written language than previously thought (cf. Platt & Adams, 2012).

Bains (2012) has since demonstrated that Grainger et al.'s (2012a) baboons could have discriminated words from nonwords by recognizing single letters (i.e., shapes) in specific serial positions. Therefore, the evidence does not force the conclusion of human-like orthographic processing. In response, Grainger, Dufau, Montant, Ziegler, and Fagot (2012b) argued that the more limited processing of Bains's (2012) model is still consistent with their original claim that the baboons were discriminating words from nonwords by recognizing letters in position, and that, in any case, Bains (2012) did not demonstrate whether his model would generalize the discrimination to novel words and novel nonwords, nor whether it would show the same orthographic similarity effects with nonwords.

The argument of Grainger et al. (2012a, 2012b) rests on the assumption that their baboons treated the stimuli as horizontal arrays of discrete symbols rather than as whole pictures. But, they offered no direct evidence to corroborate that assumption. Furthermore, the experimental design confounds category (i.e., word/nonword) with category frequency: each word in a given baboon's list of learned words was presented at least 80 times whereas each nonword was presented often only once. This frequency confound could allow the baboons to discriminate words and nonwords by simple visual familiarity to known words, and without orthographic processing. The subsequent research of Ziegler, Hannagan, et al. (2013) was intended to challenge that possibility.

Ziegler, Hannagan, et al. (2013) extended the work of Grainger et al. (2012a) by presenting the same baboons with two further sets of nonword stimuli. The first set compared responding to transposed letter (TL) nonwords with that to double-substitution (DS) nonwords. The TL items were created by transposing the two internal letters of learned words (e.g., DONE → DNOE) to produce a nonword, and DS items were created by substituting each of the two internal letters of learned words with letters of the same kind (i.e., vowels and consonants) that produced a nonword (e.g., DONE → DAGE). A higher false-positive rate to TL nonwords than to DS words is referred to as the *TL effect* and is said to be indicative of orthographic processing in humans (Grainger, 2008). The second new set of nonwords compared responding to visually similar (VS) nonwords with that to visually dissimilar (VD) nonwords. VS items were formed by selecting at random one of the two internal letters of learned words and replacing it with the most visually similar letter that produced a nonword

---

This work was supported by Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery grants to each of the authors. Requests for reprints should be sent to vokey@uleth.ca, as should requests for program code, algorithms, and the exact materials used.

(e.g., DONE → DQNE). VD items were formed similarly, except that the least visually similar letter that produced a nonword was substituted (e.g., DONE → DFNE). The comparison of VS with VD nonwords was intended to control for simple visual similarity as the basis of responding. If the baboons evinced a TL effect but no *similarity* effect (i.e., similar responding to both VS and VD items) then that would strengthen the case that the baboons were responding to the orthographic rather than the simple visual similarity of the items. And, that is precisely what Ziegler, Hannagan, et al. (2013) found: a large TL effect and no similarity effect.

A Visual Familiarity Account

Although the results of both Grainger et al. (2012a) and Ziegler, Hannagan, et al. (2013) are consistent with the possibility that the baboons were engaging in some form of orthographic processing (see Ziegler, Dufau, et al., 2013—written in response to a critique by Frost & Keuleers, 2013—for the latest, precise adumbration of the claims), we are still not convinced that a similar pattern of results could not be obtained with simple visual familiarity. After all, this one example of a negative similarity effect may reflect nothing more than that this method of creating the VS and VD items fails to capture the sources of visual similarity the baboons actually use. Furthermore, although the TL effect is generally thought to reflect orthographic processing, it is still possible that some model of visual familiarity can do so, as well. We constructed just such a visual familiarity model, and used it to simulate the results of the baboons in both experiments.

