Superstitious perceptions reveal properties of internal representations

Frédéric Gosselin, Université de Montréal,

and Philippe G. Schyns, University of Glasgow

Abstract

We have all seen a human face in a cloud, a pebble or in blots on a wall. Evidence of superstitious perceptions has been documented since classical antiquity (Gombrich, 1960; Janson, 1973) but has received little scientific attention. Here, we used superstitious perceptions in a new principled method to reveal the properties of unobservable object representations in memory. We stimulated the visual system with unstructured white noise. Observers firmly believed that they perceived an ‘S’ letter in Experiment 1 and a smile on a face in Experiment 2. Using reverse correlation (Ahumada & Lovell, 1971) and computational analyses we visualized the memory representations subtending these superstitious perceptions.

Superstitious perceptions reveal properties of internal representations

For several decades, face, object and scene recognition researchers have sought to understand the properties of representations in memory. However, the relationship between representations and behavior is tenuous, leaving researchers with few options except to test the validity of hypothesized representational schemes.

A few decades ago, Wiener (1958) showed that noise could be used to analyze the behavior of a black box, even suggesting that the brain could be studied this way. Here we propose a principled method, combining Wiener's idea with visual perception, to reconstruct the internal representation of an observer. We start from an unstructured external stimulus (white noise) and we lead the observer to believe that the stimulus comprises a signal. As white noise does not represent coherent structures in the image plane, the superstitious perception of a signal must arise from the observer's share. To characterize these internal representations, we reverse correlate (e.g., Ahumada & Lovell, 1971; Beard & Ahumada, 1998; Gold, Murray, Bennett & Sekuler, 2000; Neri, Parker & Blakemore, 1999; Oshawa, DeAngelis & Freeman, 1990) the observer's detection and rejection responses with the corresponding white noise stimuli.

