Report

Current Biology

A general-purpose mechanism of visual feature association in visual word identification and beyond

Highlights

- Readers presented with orthographic-like stimuli are sensitive to bigram frequencies
- An analogous effect emerges with images of made-up objects and visual gratings
- These data suggest that the reading system might rely on general-purpose mechanisms
- This calls for considering reading in the broader context of visual neuroscience

Authors

Yamil Vidal, Eva Viviani,
Davide Zoccolan, Davide Crepaldi

Correspondence

hvidaldossantos@gmail.com (Y.V.),
dcrepaldi@sissa.it (D.C.)

In Brief

Vidal et al. show that an effect usually studied in the context of reading—sensitivity to bigram frequencies—is also found when participants are presented with images of objects and circular sinusoidal gratings. This suggests that some mechanisms implied in the processing of novel words are in fact of general purpose.

Vidal et al., 2021, Current Biology *31*, 1–7
March 22, 2021
https://doi.org/10.1016/j.cub.2020.12.017

CellPress

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

Current Biology

CellPress

Report

A general-purpose mechanism of visual feature association in visual word identification and beyond

Yamil Vidal,[1,3,*] **Eva Viviani,**[1] **Davide Zoccolan,**[2] **and Davide Crepaldi**[1,4,5,*]
[1]Language, Learning and Reading Lab, International School for Advanced Studies (SISSA), Cognitive Neuroscience Sector, Via Bonomea 265, Trieste 34136, Italy
[2]Visual Neuroscience Lab, International School for Advanced Studies (SISSA), Cognitive Neuroscience Sector, Via Bonomea 265, Trieste 34136, Italy
[3]Twitter: @YamilVidal_arg
[4]Twitter: @CrepaldiDavide
[5]Lead Contact
*Correspondence: hvidaldossantos@gmail.com (Y.V.), dcrepaldi@sissa.it (D.C.)
https://doi.org/10.1016/j.cub.2020.12.017

SUMMARY

As writing systems are a relatively novel invention (slightly over 5 kya),[1] they could not have influenced the evolution of our species. Instead, reading might recycle evolutionary older mechanisms that originally supported other tasks[2,3] and preceded the emergence of written language. Accordingly, it has been shown that baboons and pigeons can be trained to distinguish words from nonwords based on orthographic regularities in letter co-occurrence.[4,5] This suggests that part of what is usually considered reading-specific processing could be performed by domain-general visual mechanisms. Here, we tested this hypothesis in humans: if the reading system relies on domain-general visual mechanisms, some of the effects that are often found with orthographic material should also be observable with non-orthographic visual stimuli. We performed three experiments using the same exact design but with visual stimuli that progressively departed from orthographic material. Subjects were passively familiarized with a set of composite visual items and tested in an oddball paradigm for their ability to detect novel stimuli. Participants showed robust sensitivity to the co-occurrence of features ("bigram" coding) with strings of letter-like symbols but also with made-up 3D objects and sinusoidal gratings. This suggests that the processing mechanisms involved in the visual recognition of novel words also support the recognition of other novel visual objects. These mechanisms would allow the visual system to capture statistical regularities in the visual environment.[6–9] We hope that this work will inspire models of reading that, although addressing its unique aspects, place it within the broader context of vision.

RESULTS

In this work, we propose that, if the human reading system relies on domain-general visual mechanisms, effects that are often found within the domain of visual word processing should also be observable with non-orthographic visual stimuli. We tested whether an effect typically studied in orthographic processing, i.e., participants' sensitivity to bigram frequencies, is also found when the stimuli consist of sequences of pseudofonts or arrangements of non-orthographic visual features. Bigrams have been proposed to serve as an intermediate step between single graphemes and words.[10–12] Although the role of bigram frequencies in reading is open to debate (see Chetail[13] and Schmalz and Mulatti[14] for two opposite views), fMRI[15,16] and human intracranial recordings[17] have shown that the left fusiform gyrus is sensitive to this factor. Together with the aforementioned animal research literature, which manipulated bigram frequencies as a variable of interest,[4,5,18] this makes sensitivity to bigram frequencies a suitable effect to be replicated with non-orthographic visual stimuli.

Participants are sensitive to the bigram frequencies of orthographic-like stimuli

Experiment 1 used words written using pseudofont as stimuli, which, although resembling real letters, are completely unfamiliar to the participants. Concretely, we used the Brussels artificial character set (BACS)[19] (specifically BACS-2 serif; Figure 1A). Although the use of novel words written with pseudofont is a considerable simplification compared to real-life reading, it also implies a number of important advantages. As pseudofonts lack any connection to phonology or meaning, results only depend on visual/orthographic processing. Furthermore, the use of an unfamiliar script allowed us to avoid the influence of participants' particular history of exposure to orthographic material. This approach has been widely and successfully used in the field of psycholinguistics to study diverse aspects of reading

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

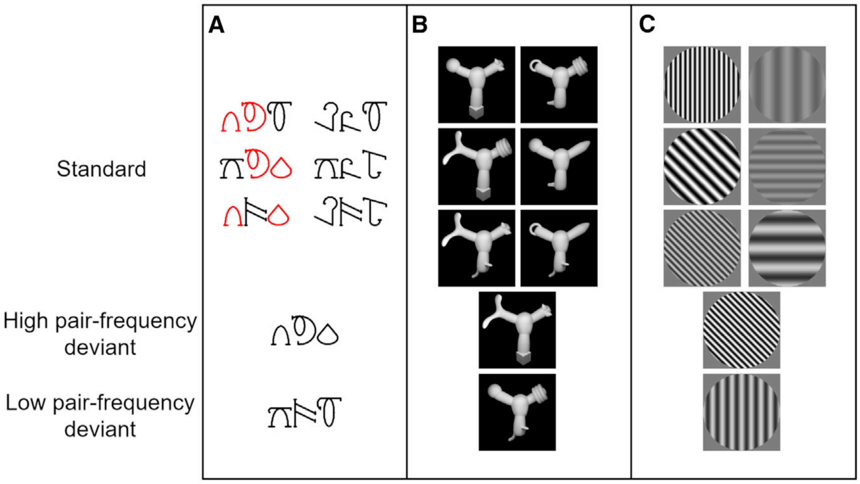

Figure 1. Examples of representative stimulus sets

(A) Three-character words used in experiment 1. Character pairs making up high-pair-frequency deviants are shared with standard words (here marked in red for illustration purposes).

