# Paper features: A neglected source of information for letter recognition

**Xavier Morin Duchesne**     Department of Psychology, University of Montreal, Montreal, Canada

**Frédéric Gosselin**     Department of Psychology, University of Montreal, Montreal, Canada

**Daniel Fiset**     Department of Psychoeducation and Psychology, University of Quebec in Outaouais, Gatineau, Canada

**Nicolas Dupuis-Roy**     Department of Psychology, University of Montreal, Montreal, Canada

**Adult observers have surprisingly low calculation efficiencies for letter recognition (see, e.g., Pelli, Burns, Farell, & Moore-Page, 2006). Here, we examine the possibility that this is partly due to observers' neglecting *paper features* (e.g., the absence of ascenders and descenders in 'o'). Each of 16 observers completed 5,000 trials of a single-letter two-alternative forced-choice detection task. Using a combination of classification image analyses and Bayesian statistical analyses, we argue that between 60% and 75% of our participants indeed neglected paper features.**

## Introduction

After years of reading and having to recognize millions of words and letters, one would expect adults to have become letter identification experts. Yet studies have shown that letter discrimination calculation efficiency is surprisingly low—ranging from 8% to 16% (Gold, Bennett, & Sekuler, 1999; Parish & Sperling, 1991; Pelli, Burns, Farell, & Moore-Page, 2006; Tjan, Braje, Legge, & Kersten, 1995).[1] This low calculation efficiency, Pelli et al. (2006) have argued, is the result of human observers' use of a suboptimal letter identification strategy. Specifically, they suggest that rather than the optimal template matching strategy, human observers recognize letters by detecting only a few features (i.e., diagnostic regions of letter images).

These features by which letters are recognized have been revealed by Fiset et al. (2008) using the Bubbles technique. Notably, they showed that these letter features coincide mostly with line terminations and, to a lesser extent, with horizontal lines (see also Chung, Tjan, & Lin, 2008; Lanthier, Risko, Stolz, & Besner, 2009; Szwed, Cohen, Qiao, & Dehaene, 2009). However, Fiset et al. (2008) overlooked an important aspect of these letter features: They fall almost exclusively on ink regions of letter images (see Fiset et al., 2008, figure 2; Fiset et al., 2009). This, as well as the fact that it was overlooked, suggests that observers typically do not use letter features falling on paper regions of letter images in letter recognition. In other words, it appears that observers rely on ink features—regions of ink in letter images—but not on paper features—regions of paper (without ink) in letter images. For example, observers rely on the presence of an ink curve to recognize the letter 'o'—an ink feature—presumably to distinguish it from the letter 'c,' but do not use the lack of ascenders or descenders—paper features—to distinguish it from the letters 'b,' 'd,' 'p,' 'q,' and 'g.'

Neglecting paper features could represent a bottleneck for letter recognition. However, the fact that no paper features attained statistical significance does not provide support for this hypothesis: it remains unclear whether observers did in fact fail to use paper features, because not rejecting the null hypothesis is different from supporting it. To circumvent this statistical shortcoming, Bayesian statistics could be employed to pit the null hypothesis against the alternative hypoth-

esis (humans do not use paper features vs. humans do use paper features; see Gallistel, 2009). Having said that, the data from Fiset et al. (2008) lack the necessary statistical power to obtain reliable individual results, and group averages risk underestimating individual differences. One could imagine, for example, that all of the observers used paper features but that each used a different set of paper features. In this case, the group average would lead to the false conclusion that observers did not use paper features. Conversely, it could be that only a few of the observers used paper features, but systematically so. Group averages would then lead to the erroneous conclusion that all observers used paper features.

In this study, 16 observers completed 5,000 trials of a single-letter two-alternative forced-choice detection task. We chose a detection task—with simple targets—to maximize statistical power and obtain reliable individual results on which to perform Bayesian statistical analyses. Although the features that we revealed are probably different from those that we would have revealed in a letter identification task, they allowed us to test the hypothesis that observers typically neglect paper features.

## Methods

### Subjects

Sixteen psychology students between 19 and 27 years of age participated in the experiment. All observers had normal or corrected-to-normal vision.

### Apparatus

The experimental programs were run on a Macintosh G4 computer in the Matlab environment. All stimuli were presented on a Sony Trinitron monitor (1024 × 768 pixels at 85 Hz), calibrated using a Samsung SyncMaster 753df photometer to allow linear manipulation of luminance. The resulting corrected table contained 101 luminance levels, ranging from 2.25 cd/m$^2$ to 98.84 cd/m$^2$. Observers were tested binocularly and their head position was maintained at a distance of 50 cm from the computer monitor by a chin rest.