Experiment 1: 'S' as in Superstitious

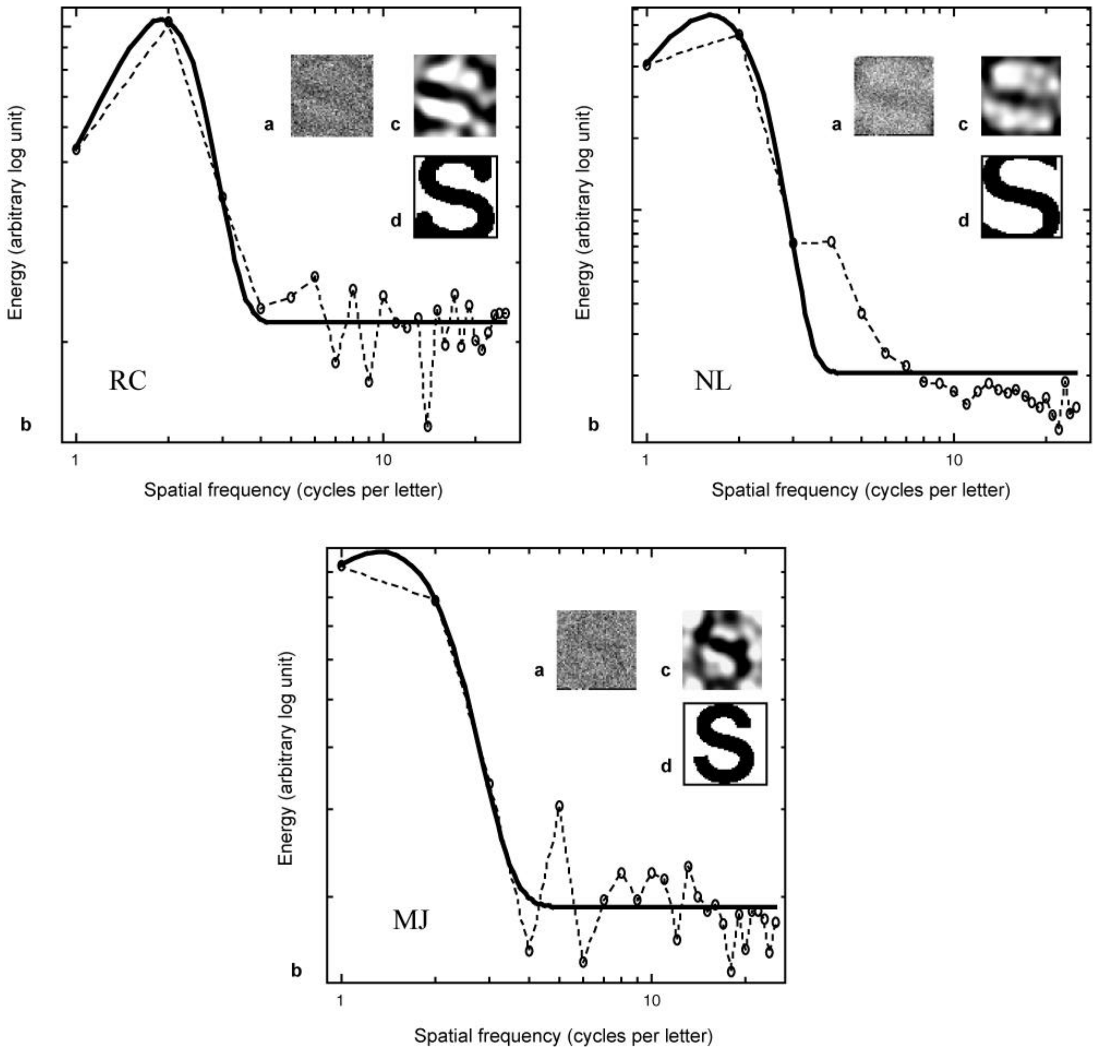
Method

In Experiment 1, we instructed three paid naive observers (RC, NL and MJ, aged 21-24) to detect in white noise the presence of a target black letter 'S' on a white background filling the image. The observers were told that the letter 'S' (for Superstitious) was present on 50% of 20,000 trials equally divided into 40 blocks and completed over a fortnight. No more detail was given regarding the shape of the letter. Each trial consisted in the presentation of one 50 x 50 pixels (2 x 2 deg of visual angle) static bit noise image with a black-pixel density of 50%. No signal was ever presented. The experiment ran on a G4 Macintosh using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997).

Results

The observers detected an ‘S’ in noise on 22.7% (RC), 45.9% (NL), and 11% (MJ) of the trials, respectively. They claimed that they responded positively whenever they saw an ‘S’ and estimated the quantity of added noise to vary between 30% and 50%. Observer RC described her response strategy as: “I simply waited to see if the S ‘jumped out at me’”.

To depict the information eliciting these superstitious perceptions, we applied reverse correlation. For each observer, we computed a ‘yes image’ (vs. ‘no image’) by adding together all the stimuli leading to detections (vs. rejections). We then subtracted the ‘no image’ from the ‘yes image’ to produce a classification image (see Figure 1a, RC, NL, and MJ). For each observer, the classification image represents the template of information that drove the detection of the target ‘S’ letter–formally, it is the best least square linear fits to the detection data.[1]



<u>Figure 1.</u> Results of Experiment 1. A trial consisted in the presentation of one 50 x 50 pixels static white noise image (black-pixel density = 50%) spanning 2 x 2 deg of visual angle. (a) Raw classification images. (b) Distributions of the average squared amplitude energy for different spatial frequencies (collapsed across all orientations) of the raw classification images (expected energy = constant). The solid lines are the best Gaussian fits (RC, mean = 1.9, std =.4 cycles per letter, R^2 = .97; NL, mean = 1.61, std =.4, R^2 =.99; MJ, mean = 1.3 std =.8, R^2 =.968). (c) Classification images filtered with a smooth low-pass (Butterworth) filter with a cutoff at 3 cycles per letter. We squeezed pixel intensities within two standard deviations from the mean. Black 'S's on a white background filling the image are revealed. (d) Best matches between the filtered classification images and 11,284 letters. For RC, the best match is an uppercase ‘S’, in Bold Courier New (*r* = 0.557); for NL, a lowercase ‘S’, in Regular Verdana (*r* = 0.553); and, finally, for MJ, an uppercase Bold Arial ‘S’ (*r* = 0.704).