Simulating Grainger et al. (2012a)

To evaluate a visual familiarity account, we used a principal components analysis (PCA), autoassociative neural network model of memory that has been used with many different kinds of visual materials—from human faces (Abdi, Valentin, & Edelman, 1999; Turk & Pentland, 1991; Vokey & Hockley, 2012) and chimpanzee faces (Vokey, Rendall, Tangen, Parr, & de Wall, 2004) to fingerprints (Vokey, Tangen, & Cole, 2009) and artificial grammar letter-strings (Vokey & Higham, 2004). As with the artificial grammar letter-strings of Vokey and Higham (2004), we applied the same approach to the letter-string materials of Grainger et al. (2012a) and Ziegler, Hannagan, et al. (2013). In this apprach, letter-strings are represented as pictures that were constructed by drawing an item into a 28 by 5 black-and-white pixel map—where each letter appeared as a 7 by 5 dot-matrix character. These pixelmaps were then converted into 140-element column vectors by assigning values 1 and 0 to elements corresponding to filled and unfilled pixels, respectively. Next, for each baboon in Grainger et al. (2012a), we constructed an autoassociative memory of the n learned words for a given baboon by (a) forming a $140 \times n$ stimulus matrix, $\mathbf{X}$, that stored the representations of all n words the baboon had learned according to the criteria in Grainger et al. (2012a), (b) performing singular value decomposition (SVD) of $\mathbf{X}$ to obtain the matrix, $\mathbf{U}$ (the matrix of eigenvectors of $\mathbf{XX}^{\mathrm{T}}$), and (c) forming the autoassociative memory matrix, $\mathbf{W}$, where $\mathbf{W} = \mathbf{UU}^{\mathrm{T}}$. The model is equivalent to a linear autoassociative neural network trained with Widrow-Hoff learning, and in statistics is equivalent to the PCA of the original data matrix (e.g., Abdi et al., 1999).

Finally (see Vokey & Hockley, 2012, for more details), the familiarity of each tested letter string, $\mathbf{x}_i$, for a given baboon, was computed relative to its $\mathbf{W}$ as $\cos(\mathbf{x}_i, \hat{\mathbf{x}}_i)$, where $\hat{\mathbf{x}}_i = \mathbf{U}_{1:m}(\mathbf{U}^{\mathrm{T}}_{1:m}\mathbf{x}_i)$ is the projection of $\mathbf{x}_i$ into the space defined by the $1{:}m$ eigenvectors in $\mathbf{U}$ and where the eigenvectors in $\mathbf{U}$ are ordered from first to last according to the descending magnitudes of their associated eigenvalues (Abdi et al., 1999). If the cosine familiarity for a given letter-string exceeded the criterion defined by the midpoint between the mean cosine familiarity of words and that of nonwords for a given baboon, the item was identified as a word; else, it was identified as a nonword for that baboon. When computing the cosine familiarity of a word, the "leave one out" technique was used (cf. Abdi et al., 1999): the word was removed from $\mathbf{X}$ prior to constructing the autoassociative memory, leaving the remaining words to serve as the simulated baboon's memory for words in the experiment; for nonwords, $\mathbf{X}$ was left intact. Thus, $\cos(\mathbf{x}_i, \hat{\mathbf{x}}_i)$ represents an item's familiarity as a novel item in the experiment for both words and nonwords for that baboon.

Results from the simulation of Grainger et al. (2012a) are presented in Figures 1, 2, and 3. Figure 1 shows the model's discrimination of words from nonwords for each of the six simulated animals as a function of the number, m, of eigenvectors ($1{:}m$) used to reconstruct the test probe, $\hat{\mathbf{x}}_i$, from each simulated animal's autoassociative memory of learned words. As shown, the functions for each animal appear to asymptote at a high level of discrimination (i.e., A' of roughly .80) with as few as 10 eigenvectors. Accordingly, the first two bars in Figure 2 depict the mean hit and false-positive rates for novel words and novel nonwords, respectively, derived from the autoassociative memory for each simulated animal based on the first 10 eigenvectors. As shown, even with as few as 10 eigenvectors to reconstruct the test probe, the model successfully discriminated novel words from novel nonwords [$F(1, 5) = 1447.20$, MSE $= .0005, p < .0001, r^2_{pb} = .97$]—a result that Grainger et al. (2012a, 2012b) interpreted as evidence of orthographic processing, but one that the model reproduces without orthographic processing.