### Procedure

Observers each did a total of five blocks of 1,000 trials. On each trial, they were presented simultaneously with a noisy target and a noisy nontarget. The remainder of the screen was midgray (50.55 cd/m$^2$). The target and nontarget each spanned 34 × 34 pixels (about 1° of visual angle); they were centered along the y-axis and had an eccentricity of ±34 pixels (about ±1° of visual angle) along the x-axis. The stimulus remained on the computer monitor until a response was given. Observers had to indicate whether the target was on the left or right by pressing on the appropriate keyboard response key. No feedback was provided to observers.

Unbeknownst to observers, none of the experimental stimuli contained the target images: They only contained white Gaussian noise ($M = 50.55$ cd/m$^2$; $SD = 20.38$ cd/m$^2$). On any given trial, however, one of the two noise fields was always more correlated with the target than the other. Such a no-signal procedure has been used to probe the internal memory representation of participants (see Dupuis-Roy & Gosselin, 2007; Gosselin & Schyns, 2002, 2003; Smith, Gosselin, & Schyns, 2012).

### Stimuli

Thirteen subjects were told verbally at the beginning of every 1,000-trial block that the target was a black letter 'X' printed on a white square, that the letter was formed by two thick diagonal bars crossing at the center of the square and each touching corners of the square, and that the black and white regions of the target had exactly the same number of pixels. The only image matching this target description is represented in Figure 1a. They were also told that the nontarget was a midgray square of the same luminance as the region of the screen outside the stimulus. The other three observers were shown the target reproduced in Figure 1a—a black letter 'O' printed on a white square—at the beginning of every 1,000-trial block. They were also shown an "easy" trial that clearly contained the target 'O' and the nontarget homogenous midgray square of the same luminance as the region of the screen outside the stimulus.

## Results

### Individual classification images

The optimal strategy to resolve this task consists in cross correlating the difference between the target and the nontarget—henceforth the optimal template—with the difference between the left and right noisy squares of the stimuli, and responding that the target is on the left when the cross correlation exceeds a criterion and that the target is on the right otherwise (Tjan et al., 1995). In the present case, the optimal templates and
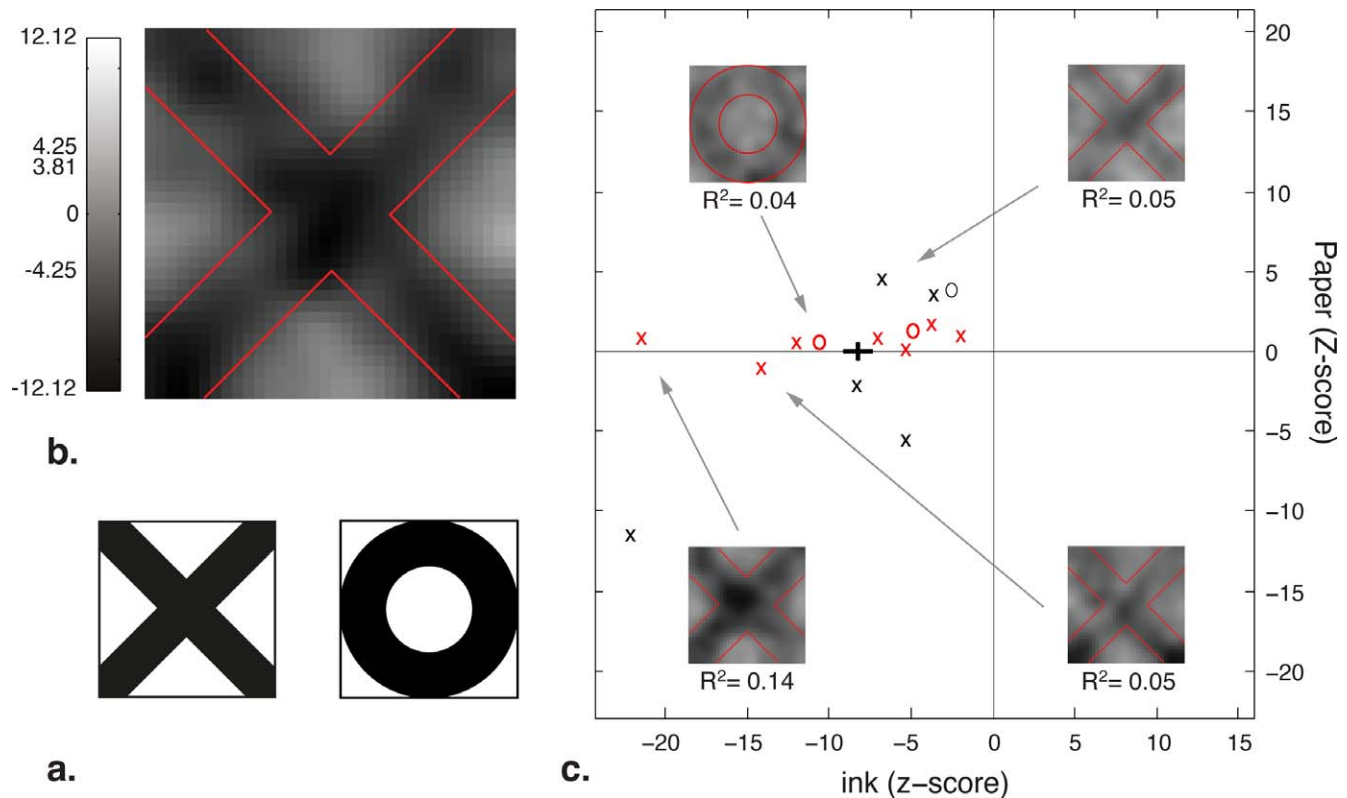
Figure 1. (a) 'X' and 'O' targets. (b) 'X' group classification image (CI) smoothed with a Gaussian kernel with a standard deviation of 2 pixels and transformed into z-scores. Key z-scores are shown on the color bar. The statistical thresholds were obtained using a two-tailed pixel test (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). The red lines superimposed on the group CI trace the boundaries between the ink and paper regions of the 'X' target. (c) Scatter plot of the individual composite z-scores calculated using the best-fitted targets. The x-axis and y-axis give the ink and the paper composite z-scores, respectively. The overall composite z-scores of individual observers are depicted by their target letters, and the average of all observers by the intersection of the two standard error bars. Red letters correspond to observers who used ink features but not paper features. Four individual CIs are also shown, with red lines that trace the boundaries between the ink and paper regions of their best-fitted targets and with associated goodness of fits.