Input white noise has equal energy at all spatial frequencies. It is therefore unbiased and the expected energy of the classification image is constant across the whole spatial frequency spectrum. Such an “empty” classification image would occur if the observer responded randomly to the white noise stimuli, either because the observer ignores the stimuli, or hallucinates ‘S’s without any systematicity. A superstitious[2] (as opposed to a blind or hallucinating) observer will respond positively to white noise fields when these correlate (even really weakly) with the observer’s internal representation of an ‘S’. Consequently, any bias appearing in the spectral analysis of the raw classification images (see the curves in Figure 1 for RC, NL, and MJ) indicates the presence of structures that underlie the superstitious perceptions of the letter ‘S’. It also provides the means to visualize the observer’s share.

We found such biases for information at slightly different bandwidth, depending on observer (RC, 1.5 to 2.3 cycles per letter, peak = 1.9 cycles; NL, 1.21 to 2.01 cycles per letter; peak = 1.61 cycles; MJ, .05 to 2.1 cycles per letter, peak = 1.3 cycles). Technically, for each observer, we bestfitted a Gaussian density function (see Figure 1, the solid lines) to the energy distribution of his/her raw classification image (Figure 1, the open circles). To determine the observer-specific bias, we computed the mean of each best Gaussian fit and included all spatial frequencies one standard deviation away—that is, a bandwidth comprised between 0 and 3 cycles per letter. We visualized this information by filtering the classification image with a smooth low-pass filter (Butterworth) with a cutoff at 3 cycles. The outcomes are black 'S's on a white background filling the image (see Figure 1c, RC, NL, and MJ). Their spectral compositions are consistent with psychophysical findings that letter identification is most efficient at about 3 cycles per letter (Solomon & Pelli, 1994; Pelli, Burns, Farell & Moore, in press).

To test that the ‘S’s in the classification images were not simply the result of the experimenters’ own superstitious perceptions, we correlated these classification images with the 26 letters of the alphabet from 31 fonts[3], 7 styles (Normal, Italic, Bold, Underline, Outline, Condense, Extend) in upper and lower case, for a total of 11,284 Pearson correlations. All letters were 50 x 50 pixels black on white templates that were horizontally and vertically stretched to maximize correlation with the filtered classification images. For RC, the highest correlation were obtained with an upper-case New Courier font, Bold style ‘S’ letter scaled horizontally (see Figure 1d), for NL, with a lowercase Verdana font, Regular style ‘S’ letter scaled horizontally, and for ML, with an uppercase Arial font, Bold style ‘S’ letter scaled vertically. On average, confounding font, style, case, and observer, the largest correlation between the classification images and the 26 letters of the alphabet was found for the ‘S’ (see Figure 2), a pattern true for each observer.