Figure 3 shows the model's false-positive rate (based on the same first 10 eigenvectors) for each simulated animal to the 7,832 nonwords each animal received as a function of orthographic similarity [using the same measure of Orthographic Levenshtein Distance (OLD20) (Keuleers, 2011; Yarkoni, Balota, & Yap, 2008) that Grainger et al. (2012a) used] to the learned words for that simulated animal. As shown, the more

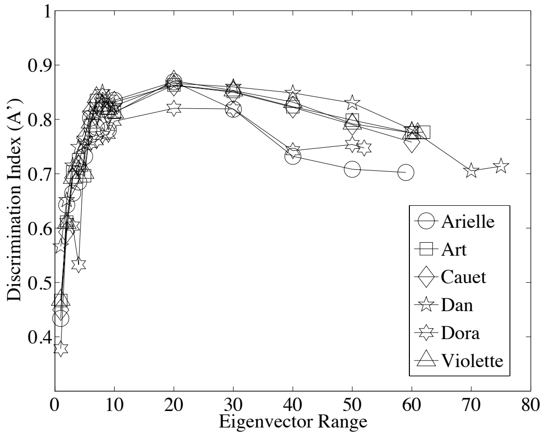

Figure 1. Discrimination of novel words and novel nonwords as a function of the eigenvector range (1:*m*) used to construct the autoassociative memory for each of the six simulated animals. Because each of the animals learned different numbers of words (e.g., Dan the most and Dora the fewest), the total number of eigenvectors available to construct the autoassociative memory for each also necessarily varied. Discrimination is expressed as A', an unbiased non-parametric measure of discrimination. A' varies from 0 and 1, with values greater than .5 indicating increasingly good discrimination.

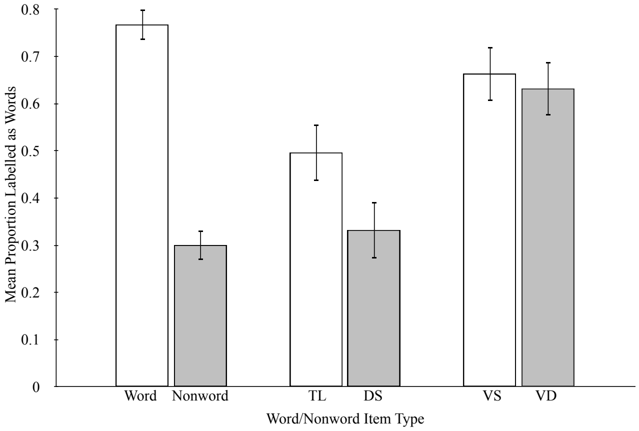

Figure 2. Mean hit and false-positive rates to novel words and the different types of novel nonwords derived from the autoassociative memory for each simulated animal based on the first 10 eigenvectors. Error-bars are Fisher LSD$_{.05}$ values for each comparison.

orthographically-similar (in the OLD20 sense) a nonword was to the learned words, the more false-positive responses it elicited—a result that Grainger et al. (2012a, 2012b) took as evidence for orthographic processing, but that, again, also falls out of familiarity-based discrimination.

Simulating Ziegler, Hannagan, et al. (2013)

We applied the same model to the new materials in Ziegler, Hannagan, et al. (2013), using the previously-established decision criterion for each simulated animal from the simulation of the original Grainger et al. (2012a) experiment. TL and DS items were constructed as described in Ziegler, Hannagan, et al. (2013). For many of the TL items, the transposition operation produces a word (e.g., SANG → SNAG, or BOOK → BOOK), so they had to be removed from the simulation for a given baboon, reducing the number of VS nonwords for the simulation of that baboon. For DS items, the substitution operation also could still result in a word, so the algorithm was filtered by a dictionary (the British Dictionary 2.2 for the Excalibur spell-checker) to iterate to produce just nonwords.