the targets are linearly related, and all the pixels of the optimal template—those corresponding to the ink and paper regions of the targets—contribute equally to the cross correlations.

To estimate the template of each observer for each 1,000-trial block, we performed a least-squares multiple linear regression on the difference between every pair of white Gaussian noise fields and corresponding responses (for details, see Dupuis-Roy & Gosselin, 2007). The output of this regression was a plane of 34 × 34 regression coefficients—henceforth referred to as a classification image (CI)—indicating the strength of the association between the contrast value of a given pixel in the noise field and the detection of the target letter. For a linear observer, the expected value of the CI is proportional to the template (e.g., Murray, Bennett, & Sekuler, 2005). Light gray regions of the CIs represent *positive* regression coefficients: pixels for which a high contrast was correlated with the tendency to detect the target. Dark gray regions represent *negative* regression

coefficients: pixels for which a low contrast was correlated with the tendency to detect the target. Midgray regions indicate regression coefficients near zero; these had little to no impact on the detection of the target letter. In other words, the nonzero elements of a block CI depict the elements of the noise fields which were used to determine the presence of the target letter. Four individual CIs are represented in Figure 1c.

## A lesson from the 'X' group classification image

Next, we computed the group CI for the 12 observers that detected the target 'X' by summing all their block CIs, irrespective of observer, by smoothing the (mirror-padded) result slightly (full width at half maximum = 4.71 pixels), and by transforming it into z-scores. Figure 1b shows this group CI. The pixels range from black (z-score = −12.12) to white (z-score = 12.12). The contour of the target 'X' is superimposed in red on the