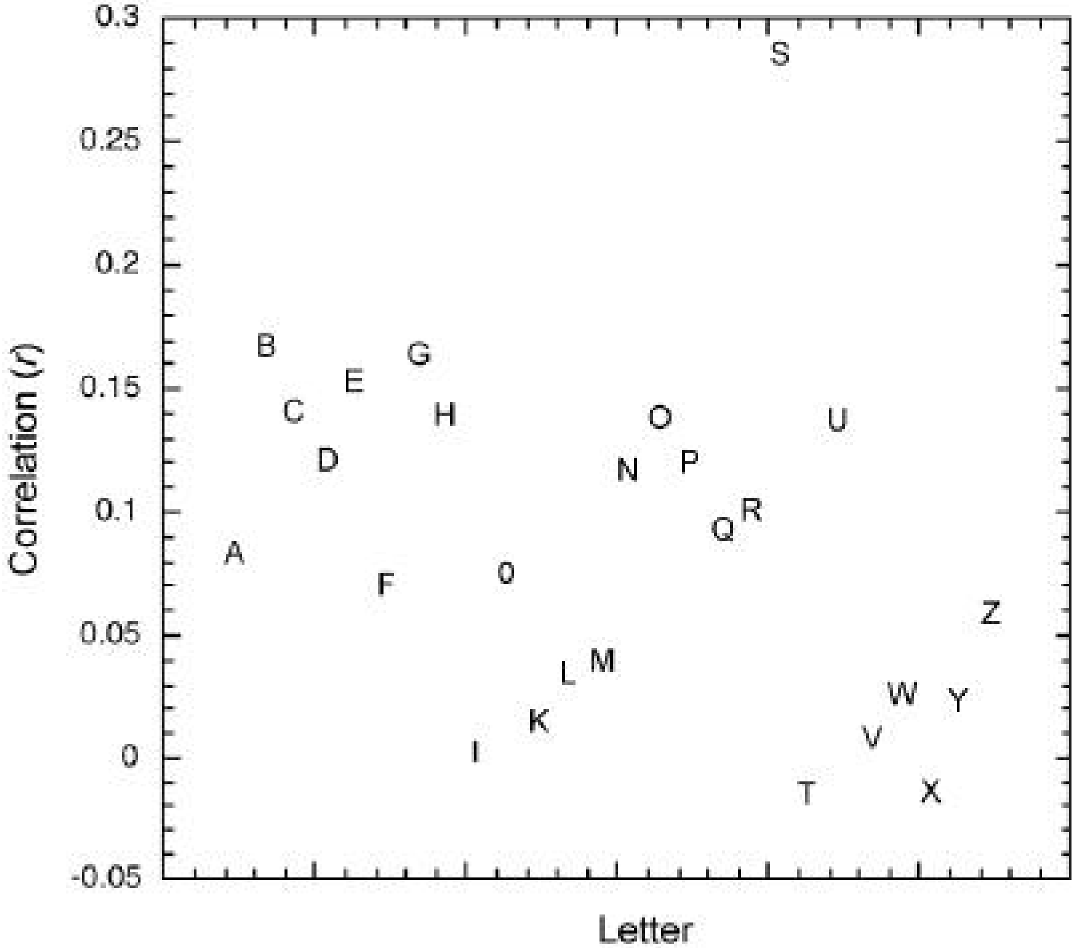


Figure 2. Average Pearson correlation coefficients between the three filtered classification images depicted in Figure 1c, and the 26 letters of the alphabet coming in 31 fonts, 7 styles, upper and lower case for a total of 11,284 stimuli.

In sum, we have induced superstitious perceptions of ‘S’s by instructing three observers to detect this letter in noise. Unknown to them, the stimuli never comprised the letter, but only white noise. Thus, if the observers had been performing only according to the stimulus (i.e., in a bottom-up manner), their classification images should have had the same properties as averaged white noise—i.e., constant energy across all spatial frequencies. However, there were marked peaks of energy below 3 cycles per letter. These must have arisen from top-down influences on the interpretation of white noise—very low correlations between input noise and the memory representations of the letter. Further analyses revealed the shape of the letters that the observers thought they saw.

Experiment 2: Simile smile

Method

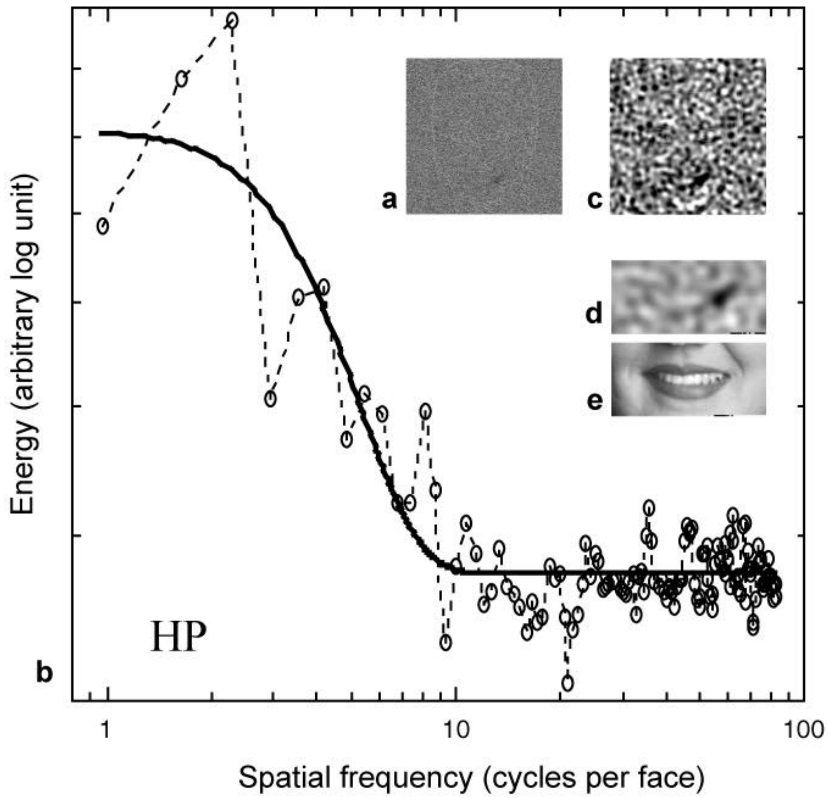
Experiment 2 generalizes this rendering of represented visual information to a more complex representation, using two other observers. We instructed two female observers (AR, aged 26; HP, aged 23) to discriminate between a smiling and non-smiling face embedded in noise. Each observer was told that the smiling face was present on 50% of 20,000 trials equally divided into 40 blocks and ran over a fortnight. No detail was given regarding the alternative expressions to ensure that the observer focussed on detecting the features of a smile. In each trial, one sparse image spanning 256 x 256 pixels (5.72 x 5.72 deg of visual angle) was presented. This image comprised 27.5% of the randomly sampled black pixels of the contours of a face without a mouth (indicated with a red marker in figures 3a and 3c) and, for the remainder, bit noise with the same density of black pixels. No signal was therefore presented in the mouth area. The experiment ran on a Macintosh G4 using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997).