(B) Objects used in experiment 2. Each pair of shapes composing high-pair-frequency deviant objects was shared with a standard object.

(C) Circular gratings used in experiment 3. Each pair of visual feature values (spatial frequency, orientation, and contrast) defining high-pair-frequency deviant gratings was shared with a standard grating.

See also Figure S1.

(e.g., Chetail,[13] Vidal et al.,[19] Maurer et al.,[20] and Taylor et al.[21],[22]).

A different subset of 9 BACS characters was randomly selected to construct the stimuli for each participant. These characters were used to build 6 three-character combinations (e.g., ABC), which were used as *standard* words. Next, two different *deviant* words were constructed. A "high-pair-frequency" deviant was constructed using bigrams (pairs of characters) that were present in the standard words. For example, the deviant DEF is made up of the first bigram from the standard word **DE**GDEG, the second bigram from the standard word H**EF**, and the open bigram[10] from the standard word **D**I**F**. A "low-pair-frequency deviant" was instead constructed using bigrams that were never present in standard words. For example, the deviant HIG is made of characters present in the aforementioned standard words, but in a unique combination. Therefore, all words used the same characters, but although standard words and high-pair-frequency deviants shared bigrams, low-pair-frequency deviants did not.

Participants first completed a learning block in which only standard stimuli were presented. They were instructed to pay attention and try to learn them. After this, the experiment followed a visual oddball design. Standard stimuli were presented intermixed with deviants, and participants were asked to classify them as either "correct" (standard) or "mistaken" (deviant). Standard words were presented in 90% of the trials (15% each token), and deviants were presented in 5% of the trials each.

This design allowed us to separately manipulate two variables. On one hand, the frequency of occurrence of each individual word (token frequency) was 15% for the standard stimuli and 5% for each deviant stimulus. On the other hand, the mean pair frequency (mean of the frequency of occurrence of the composing bigrams) of each word was high for standard stimuli (5.27%) and high-pair-frequency deviants (6.66%) but low for low-pair-frequency deviants (1.66%). This is because, although high-pair-frequency deviants shared bigrams with standard stimuli, the bigrams composing low-pair-frequency deviants were unique to them.

Our task requested participants to distinguish words by their token frequency; therefore, high- and low-pair-frequency deviants should be equally rejected. However, if participants are sensitive to the frequency with which pairs of characters appear together, the detection of high-pair-frequency deviants should be harder.

We characterized participants' performance by computing their sensitivity index or d-prime (*d'*).[23] All effect sizes reported are Hedges' *g*.[24],[25] Confidence intervals (CIs) reported between square brackets are 95% CI.

While the mean *d'* for the high-pair-frequency deviant was 0.84 [0.24, 1.45], it was 2.02 [1.44, 2.59] for the low-pair-frequency deviant. Therefore, high-pair-frequency deviants were harder to detect. The difference in *d'* between deviants was 1.17 [0.60, 1.74] ($t_{(21)}$ = 4.30; p = 0.00016; g = 0.86 [0.36, 1.37]). This effect was present in the majority of the participants (86.36% [65.09%, 97.09%] or 19 out of 22; one-sided binomial test: p = 0.00043), which implies that the effect is highly reliable (Figure 2A).

These results were replicated in an additional experiment, in which the novel words presented to the participants were six characters long (Figure S1). The effect found in this experiment was of similar magnitude as the one reported in experiment 1 (Table 1). This is particularly notable, because longer words were presumably more difficult to learn as chunks, and therefore, one would have expected participants to rely more heavily on letter statistics. The fact that an effect of equivalent magnitude emerges regardless of word length suggests that bigram coding is not a strategic, task-specific effect; rather, it seems to be an intrinsic aspect of novel word coding.

In brief, the detection of deviants that shared pairs of characters with the standard words resulted more challenging, which implies that participants were sensitive to words' bigram frequencies.

Participants are sensitive to the co-occurrence of shape features in visual objects

Experiment 2 had exactly the same design as the preceding experiment, with the exception that the stimuli consisted now of 2-dimensional renderings of 3-dimensional objects. These objects were composed by a central Y-shaped body and a distinctive shape attached to each of three branches (similar to the stimuli in Baker et al.[29]; Figure 1B). While the overall objects play the role of words in experiment 1, the terminal shapes play a role analogous to characters.

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

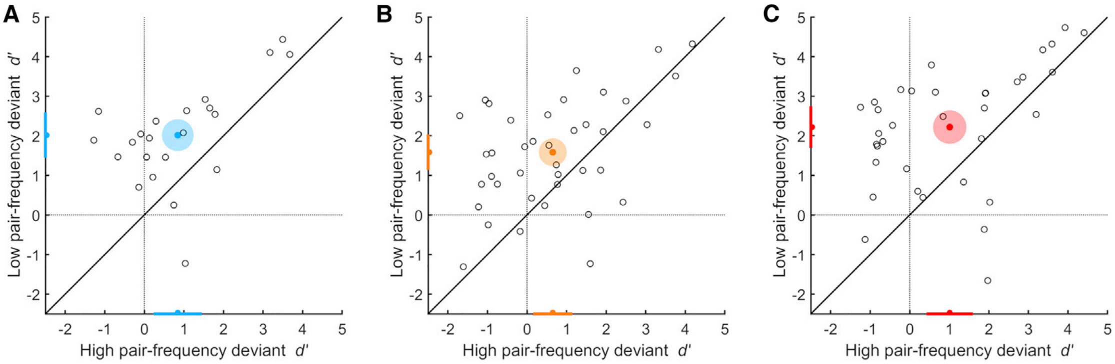

Figure 2. Participants are sensitive to the co-occurrence of visual features across different types of stimuli

On each graph, the x and y axes represent sensitivity (d') to high-pair-frequency and low-pair-frequency deviants, respectively. While each dot represents a participant, the colored dot represents the mean of the group. The shaded area around the mean denotes group level within participants 95% CI. Projected on each axis, a colored dot indicates the mean performance for the respective deviant, and the error bar represents 95% CI. Note that the majority of the participants in all of the experiments lay above the diagonal.