| # | Target | Ink | | | Paper | | |
|---|---|---|---|---|---|---|---|
| | | Statistical model standard deviation | Range of alternative prior | Odds | Statistical model standard deviation | Range of alternative prior | Odds |
| 1 | 'X' | 1.50 | −7.68 to −3.81 | $1.55 \times 10^{13}$ | 0.97 | −0.74 to 1.63 | 0.11 |
| 2 | | 0.66 | −1.68 to 0.10 | 4.14 | 0.66 | −0.21 to 1.33 | 0.22 |
| 3 | | 1.08 | −4.46 to −1.76 | $3.44 \times 10^{7}$ | 1.22 | 0.23 to 3.16 | 108.09 |
| 4 | | 0.81 | −2.48 to −0.49 | $4.50 \times 10^{3}$ | 0.83 | −0.51 to 1.750 | 0.73 |
| 5 | | 1.60 | −2.80 to 1.54 | 0.20 | 1.23 | −1.12 to 2.20 | 0.18 |
| 6 | | 1.40 | −6.14 to −2.68 | $5.37 \times 10^{6}$ | 0.54 | −1.72 to 0.34 | 164.05 |
| 7 | | 1.09 | −8.17 to −5.19 | $7.36 \times 10^{35}$ | 1.20 | −1.48 to 1.45 | 0.18 |
| 8 | | 0.79 | −3.36 to −1.39 | $5.73 \times 10^{8}$ | 3.87 | −8.72 to 0.44 | 0.99 |
| 9 | | 1.38 | −3.42 to 0.44 | 4.11 | 1.35 | −0.32 to 2.63 | 4.02 |
| 10 | | 1.31 | −3.91 to −0.72 | 422.82 | 1.01 | −1.10 to 1.36 | 0.10 |
| 11 | | 2.85 | −11.83 to −4.92 | $6.26 \times 10^{11}$ | 1.76 | −6.81 to −2.18 | $1.49 \times 10^{8}$ |
| 12 | | 1.13 | −11.11 to −8.44 | $6.52 \times 10^{76}$ | 1.61 | −1.53 to 2.68 | 0.17 |
| 13 | | 1.18 | −4.38 to −1.80 | $9.03 \times 10^{6}$ | 0.58 | −0.25 to 1.25 | 0.15 |
| 14 | 'O' | 1.38 | −2.39 to 0.80 | 0.64 | 0.63 | 1.01 to 2.69 | $9.47 \times 10^{6}$ |
| 15 | | 0.63 | −5.41 to −3.87 | $4.41 \times 10^{59}$ | 1.54 | −2.09 to 2.08 | 0.15 |
| 16 | | 1.67 | −4.12 to −0.29 | 11.49 | 2.79 | −3.73 to 3.61 | 0.28 |

Table 1. Summary of the individual Bayesian analyses.

group CI. A dark 'X' printed on a midgray square is distinctly visible. This suggests, as predicted by our hypothesis, that observers used ink but not paper features to detect the 'X' target. More specifically, like the subjects for Fiset et al. (2008), our observers appear to have used the line intersection and the four line terminations. However, regions of significant negative correlations in the group CI spread outside of the target 'X' ink region into its paper regions. This suggests that some participants exaggerated the size of the ink region. In other words, some participants appear to have had trouble with the constraint dictating that the black and white regions of the target had exactly the same number of pixels. This could interfere with the testing of our hypothesis—that paper features are neglected—because the summing of dark pixels (negative z-scores) from some observers with bright pixels (positive z-scores) from other participants could artificially result in an average z-score near zero. We would thus be more likely to erroneously conclude that paper features were not used to perform the task.

## Individual Bayesian analyses

Therefore, we proceeded to test our hypothesis controlling for this particular source of variance. Three steps were required. First, we needed to account for individual variations in ink-to-image-area ratio. To do this, we computed individual CIs by summing the five block CIs of each individual. Figure 1 shows four of these individual CIs: They were smoothed and trans-

formed into z-scores. Then we best-fitted each subject's individual CI with an image identical to her or his target letter, but with a varying stroke width, resulting in a *best-fitted target* ($R^2$: $M = 0.03$; $SD = 0.03$; min = 0.003; max = 0.14).

Second, for each of the 16 subjects and each of the five blocks, we computed two composite z-scores (i.e., the sum of all z-scores within a region divided by the square root of the number of z-scores within this region): one composite z-score for the best-fitted ink region and another for the complementary paper region. The odds for one subject (#5 in Table 1) favored the null hypothesis over the alternative hypothesis for both ink and paper regions, which indicates that he or she did not perform the task properly. The two overall composite z-scores of the other subjects (i.e., the sum of the composite z-scores of the five blocks divided by $\sqrt{5}$) are represented on the scatter plot of Figure 1c. The x-axis and y-axis correspond, respectively, to the overall composite z-score of the best-fitted ink area and to the complementary paper area overall composite z-score. Individual observers are depicted by a letter 'X' or 'O' depending on their target letter, and the average of all observers by the intersection of the two standard error bars.