Results

The observers detected a smile on 7.07% (HP) and 48.4% (AR) of the trials. Observer HP explained that she had been very conservative and only responded 'yes' when she was absolutely certain that the face was indeed smiling. She added that she looked for teeth and used the eyes and the nose to locate the mouth. Observer AR reported that she was focusing mostly on the junctions of the lips.

To visualize their internal representations of a smile we first computed the raw classification images as explained above (see Figure 3a, HP and AR). Following spectral analyses of the classification images and Gaussian fits of their energy distributions, an information bias appeared for HP between 0 to 13.69 cycles per face, with a peak at 12.83 cycles (see Figure 3b, HP). For AR, the critical bandwidth was between 0.92 and 5.47 cycles per face, with a peak at 3.192 cycles (see Figure 3b, AR).

These biases are consistent with the most efficient bandwidth described in the identification of face expressions literature–i.e., maximum efficiency centred at 8 cycles per face (Bayer, Schwartz & Pelli, 1998). We revealed this information by filtering the classification images with a low-pass cutoff (at 14.27 cycles per face for HP and 5.84 cycles for AR, see Figure 3c, HP and AR). The outcomes render the internal representation of a smile revealing the teeth for HP, and a smile with well defined junctions of the lips for AR.



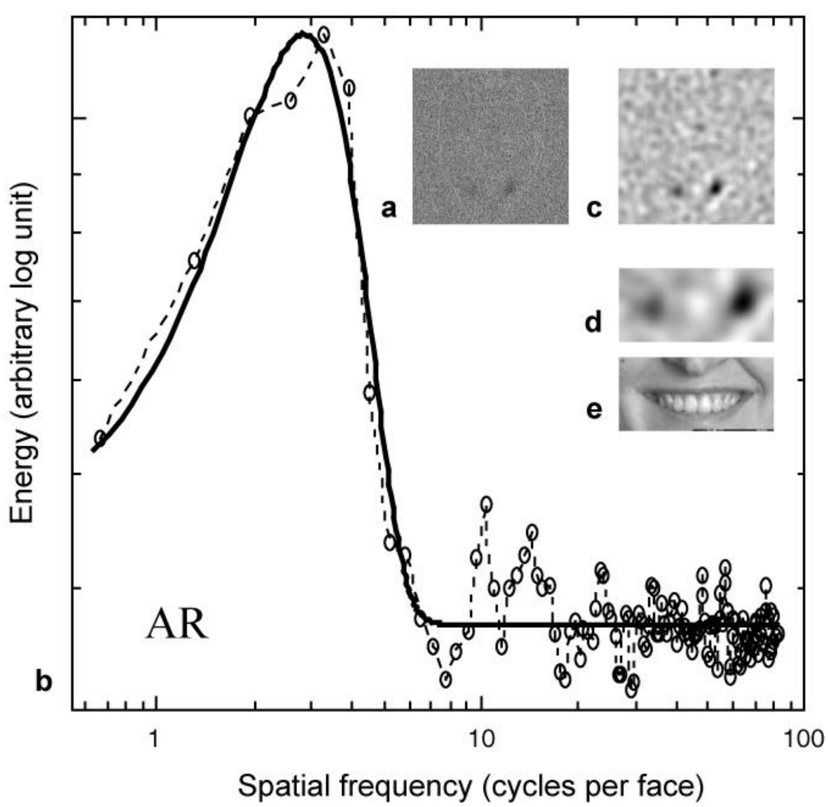


Figure 3. Results of Experiment 2. A trial consisted in the presentation of a 256 x 256 pixels image (5.72 x 5.72 deg of visual angle). This image comprised 27.5% of the randomly sampled black pixels of