(A) Experiment 1 (three-character word-like stimuli).

(B) Experiment 2 (visual objects).

(C) Experiment 3 (sinusoidal gratings).

See also Figure S1.

These objects differed from written words in two ways. First, rather than being formed by adjacent but independent graphemes, the parts conforming the objects were physically connected. Second, the constituent parts followed a radial spatial arrangement rather than the linear arrangement typical of orthographic material.

We selected 9 distinctive shapes, with which we constructed 6 standard objects and the same two different types of deviants employed in experiment 1. The first of such deviant objects was composed of pairs of shapes that were all present in the standard objects (high-pair-frequency deviant). The second deviant object was instead constructed with pairs of shapes that were not present in any standard object (low-pair-frequency deviant). Two different sets of images were created, and each participant was exposed to one of them. An example stimulus set can be seen in Figure 1B.

As in experiment 1, high-pair-frequency deviants were harder to detect. The average d' for the high-pair-frequency deviant was 0.65 [0.16, 1.15], and it was 1.58 [1.14, 2.03] for the low-pair-frequency deviant. The difference in d' between deviants was 0.93 [0.44, 1.43] ($t_{(38)} = 3.81$; p = 0.00025; $g = 0.64$ [0.27, 1.01]). The effect of interest was present in 74.36% [57.87%, 86.96%] of the participants (29 out of 39; one-sided binomial test: p = 0.0017; Figure 2B).

These results show that an effect akin to sensitivity to bigram frequencies can also be observed when participants are presented with novel visual stimuli that are clearly non-orthographic.

Participants are sensitive to the co-occurrence of low-level visual features

These results of experiment 2 show that participants' sensitivity to bigram frequencies can be observed outside the domain of orthographic processing. Yet the stimuli used were similar to reading material insofar as higher-level units (words or objects) were made up of lower-level parts (characters or shapes) arranged in space. Therefore, to test the generality of our findings, we performed a third experiment, where the stimuli were instead circular sinusoidal gratings defined by combinations of low-level visual features. These features (which played the same role as the characters in experiment 1) were spatial frequency, orientation, and contrast (Figure 1C).

In experiment 1, each of the 3 character positions defining a word could be occupied by 1 out of 3 possible characters. In the same way, each low-level visual feature defining the sinusoidal gratings could take 3 different values. We used these values to construct stimuli with a statistical structure in all identical to that of experiment 1. For example, the high-pair-frequency deviant in Figure 1C shares spatial frequency and contrast with the top left standard, orientation and contrast with the middle left standard, and spatial frequency and orientation with the bottom left standard. So this deviant shares all its "bigrams" with standard stimuli. Instead, the low-pair-frequency deviant is defined by feature values in a unique combination. A different shuffling of feature values was used for each participant.

In experiment 3, mean d' was 1.01 [0.42, 1.60] for the high-pair-frequency deviant and 2.22 [1.69, 2.74] for the low-pair-frequency deviant. Once more, high-pair-frequency deviants were harder to detect. The difference in d' between deviants was 1.21 [0.61, 1.81] ($t_{(34)} = 4.12$; p = 0.00011; $g = 0.74$ [0.33, 1.14]). The majority of the participants (82.86% [66.35%, 93.44%]; 29 out of 35; one-sided binomial test: p = 5.8e−05) showed an effect in the direction of the hypothesis (Figure 2C).

These results show that a "bigram frequency" effect is even present when the features that define the stimuli are non-spatially segregated, low-level visual properties.

Participants' sensitivity to feature co-occurrence across stimuli types

We performed a Bayesian analysis to compare the magnitude of the effects across experiments, which showed that participants' performance was biased by the co-occurrence of features to an equivalent extent, irrespective of the type of stimuli

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

Table 1. Comparison of participants' sensitivity to feature co-occurrence across experiments

Comparison across stimuli types	BF_{01}	Hedges' g
Three-character words versus objects	4.21	0.16 [−0.36, 0.68]
Three-character words versus sinusoidal gratings	4.89	−0.02 [−0.56, 0.51]
Objects versus sinusoidal gratings	4.43	−0.16 [−0.62, 0.29]
Six-character words versus objects	3.37	0.24 [−0.27, 0.76]
Six-character words versus sinusoidal gratings	4.86	0.05 [−0.47, 0.58]
Three-character words versus six-character words	4.31	−0.09 [−0.68, 0.49]

Because we were also interested in evidence for the null (i.e., the difference between high- and low-pair-frequency detection might not differ across stimulus types), we performed a series of Bayesian independent-samples tests (JZS Bayes factor),[26–28] comparing the effect of interest (i.e., d' to low-pair-frequency deviant minus d' to high-pair-frequency deviant) across experiments. All comparisons across experiments yielded BF_{01} values above 3, which implies substantial evidence in favor of the null compared to the alternative hypothesis.