Finally, we tested our hypothesis. We predicted that observers would use ink features but not paper features. Standard statistical tools are inappropriate because, at best, they would allow us to conclude that (a) the composite z-scores are significantly greater in the best-fitted ink area than in the complementary paper area or (b) the composite z-scores are signifi-

cantly greater than zero in the best-fitted ink area but not in the complementary paper area. The first would not eliminate the possibility that paper features are also used, and the second would not support the null hypothesis that paper features are not used. Thus, we turned to Bayesian statistics, which have the power to test our hypothesis (see, e.g., Gallistel, 2009; Kruschke, 2010).

For each observer, we performed two Bayesian statistical analyses: first, for the composite z-score of the best-fitted ink region, and second, for that of the complementary paper region. These analyses closely followed Gallistel (2009). To summarize, we computed the posterior likelihood probability that ink features were used given the ink region composite z-scores, and the posterior likelihood probability that ink features were not used given the ink region composite z-scores. The ratio of the sum of these posterior likelihood probabilities indicates how many times it is more likely that ink features were used rather than not, given the data. Odds greater than 1 favor the alternative hypothesis, and odds smaller than 1 favor the null hypothesis. We performed the same analysis for the composite z-scores of the paper region. Results are summarized in Table 1. The fifth column shows the odds for the ink region, and the eighth column for the paper region. The other columns detail the parameters used for the analyses.[2]

## Conclusion

Nine out of the 15 observers that performed the task properly used ink features but not paper features to detect their targets. The data of one of the remaining observers (#8 in Table 1) who used ink features to detect the 'X' target does not allow us to say if he or she also used paper features (i.e., odds ~ 1). The other five observers did use paper features, contrary to our hypothesis. In fact, one of these observers (#14) used paper features but not ink features to detect the 'O' target. The four remaining observers (#3, #6, #9, and #11) used both ink and paper features. Notably, however, for two of these four observers (#6 and #11), paper features were represented in dark shades, albeit in lighter shades than their ink features. We propose that these dark paper features may have originated from within-subject variations of the internal memory representation of the target. An observer who neglected paper features could, for example, have changed her or his memory representation of the target after two thirds of trials to one with larger strokes. In this case, the best-fitted target would match the memory representation used for the greatest number of trials, and some of the ink features of the target memory representation with

larger strokes would incorrectly be identified as *dark* paper features. Therefore, we take the evidence from these two dark-paper-feature observers to be ambiguous relative to our hypothesis. In sum, 9 out of the 12 observers who obtained unambiguous results in our experiment neglected paper features in detecting their targets.

One caveat is that some of the observers who appear to have neglected paper features might have misunderstood the task: They might have looked for a white square nontarget rather than the midgray square designated by our instructions. Under this interpretation, the paper region of the optimal template is no longer diagnostic for the task, and subjects would not be expected to have used it. In fact, in work by Tjan and Nandy (2006, experiment 2), three subjects performed a letter 'o' detection task in which the nontarget was of the same color as the paper region of the target, and indeed, none of those subjects used paper features. Tjan and Nandy also asked three subjects to participate in letter 'o' versus 'x' discrimination tasks (experiments 1 and 3). In this case, paper regions of the optimal template are diagnostic of the task, and all subjects did use paper features in addition to ink features. These two sets of results from Tjan and Nandy give some credence to this "misunderstanding" possibility.

That being said, the Tjan and Nandy letter discrimination results are not necessarily incompatible with ours, because some of our observers—25% to 40%—did use paper features as well. Furthermore, the five subjects tested by Fiset et al. (2008) in two 26-letter identification tasks and the four subjects tested by Fiset et al. (2009) in a 26-letter identification task appear to have neglected paper features. Altogether, we feel confident concluding that a majority of observers neglect paper features in letter recognition tasks.

As we point out in the Introduction, the absence of ascenders and descenders—paper features—in the letter 'o' would help distinguish it from 'b,' 'd,' 'p,' 'q,' and 'g.' But the letter 'o' is far from being an exception. We have computed the pixel-by-pixel differences between every letter image and every other letter image (i.e., 'a' vs. 'b,' …, 'z'; 'b' vs. 'a,' 'c,' …, 'z,' etc.) and, using the resulting templates, calculated the average proportion of nonzero pixels falling in the ink region. This computation indicates that paper features account for more than 50% of all available information for letter identification. Adding spatial uncertainty and using different font families had little impact on the results.[3] In all fairness, this estimation somewhat exaggerates the importance of paper features relative to ink features for letter recognition. In addition to the letter identification information that they carry, ink features allow the localization of letters: Where there's ink, there's a letter. Paper features do not provide this kind

of information. Once the letter has been coarsely localized, however, paper features become as important as, if not more important than, ink features for letter identification. Therefore, neglecting paper features represents a major efficiency bottleneck for letter identification.