The creation of VS and VD items presented a different problem: how to define the visual similarity of the letters. All Ziegler, Hannagan, et al. (2013) provided by way of explanation (in the supplementary materials) was that they had used the *Hausdorff* distance measure, but no further information as to what representations of the letters the measure was applied (e.g., scans of the letters they used with the baboons as pixel-maps, the actual vector graphic computer definition of the letters, etc., some generic letter definitions, and so on). The *Hausdorff* distance measure can be applied to many such representations (it requires only that the points of the objects be scored on the same dimensions)—some of which could be tightly-tied to the actual spatial qualities of a specific font at a specific size, and some more abstract. Clearly, we could have used our simple cosine familiarity metric applied to the raw pixel-maps we defined for the individual letters (or the *Hausdorff* distance measure applied to the same representations), and such approaches likely would produce a large similarity effect. But, it is not clear that such metrics at the level of individual letters is relevant to the visual processing of the baboons (or humans, for that matter) for whole items. For example, in our model, the reconstruction of an item that differs in only one letter from a learned item does *not* differ from that item in just the reconstruction of that one letter—the effect is spread over all of the reconstructed pixels of the item. Indeed, the metric of individual letters of Ziegler, Hannagan, et al. (2013) appears to assume that the baboons *do* process the items as individual letters, as does the VS vs. VD comparison more generally. From that perspective, the null similarity effect in Ziegler, Hannagan, et al. (2013), which they took as evidence of orthographic processing, could just as easily be seen as evidence that the baboons do not respond to the items as being composed of individual letters.

Given that ambiguity, and the fact that baboons may evince differential sensitivity to different measures of the visual simi-

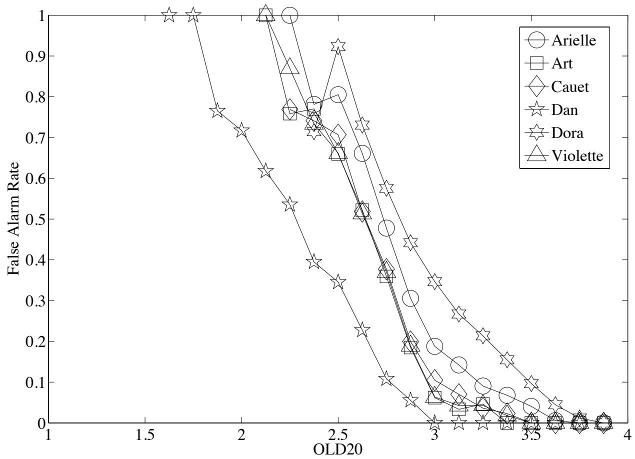

Figure 3. False-alarm rate (based on the first 10 eigenvectors) for nonwords (i.e., identification of nonwords as words) for each of the six simulated animals as a function of Orthographic Levenshtein Distance (OLD20) (Keuleers, 2011; R Core Team, 2012; Yarkoni et al., 2008) of the 7,832 nonwords to the learned set of words for each animal. OLD20 indexes a string's similarity to a comparison set as the mean edit distance of the nearest 20 exemplars in the comparison set. The greater a string's OLD20 is, the more different (in this case, the less word-like) it is from the comparison set.

larity of the letters in items (and that there is nothing privileging *Hausdorff* distance over any other), and the further fact that it is not yet established that the baboons actually respond to the letters of the items qua letters in the Ziegler, Hannagan, et al. (2013) task, we searched for something slightly less tied to the actual letter representations as we, or Ziegler, Hannagan, et al. (2013), used them. Ultimately, we settled on the (human) recognition letter confusion matrix of upper-case letters extensively compiled by Gilmore, Hersh, Caramazza, and Griffin (1979). We are *not* arguing that this matrix represents in any way the letter confusion matrix of baboons in the Ziegler, Hannagan, et al. (2013) task (if there even is one), only that it does capture the inter-letter confusions (for a particular representation of uppercase letters) of the species that Ziegler, Hannagan, et al. (2013) claim baboons emulate in this regard to some degree. From this matrix, we computed the rank ordering of similarity for each letter to each of the other letters. For each learned word for each baboon, we constructed VS and VD items as described in Ziegler, Hannagan, et al. (2013) by randomly choosing one of the two internal letters, and substituting the most (least) similar letter to that letter. If that resulted in a word (as filtered by the dictionary), the process was repeated, substituting the next most (least) similar letter until a nonword was produced.

The results are shown as the last four bars in Figure 2, which depict the false-positive rates for the simulation with the nonword item types in Ziegler, Hannagan, et al. (2013). As shown, and as in Ziegler, Hannagan, et al. (2013), there was a large TL effect, $F(1,5) = 48.74$, MSE $= .0017$, $p = .0009$, $r^2_{pb} = .91$, and no significant similarity effect, $F(1,5) = 2.01$, MSE $= .0015$, $p = .2153$, $r^2_{pb} = .29$.