Table 2. Hit rates and false alarm rates in all experiments

Experiment	High P-F deviant hits	Low P-F deviant hits	Standard false alarms
Three-character words	0.47 [0.33, 0.62]	0.78 [0.68, 0.89]	0.23 [0.16, 0.30]
Objects	0.46 [0.35, 0.58]	0.70 [0.60, 0.81]	0.25 [0.20, 0.30]
Sinusoidal gratings	0.45 [0.32, 0.58]	0.76 [0.66, 0.86]	0.17 [0.11, 0.22]
Six-character words	0.40 [0.26, 0.53]	0.74 [0.66, 0.83]	0.26 [0.18, 0.33]

Hit rates and false alarm rates were comparable across experiments.

used (Table 1; a summary of participants' hit and false alarm rates in all experiments can be found in Table 2). Despite different mechanisms could in principle produce effects of equivalent magnitude purely by chance, this evidence may suggest that the data described here are the result of a unique, domain-general visual mechanism.

DISCUSSION

Letter co-occurrence and models of visual word identification

As we mentioned earlier, the role of bigram frequencies in reading is at the center of a heated debate. Recently, Grainger[30] proposed the coexistence of location-invariant and location-specific coding as the marker of orthographic processing—bigram and bigram frequencies were not considered as a defining feature. Schmalz and Mulatti[14] have reviewed and studied the effect of bigram frequencies in reading times. Using Bayesian analysis, the authors conclude that there is evidence against a modulation of reading times by bigram frequencies. Chetail[13] conducted instead a systematic revision covering 20 papers on bigram frequency and arrived at the conclusion that sensitivity to orthographic regularities, such as bigram frequencies, may influence visual word recognition at all levels of processing.

It is worth mentioning that the putative bigram effect seems to be easier to obtain in experiments using pseudowords or unfamiliar scripts that are novel to the participants (e.g., Chetail[31] and Lelonkiewicz et al.[32]), even if it is not clear how much familiarity with a novel alphabet/lexicon is required before this and other signs of mature orthographic processing emerge.[33] In experiment 1, we provide new evidence in this line, showing that, when readers are confronted with novel words written in an unfamiliar script, they encode for the statistics of co-occurrence between letters quite soon and capitalize on it to identify such words.

Several models of visual word identification propose that orthographic processing builds on letter co-occurrence statistics. Seminal work on parallel distributed processing models explicitly suggested that the nature of sublexical representations depend on orthographic regularities.[34] More recently, models were proposed that suggest a hierarchy where higher-order units are based on the statistics of co-occurrence between lower-level units (e.g., Dehaene et al.[11]). Some of these models have been implemented computationally[35,36] and have proven to account for several experimental findings in visual word identification. Some of these models explicitly commit to the existence of bigram representations[11]; this is obviously in line with the results illustrated here, although we did not contrast different coding schemes (e.g., edit distance; spatial coding),[37] and therefore, we cannot speak in this respect. It is worth noting that high-pair-frequency deviants shared bigrams with the standard words/objects, which might have contributed, in addition to bigram frequency, to make these kinds of stimuli more confusable. Note, however, that, regardless of the coding scheme, our results unambiguously show that participants are sensitive to the presence of orthographic regularities.

Visual "word" identification of non-orthographic stimuli

Although the models mentioned in the previous section successfully account for a number of experimental findings in the context of reading, they share a common limitation. By dealing only with orthographic stimuli, they do not place reading within the broader context of visual perception. As a result, they consider reading as domain specific (either explicitly or implicitly). In sharp contrast, the results of our experiments 2 and 3 strongly suggest that at least some of the mechanisms at play during visual word identification are not specialized for reading.

The presence of an effect akin to sensitivity to bigrams' frequencies in the case of stimuli radically different from orthographic material implies that the mechanism at play might be domain general. This is in agreement with the proposal that cultural inventions, such as writing, recycle pre-existing domain-general cognitive processes.[2,3] Furthermore, our results suggest that, in the case of reading, one mechanism that is recycled is the visual system's ability to extract statistical regularities in the co-occurrence of lower-level visual features and integrate them into higher-level multifeatural representations.

Our intention is not to claim that every aspect of visual word identification is domain general. Certainly, some effects found in the literature seem to reveal letters' special status. For example, transposed-letter strings (e.g., NDTF for NTDF) are considerably more confusable than strings of non-letters symbols (e.g., %$&£ for %&$£);[38] the accuracy profile in letter identification across different positions within strings is unique to letters (e.g., Tydgat and Grainger[39]); and attention is preferably deployed to the beginning of letter strings, but not to the beginning of strings of other symbols (e.g., Scaltritti and Balota[40]). In experiment 2, the shapes defining the novel objects were arranged radially, and the notion itself of spatial arrangement cannot be applied at all to the sinusoidal gratings used in experiment 3. Therefore, some of the above-mentioned letter-specific effects, which require a linear spatial arrangement, are fundamentally not replicable with these stimuli. However, the existence of effects that are specific to orthographic material does not justify the conception of visual word identification as purely domain specific. A successful model of reading should account for these letter-specific effects but consider the broader framework of vision.

Word recognition in the bigger context of vision

Our results are in agreement with two major theoretical frameworks in visual neuroscience. The first asserts that visual object information is extracted along a largely feedforward hierarchy, where tuning for progressively more-complex visual features is built up incrementally.[41] That is, units in a given layer of the hierarchy integrate inputs from units of the previous layer, gaining selectivity for the combination of features encoded by the input neurons. For instance, inputs from simple edge detectors can be combined to produce tuning for corners and curved boundaries.

These ideas have been instantiated in a number of neural network models, starting from Fukushima's Neocognitron[42] and Riesenhubber's and Poggio's HMAX model,[43] to arrive to modern deep convolutional neural networks.[44] These models not only achieve extraordinary accuracies at classifying visual images but can also account for key trends in the tuning properties of ventral neurons in humans, monkeys, and rats,[45–50] as well as human and monkey performance in object recognition tasks.[51–56] Our findings are highly consistent with these models, because they show how distinct visual features (no matter whether characters, shapes, or low-level visual properties) are hierarchically integrated into progressively more-complex combinations (e.g., bigrams or pairs of features) before being represented as full "objects."