## Footnotes

[1] Parish and Sperling (1991) actually report a 42% letter recognition efficiency. Pelli et al. (2006) indicate, however, that this is a mistake and that the actual calculation efficiency was around 16%.

[2] More specifically, the prior probability distribution of the hypothesis that the composite z-scores of the ink region (vs. paper region) for each individual are equal to zero (the null hypothesis) is a Dirac delta function centered on 0; and the prior probability distribution of the hypothesis that the composite z-scores of the ink region (vs. paper region) for each individual are different from zero (the alternative hypothesis) is a rectangular probability distribution function. The range of this probability distribution was set to the range of the individual empirical data (see columns 4 and 7 in Table 1). We assumed a normal statistical model and estimated its standard deviation from the empirical data (see columns 3 and 6 in Table 1). We then computed the likelihood function given our experimental data. We multiplied point by point the likelihood function of the ink region (vs. paper region) with the two prior probability distributions to get the posterior likelihood functions of the ink region (vs. paper region). We integrated the two posterior likelihood functions of the ink region (vs. paper region) to get the marginal likelihoods. The ratio of the marginal likelihood of the alternative hypothesis and the marginal likelihood of the null hypothesis tells us to what extent the hypothesis that the composite z-scores of the ink region (vs. paper region)

are different from zero is more likely than the hypothesis that the composite z-scores of the ink region (vs. paper region) are equal to zero (see columns 5 and 8 in Table 1).

[3] The proportions of diagnostic pixels occupying paper regions of the letter images with and without Gaussian position uncertainty ($SD = 0.25 \times$ letter width) for different font families: Künstler = 0.573 and 0.581, Bookman = 0.573 and 0.581, Courier = 0.576 and 0.584, Sloan = 0.573 and 0.581, Baskerville = 0.615 and 0.627, Arial = 0.572 and 0.587, Gill Sans = 0.574 and 0.586, Futura = 0.556 and 0.561, Times = 0.593 and 0.602, Helvetica = 0.573 and 0.581. On average, with position uncertainty = 0.578 and without = 0.587.

## References

Chauvin, A., Worsley, K. J., Schyns, P.G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision, 5*(9):1, 659–667, http://www.journalofvision.org/content/5/9/1, doi:10.1167/5.9.1. [PubMed] [Article]

Chung, S., Tjan, B., & Lin, Y. (2008). Feature maps for letters. *Journal of Vision, 8*(6): 970, http://www.journalofvision.org/content/8/6/970, doi:10.1167/8.6.970. [Abstract]

Dupuis-Roy, N., & Gosselin, F. (2007). Perceptual learning without signal. *Vision Research, 47,* 349–356.

Fiset, D., Blais, C., Arguin, M., Tadros, K., Éthier-Majcher, C., Bub, D., & Gosselin, F. (2009). The spatio-temporal dynamics of visual letter recognition. *Cognitive Neuropsychology, 26*(1), 23–35.

Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for uppercase and lowercase letter identification. *Psychological Science, 19*(11), 1161–1168.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116*(2), 439–453.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature, 402*(6758), 176–178.

Gosselin, F., & Schyns, P. G. (2002). You are about to see pictorial representations! *Behavioral & Brain Sciences, 25*(2), 191–192.

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science, 14*(5), 505–509.

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(5), 658–676.

Lanthier, S. N., Risko, E. F., Stolz, J. A., & Besner, D. (2009). Not all visual features are created equal: Early processing in letter and word recognition.

*Psychonomic Bulletin & Review, 16*(1), 67–73, doi: 10.3758/pbr.16.1.67.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision, 5*(2):5, 139–149, http://www.journalofvision.org/content/5/2/5, doi:10.1167/5.2.5. [PubMed] [Article]

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31*(7–8), 1399–1415.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring internal representations for behavioral and brain data. *Current Biology, 22*(3), 191–196.

Szwed, M., Cohen, L., Qiao, E., & Dehaene, S. (2009). The role of invariant line junctions in object and visual word recognition. *Vision Research, 49*(7), 718–725, doi:10.1016/j.visres.2009.01.003.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research, 35*(21), 3053–3069.

Tjan, B. S., & Nandy, A. S. (2006). Classification images with uncertainty. *Journal of Vision, 6*(4):8, 387–413, http://www.journalofvision.org/content/6/4/8, doi:10.1167/6.4.8. [PubMed] [Article]