Discussion

A standard, autoassociative model of memory applied to the materials of both Grainger et al. (2012a) and Ziegler, Hannagan, et al. (2013) reproduces the principal results that Grainger et al. (2012a), Grainger et al. (2012b), Ziegler, Hannagan, et al. (2013), and Ziegler, Dufau, et al. (2013) cited as evidence of orthographic processing in their baboons (i.e., recognition of letter identities in serial positions). Although we cannot rule out the possibility that their baboons performed orthographic processing, our demonstration of simple visual familiarity shows that the principal results can be explained equally well as an example of familiarity-based visual discrimination, and that, therefore, the results of both Grainger et al. (2012a) and Ziegler, Hannagan, et al. (2013) do *not* force the conclusion of orthographic processing in their baboons. But, we would argue, and contra Frost and Keuleers (2013), related results in the human domain are subject to the same conclusion. Ironically, we agree with Grainger et al. (2012a, 2012b), Ziegler, Hannagan, et al. (2013), and Ziegler, Dufau, et al. (2013) that skilled readers and baboons are likely to be relying on similar processes in these tasks, only because skilled readers are also probably not routinely relying on orthographic processing either, given the ease with which simple visual familiarity can accomplish lexical decisions (e.g., Brooks, 1978; Vokey & Brooks, 1992).

References

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural networks*. Thousand Oaks, CA: Sage Publications.

Bains, W. (2012). Comment on "Orthographic processing in baboons *(Papio papio)*". *Science*, *337*(6099), 1173. doi: 10.1126/science.1224508

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. *Cognition and categorization*, *3*, 170–211.

Frost, R., & Keuleers, E. (2013). What can we learn from monkeys about orthographic processing in humans? A reply to Ziegler et al. *Psychological science*.

Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, *25*, 425–431.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, *23*, 1—35.

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

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012b). Response to comment on "Orthographic Processing in Baboons *(Papio papio)*". *Science*, *337*(6099), 1173. doi: 10.1126/science.1224939

Keuleers, E. (2011). vwr: Useful functions for visual word recognition research [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=vwr` (R package version 0.1)

Platt, M. L., & Adams, G. K. (2012). Monkey see, monkey read. *Science*, *336*(6078), 168-169. doi: 10.1126/science.1221402

R Core Team. (2012). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/` (ISBN 3-900051-07-0)

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*, 71–86.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 328.

Vokey, J. R., & Higham, P. A. (2004, Sep). Opposition logic and neural network models in artificial grammar learning. *Consciousness and Cognition*, *13*(3), 565–578. doi: 10.1016/j.concog.2004.05.008

Vokey, J. R., & Hockley, W. E. (2012). Unmasking a shady mirror effect: recognition of normal versus obscured faces. *Quarterly Journal of Experimental Psychology*, *65*(4), 739–759. doi: 10.1080/17470218.2011.628399

Vokey, J. R., Rendall, D., Tangen, J. M., Parr, L. A., & de Wall, F. B. M. (2004). Visual kin recognition and family resemblance in chimpanzees *(Pan troglodytes)*. *Journal of Comparative Psychology*, *118*, 194–199.

Vokey, J. R., Tangen, J. M., & Cole, S. A. (2009, May). On the preliminary psychophysics of fingerprint identification. *Quarterly Journal of Experimental Psychology*, *62*(5), 1023–1040. doi: 10.1080/17470210802372987

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's *N*: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979. doi: 10.3758/PBR.15.5.971

Ziegler, J. C., Dufau, S., Montant, M., Hannagan, T., Fagot, J., & Grainger, J. (2013). What can we learn from humans about orthographic processing in monkeys? A reply to Frost and Keuleers (2013). *Psychological science*.

Ziegler, J. C., Hannagan, T., Dufau, S., Montant, M., Fagot, J., & Grainger, J. (2013). Transposed-letter effects reveal orthographic processing in baboons. *Psychological Science*. doi: 10.1177/0956797612474322