Interestingly, in our experiments, such hierarchical feature integration took place spontaneously, via exposure to the statistics of the stimulus sets. This is in line with another important theoretical principle that has been called into cause to explain why visual neurons develop certain kinds of tuning properties. This principle postulates that the tuning of sensory neurons is determined by adaptation to the statistics of the signals they need to encode.[7,9,57] Starting from the pioneering intuitions of Attneave and Barlow,[58,59] such efficient coding principle has been instantiated in a number of computational models that are able to learn key properties of the visual system through unsupervised exposure to the spatiotemporal regularities of the visual input.[60–62]

Solid causal evidence has been gathered to show that visual neurons adaptively change their tuning depending on the statistics of the visual stimuli they have been exposed to, both during postnatal development[63–65] and adult life.[66,67] At the perceptual level, human sensitivity to higher-order image statistics closely matches the informational content of such statistics in natural scenes,[68,69] thus suggesting a developmental adaptation of visual perception to the regularities of the visual world.

Moreover, exposure to altered object statistics has been shown to affect human performance in object-recognition tasks,[54,70,71] implying that unsupervised adaptation to visual input statistics continuously sculpt visual perception, even in adult life. Our findings add further behavioral evidence to this conclusion by showing that enhanced sensitivity to specific pairs of visual features emerges as a result of their mere frequency of occurrence within a given stimulus set.

Final conclusions

The results presented in this work suggest that a fundamental processing mechanism behind the processing of visual words also supports the recognition of other visual objects. This implies that such mechanism is indeed of general purpose and confirms the view that reading builds on evolutionarily older cognitive structures.[2,3] This mechanism would enable the statistical learning of regularities in the visual environment (e.g., Frost et al.,[6] Olshausen and Field,[7] Saffran et al.,[8] and Simoncelli and Olshausen[9]). In this view, specialization for letter and letter clusters would emerge in skilled readers via the heavy exposure to written language that is characteristic of modern society (e.g., Perea et al.[72,73]). To conclude, we hope this work will help to inspire models of reading that profit from the body of knowledge amassed in the broader field of visual neuroscience.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Participants
- METHOD DETAILS
  - ○ Procedure and experimental design
  - ○ Stimuli sets
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Data and participants exclusion criteria
  - ○ Measure of performance
  - ○ Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.12.017.

ACKNOWLEDGMENTS

This research was supported by the ERC Starting Grant no. 679010 (STATLEARN), awarded to D.C., and by the ERC Consolidator Grant no. 616803 (LEARN2SEE), awarded to D.Z.

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

AUTHOR CONTRIBUTIONS

Conceptualization, Y.V., E.V., D.Z., and D.C.; Data Curation, Y.V.; Formal Analysis, Y.V.; Funding Acquisition, D.Z. and D.C.; Investigation, Y.V. and E.V.; Methodology, Y.V.; Resources, Y.V. and E.V.; Software, Y.V.; Supervision, Y.V. and D.C.; Visualization, Y.V.; Writing – Original Draft, Y.V. and D.C.; Writing – Review & Editing, Y.V., E.V., D.Z., and D.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 14, 2020
Revised: November 12, 2020
Accepted: December 14, 2020
Published: January 7, 2021

REFERENCES

1. C. Woods, ed. (2010). Visible Language (Oriental Institute of the University of Chicago).
2. Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. Neuron *56*, 384–398.
3. Dehaene, S., and Cohen, L. (2011). The unique role of the visual word form area in reading. Trends Cogn. Sci. *15*, 254–262.
4. Grainger, J., Dufau, S., Montant, M., Ziegler, J.C., and Fagot, J. (2012). Orthographic processing in baboons (Papio papio). Science *336*, 245–248.
5. Scarf, D., Boy, K., Uber Reinert, A., Devine, J., Güntürkün, O., and Colombo, M. (2016). Orthographic processing in pigeons (*Columba livia*). Proc. Natl. Acad. Sci. USA *113*, 11272–11276.
6. Frost, R., Armstrong, B.C., and Christiansen, M.H. (2019). Statistical learning research: a critical review and possible new directions. Psychol. Bull. *145*, 1128–1153.
7. Olshausen, B.A., and Field, D.J. (2004). Sparse coding of sensory inputs. Curr. Opin. Neurobiol. *14*, 481–487.
8. Saffran, J.R., Aslin, R.N., and Newport, E.L. (1996). Statistical learning by 8-month-old infants. Science *274*, 1926–1928.
9. Simoncelli, E.P., and Olshausen, B.A. (2001). Natural image statistics and neural representation. Annu. Rev. Neurosci. *24*, 1193–1216.
10. Grainger, J., and Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? Trends Cogn. Sci. *8*, 58–59.
11. Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. Trends Cogn. Sci. *9*, 335–341.
12. Snell, J., van Leipsig, S., Grainger, J., and Meeter, M. (2018). OB1-reader: a model of word recognition and eye movements in text reading. Psychol. Rev. *125*, 969–984.
13. Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. Front. Psychol. *6*, 645.
14. Schmalz, X., and Mulatti, C. (2017). Busting a myth with the Bayes factor. Ment. Lex. *12*, 263–282.
15. Binder, J.R., Medler, D.A., Westbury, C.F., Liebenthal, E., and Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. Neuroimage *33*, 739–748.
16. Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., and Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron *55*, 143–156.
17. Lochy, A., Jacques, C., Maillard, L., Colnat-Coulbois, S., Rossion, B., and Jonas, J. (2018). Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. Proc. Natl. Acad. Sci. USA *115*, E7595–E7604.
18. Rajalingham, R., Kar, K., Sanghavi, S., Dehaene, S., and DiCarlo, J.J. (2020). The inferior temporal cortex is a potential cortical precursor of orthographic processing in untrained monkeys. Nat. Commun. *11*, 3886.
19. Vidal, C., Content, A., and Chetail, F. (2017). BACS: the Brussels artificial character sets for studies in cognitive psychology and neuroscience. Behav. Res. Methods *49*, 2093–2112.
20. Maurer, U., Blau, V.C., Yoncheva, Y.N., and McCandliss, B.D. (2010). Development of visual expertise for reading: rapid emergence of visual familiarity for an artificial script. Dev. Neuropsychol. *35*, 404–422.
21. Taylor, J.S., Plunkett, K., and Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: an artificial orthography paradigm. J. Exp. Psychol. Learn. Mem. Cogn. *37*, 60–76.
22. Taylor, J.S.H., Davis, M.H., and Rastle, K. (2019). Mapping visual symbols onto spoken language along the ventral visual stream. Proc. Natl. Acad. Sci. USA *116*, 17723–17728.
23. Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. Behav. Res. Methods Instrum. Comput. *31*, 137–149.
24. Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. J. Educ. Stat. *6*, 107–128.
25. Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. *4*, 863.
26. Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. *16*, 225–237.
27. Jarosz, A.F., and Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. J. Probl. Solving *7*, 2–9.
28. Leppink, J., O'Sullivan, P., and Winston, K. (2017). Evidence against vs. in favour of a null hypothesis. Perspect. Med. Educ. *6*, 115–118.
29. Baker, C.I., Behrmann, M., and Olson, C.R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat. Neurosci. *5*, 1210–1216.
30. Grainger, J. (2018). Orthographic processing: a 'mid-level' vision of reading: the 44th Sir Frederic Bartlett lecture. Q. J. Exp. Psychol. *71*, 335–359.
31. Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. Cognition *163*, 103–120.
32. Lelonkiewicz, J.R., Ktori, M., and Crepaldi, D. (2020). Morphemes as letter chunks: discovering affixes through visual regularities. J. Mem. Lang. *115*, 104152.
33. Fernández-López, M., Marcet, A., and Perea, M. (2020). Does orthographic processing emerge rapidly after learning a new script? Br. J. Psychol. Published online August 10, 2020. https://doi.org/10.1111/bjop.12469.
34. Seidenberg, M.S. (1987). Sublexical structures in visual word recognition: access units or orthographic redundancy? In Attention and Performance 12: The Psychology of Reading, M. Coltheart, ed. (Lawrence Erlbaum Associates), pp. 245–263.
35. Whitney, C. (2001). How the brain encodes the order of letters in a printed word: the SERIOL model and selective literature review. Psychon. Bull. Rev. *8*, 221–243.
36. Davis, C.J. (2010). SOLAR versus SERIOL revisited. Eur. J. Cogn. Psychol. *22*, 695–724.
37. Davis, C.J. (2010). The spatial coding model of visual word identification. Psychol. Rev. *117*, 713–758.
38. Massol, S., Duñabeitia, J.A., Carreiras, M., and Grainger, J. (2013). Evidence for letter-specific position coding mechanisms. PLoS ONE *8*, e68460.
39. Tydgat, I., and Grainger, J. (2009). Serial position effects in the identification of letters, digits, and symbols. J. Exp. Psychol. Hum. Percept. Perform. *35*, 480–498.

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

40. Scaltritti, M., and Balota, D.A. (2013). Are all letters really processed equally and in parallel? Further evidence of a robust first letter advantage. Acta Psychol. (Amst.) *144*, 397–410.
41. DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron *73*, 415–434.
42. Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. *36*, 193–202.
43. Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat. Neurosci. *2*, 1019–1025.
44. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444.
45. Cadieu, C., Kouh, M., Pasupathy, A., Connor, C.E., Riesenhuber, M., and Poggio, T. (2007). A model of V4 shape selectivity and invariance. J. Neurophysiol. *98*, 1733–1750.
46. Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. USA *111*, 8619–8624.
47. Khaligh-Razavi, S.M., and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput. Biol. *10*, e1003915.
48. Matteucci, G., Bellacosa Marotti, R., Riggi, M., Rosselli, F.B., and Zoccolan, D. (2019). Nonlinear processing of shape information in rat lateral extrastriate cortex. J. Neurosci. *39*, 1649–1670.
49. Bashivan, P., Kar, K., and DiCarlo, J.J. (2019). Neural population control via deep image synthesis. Science *364*, eaav9436.
50. Kietzmann, T.C., Spoerer, C.J., Sörensen, L.K.A., Cichy, R.M., Hauk, O., and Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. Proc. Natl. Acad. Sci. USA *116*, 21854–21863.
51. Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. USA *104*, 6424–6429.
52. Rajalingham, R., Schmidt, K., and DiCarlo, J.J. (2015). Comparison of object recognition behavior in human and monkey. J. Neurosci. *35*, 12127–12136.
53. Rajalingham, R., Issa, E.B., Bashivan, P., Kar, K., Schmidt, K., and DiCarlo, J.J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. J. Neurosci. *38*, 7255–7269.
54. Cox, P.H., and Riesenhuber, M. (2015). There is a “U” in clutter: evidence for robust sparse codes underlying clutter tolerance in human vision. J. Neurosci. *35*, 14148–14159.
55. Kubilius, J., Bracci, S., and Op de Beeck, H.P. (2016). Deep neural networks as a computational model for human shape sensitivity. PLoS Comput. Biol. *12*, e1004896.
56. Kheradpisheh, S.R., Ghodrati, M., Ganjtabesh, M., and Masquelier, T. (2016). Deep networks can resemble human feed-forward vision in invariant object recognition. Sci. Rep. *6*, 32672.
57. Olshausen, B.A., and Field, D.J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res. *37*, 3311–3325.
58. Attneave, F. (1954). Some informational aspects of visual perception. Psychol. Rev. *61*, 183–193.
59. Barlow, H.B. (1961). Possible principles underlying the transformations of sensory messages. In Sensory Communication, W.A. Rosenblith, ed. (MIT), pp. 217–234.
60. Földiák, P. (1991). Learning invariance from transformation sequences. Neural Comput. *3*, 194–200.
61. Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607–609.
62. Berkes, P., and Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. J. Vis. *5*, 579–602.
63. Hsu, A.S., and Dayan, P. (2007). An unsupervised learning model of neural plasticity: orientation selectivity in goggle-reared kittens. Vision Res. *47*, 2868–2877.
64. Matteucci, G., and Zoccolan, D. (2020). Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of V1 complex cells. Sci. Adv. *6*, eaba3742.
65. Arcaro, M.J., Schade, P.F., Vincent, J.L., Ponce, C.R., and Livingstone, M.S. (2017). Seeing faces is necessary for face-domain formation. Nat. Neurosci. *20*, 1404–1412.
66. Li, N., and DiCarlo, J.J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. Science *321*, 1502–1507.
67. Li, N., and DiCarlo, J.J. (2010). Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. Neuron *67*, 1062–1075.
68. Tkačik, G., Prentice, J.S., Victor, J.D., and Balasubramanian, V. (2010). Local statistics in natural scenes predict the saliency of synthetic textures. Proc. Natl. Acad. Sci. USA *107*, 18149–18154.
69. Hermundstad, A.M., Briguglio, J.J., Conte, M.M., Victor, J.D., Balasubramanian, V., and Tkačik, G. (2014). Variance predicts salience in central sensory processing. eLife *3*, 1–28.
70. Wallis, G., and Bülthoff, H.H. (2001). Effects of temporal association on recognition memory. Proc. Natl. Acad. Sci. USA *98*, 4800–4804.
71. Perry, G., Rolls, E.T., and Stringer, S.M. (2006). Spatial vs temporal continuity in view invariant visual object recognition learning. Vision Res. *46*, 3994–4006.
72. Perea, M., Moret-Tatay, C., and Panadero, V. (2011). Suppression of mirror generalization for reversible letters: evidence from masked priming. J. Mem. Lang. *65*, 237–246.
73. Perea, M., Abu Mallouh, R., Garcí A-Orza, J., and Carreiras, M. (2011). Masked priming effects are modulated by expertise in the script. Q J Exp Psychol (Hove) *64*, 902–919.
74. Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. *10*, 433–436.
75. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. *10*, 437–442.
76. Hentschke, H., and Stüttgen, M.C. (2011). Computation of measures of effect size for neuroscience data sets. Eur. J. Neurosci. *34*, 1887–1894.
77. Blender Online Community (2017). Blender - a 3D modelling and rendering package. https://www.blender.org/.
78. Endress, A.D., and Mehler, J. (2009). The surprising power of statistical learning: when fragment knowledge leads to false memories of unheard words. J. Mem. Lang. *60*, 351–367.
79. Schwarzkopf, S. (2015). Bayes Factors Matlab functions.
80. Morey, R.D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). Tutor. Quant. Methods Psychol. *4*, 61–64.

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw data	This paper	https://osf.io/3tyeu/
Stimuli used in experiments 1, 2 and 3	This paper	https://osf.io/3tyeu/
Software and algorithms		
MATLAB 2015b	MathWorks	https://www.mathworks.com/
Psychophysics Toolbox v.3.0.14	[74,75]	http://psychtoolbox.org/
Measures of Effect Size Toolbox v.1.6.1	[76]	https://it.mathworks.com/matlabcentral/fileexchange/32398-hhentschke-measures-of-effect-size-toolbox
Bayes Factors MATLAB functions	Sam Schwarzkopf's lab	https://figshare.com/articles/Bayes_Factors_Matlab_functions/1357917
Analysis scripts	This paper	https://osf.io/3tyeu/
Blender 2.79b	Blender Foundation	https://www.blender.org/
Other		
Brussels Artificial Character Set (BACS)	[19]	https://osf.io/dj8qm/
Monitor LCD 27''	BenQ	XL2720Z

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Davide Crepaldi (dcrepaldi@sissa.it, @CrepaldiDavide).

Materials availability

The stimuli used in experiments 1, S1, and 2 can be found at the following Open Science Framework (OSF) repository: https://osf.io/3tyeu/

The stimuli used in experiment 3 were generated programmatically.

Data and code availability

The data from all experiments, as well as the code used to perform the statistical analysis can be found at https://osf.io/3tyeu/.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

All participants were self-reported right handed, Italian native speakers, and were recruited from the city of Trieste via on-line advertisement. They all had normal or corrected-to-normal vision and no language-related impairments. Participants provided informed consent and received a monetary compensation of 10€. The experiment was approved by SISSA's Ethical Committee.

Twenty-two participants (5 male and 17 female) took part in experiment 1 (mean age = 23.4, $\sigma = 2.21$ years); 40 participants (12 male, 28 female) took part in experiment 2 (mean age = 24.22, $\sigma = 2.57$ years) and 36 participants (8 male, 28 female) took part in experiment 3 (mean age = 23.22, $\sigma = 3.37$ years).

METHOD DETAILS

Procedure and experimental design

The 4 experiments presented in this work followed the same procedure and used the same experimental design, differing only in the stimuli used. Participants sat in a sound-attenuated testing booth at around 70cm of a 27'' computer monitor (BenQ XL2720Z). The

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

experiments were programmed and run in MATLAB (2015b, MathWorks, Natick, MA, USA) using the Psychophysics Toolbox (v3) extensions.[74,75]

As described in the Results section, participants first completed a learning block where 200 standard stimuli were presented. During this block they were instructed to pay attention and try to learn the stimuli. Stimuli were presented one at a time and remained on screen for 1.5 to 2 s. After the learning block, participants completed 6 testing blocks, where standard stimuli were presented intermixed with deviant stimuli. On each of these blocks, participants were presented with 180 standard stimuli (90% of the trials; 15% each of the 6 standard tokens) and 20 deviant stimuli (10% percent of the trials; 5% each deviant condition).

Stimuli order of presentation was pseudorandom. Each test block started with 12 standard stimuli, after which deviants were presented allowing 6 to 12 standard stimuli in between each deviant presentation. Stimuli repetitions were only allowed after 2 other stimuli were presented (e.g., s2 s3 s1 s2).

Participants had a maximum of 2 s to classify each stimulus as either "Correct" (standard) or "Mistaken" (deviant) by pressing one of two keys on a keyboard. Key mapping was counterbalanced across participants. In case of timeout, the next trial started without any feedback. Participants were not informed about the amount and type of deviants. Overall, each participant was asked to classify 1080 instances of standard stimuli, and 60 instances of each deviant stimuli. Each block lasted on average 7 minutes and the entire experiment had an approximated duration of 50 minutes.

Stimuli sets

The stimuli used in experiment 1 (and in the supplemental experiment reported in Figure S1) were words constructed using the Brussels Artificial Character Set (BACS-2 serif[19]), whose characters have perimetric complexity, number of strokes, junctions and terminations matched to the English alphabet. We picked 23 out of the 26 available characters in BACS-2 with serifs. The three characters excluded were J, which in BACS-2 resembles a question mark; P, which is a vertical mirror flip image of the BACS-2 character B; and W, which resembles closely the BACS-2 character V.

The stimuli used in experiment 2 were not orthographic. Instead, we used images of 3D objects created using the software Blender (version 2.79b[77]). Two different sets of images with the same statistical structure were created, and each participant was exposed to one of them. This had the goal of ruling out the possibility that the effects were driven by some idiosyncratic feature of a given set of shapes.

Finally in experiment 3, the stimuli used were circular sinusoidal gratings defined by different values of three low level visual features. These features were spatial frequency (.4, 0.8 and 1.6 cycles per degree of visual angle), contrast (20%, 60% and 100%) and orientation (0, 45 and 90 degrees).

Note that in all experiments, although the high pair-frequency deviants shared 2 features with 3 of the standard stimuli, they shared none with the other three. On the other hand, the low pair-frequency deviants shared 1 feature with each of the 6 standard stimuli. Thus the average number of shared features between deviants and standard stimuli was 1 for both types of deviants. Furthermore for each participant, stimuli were constructed using a different shuffling characters (experiment 1) and feature values (experiment 3), but always respecting the same statistical structure. Similarly, in experiment 2, we constructed 2 different stimuli sets using different combinations of shapes. These measures make it impossible to correctly classify stimuli based on single characters, shapes or feature values, and extremely unlikely to be able to do it based on values of individual pixels.

The statistical structure of the stimuli sets was based on the one used in Endress and Mehler.[78]

QUANTIFICATION AND STATISTICAL ANALYSIS

Data and participants exclusion criteria

As we stated earlier, during the testing blocks participants had a time limit of 2 s to provide an answer. Trials in which participants did not provide an answer were excluded from the analyses, and participants with more than 20% of such trials for any stimulus category were excluded altogether.

In experiment 1, all participants provided enough trials in all conditions. Participants failed to provide an answer in 1.68% of the standard trials, 2.88% of the high pair-frequency deviant trials and 2.58% of the low pair-frequency deviant trials. Overall participants provided an answer in 98.21% of the trials. Two participants had to be excluded from the analysis in experiment S1. The remaining participants failed to provide an answer in 3.25% of the standard trials, 3.55% of the high pair-frequency deviant trials and 3.7% of the low pair-frequency deviant trials. Participants provided an answer in 96.71% of all trials. In experiment 2, one participant was excluded from the analysis. The rest of the participants failed to provide an answer in 2.05% of the standard trials, 2.74% of the high pair-frequency deviant trials and 2.65% of the low pair-frequency deviant trials. Of all trials, 97.88% were answered within the time limit. Finally, one participant was excluded from the analysis in experiment 3. The remaining participants failed to provide an answer in 1.42% of the standard trials, 1.76% of high pair-frequency deviant trials and 2% of the low pair-frequency deviant trials. Participants provided an answer in 98.53% of all trials.

Measure of performance

To better characterize the participants' ability to detect deviant stimuli, we resorted to Signal Detection Theory and computed an independent sensitivity index (d-prime score or *d'*) for each deviant type and for each participant. This measure of performance takes

Please cite this article in press as: Vidal et al., A general-purpose mechanism of visual feature association in visual word identification and beyond, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.017

into account possible response biases.[23] Participants' responses were classified as "hit" (deviant stimuli classified as "mistaken") or "false alarm" (standard stimuli classified as "mistaken"). Next, for each deviant type, *d'* was calculated as

$$d' = Z(\text{HitRate})\text{-}Z(\text{FalseAlarmRate})$$

where Z(…) is the inverse of the cumulative standard normal distribution. This takes into consideration the overall bias toward a "correct" or a "mistaken" response. As this function does not output a finite value if either the hit rate or the false alarm rate are either 0 or 1, and considering the total amount of trials of each type, hit rate was capped between 1/60 and 59/60, and false alarm rate was capped between 1/1080 and 1079/1080.

Statistical analysis

Statistical comparisons within each experiment were performed using paired-samples Student's t test when comparing *d'* scores across deviants. Comparisons across experiments were performed using Bayesian independent samples tests (JZS Bayes Factor[26–28]). This test measures the relative evidence between the null and alternative hypotheses, allowing to assess evidence in favor of the null. Tests were performed using a Cauchy prior with scale value of $r = 1$. The code for this analysis was written by Sam Schwarzkopf.[79]

All effect sizes reported are Hedges' *g*,[24,25] which is more precise than Cohen's *d*, as it applies a correction for small sample sizes. Effect sizes were calculated using the Measures of Effect Size Toolbox.[76] All confidence intervals reported between square brackets are 95% CIs. In the case of Figure 2, as data is paired, the CI around the dot representing the mean performance at the group level is a within participants CI calculated using the normalization method proposed by Morey.[80]