the contours of a face without a mouth (indicated with a red marker in (a) and (c)) and, for the remainder, bit noise with the same density of black pixels. (a) Raw classification images. (b) Distributions of the average squared amplitude energy for different spatial frequencies (collapsed across all orientations) of the raw classification images. (Each open circle is the average of two successive data point for subject HP.) The solid lines are the Gaussian bestfits. For HP, this bestfit has a mean =.85 cycles per face and a std = 12.83 (R^2 = .83); For AR, it has a mean = 3.19 cycles per face and a std = 2.28 (R^2 =.95). (c) Classification images filtered with a low-pass (Butterworth) filter (cutoff = 14.27 and 5.84 cycles per face, for HP and AR, respectively). We squeezed pixel intensities within two standard deviations from the mean. Smiles are revealed. (d) Areas of the filtered classification images that were correlated with the corresponding area of 48 face stimuli (3 expressions * (8 men + 8 women)). (e) The face stimuli with the largest correlation (r = .24 and .32, for HP and AR, respectively) with the filtered classification images. Note that the displayed expression is happiness.

As in Experiment 1, we validated the content of the classification image by correlating the mouth area of the classification images (see Figure 3d, HP and AR) with the mouths of 16 individuals (8 males, 8 females from Schyns & Oliva, 1999), each displaying 3 different expressions (neutral, happy and angry) for a total of 48 correlations per observer. The largest correlations were produced with happy women (see Figure 3e, HP and AR). The average correlations for the happy, neutral and angry mouths were .08, -.005, and .002, respectively, for HP; and 0.137, -0.0004, and 0.009, respectively, for AR.

Discussion

We have presented a method to reconstruct unobservable representations from superstitious perceptions. In white noise we elicited superstitious perceptions of ‘S’s in Experiment 1 and of smiles in Experiment 2. Reverse correlation rendered the internal representations underlying these perceptions. Importantly, the internal representations had spectral properties compatible with those reported in recognition studies. This is in line with recent findings demonstrating that some neurons in the human medial temporal lobe respond both to bottom-up visual inputs and to the top-down mental visualizations of these visual inputs (Kreiman, Koch & Fried, 2001). To our knowledge, it is the first time that the representations underlying object recognition are depicted in the absence of a systematic bottom-up signal.

The depiction of complex, psychologically validated object representations from white noise is genuinely new. Studies close in spirit to this research have successfully mapped the low-level receptive fields of single neurons from pseudo-random stimuli (see De Angelis, Ohzawa & Freeman, 1995, for a review). However, these representations depict a level of visual organization considerably lower than object recognition. At such higher levels, noise has been applied to study the nature of illusory contours (Gold et al., 2000), the discrimination of letters (Watson, 1998) or scenes (Ahumada & Beard, 1999; Abbey et al., 1999). However, in all of these cases the statistics of the input space comprises a signal in addition to the noise. This stimulation with a signal plus noise is biased, unlike ours, which only comprise unbiased white noise.

To illustrate this contrast, consider the study of Gold et al. (2000) in which reverse correlation depicted the illusory contours observers perceived to classify noisy Kaniza squares into concave and convex. In 20% of the trials of a within-subjects design the real contours (concave vs. convex) were presented in white noise. This signal biased the input distribution which, in turn, could have biased perception in the ILLUSORY conditions (when only noise is presented). Whether or not the classification images derived in these conditions represent more than input signal artificially internalized for the sole purpose of the experiment is a generic issue that pervades the field. To avoid these difficulties altogether, we went back to Wiener's (1958) original idea and used only white noise to depict the observer's share.

To the extent that white noise can be weakly correlated with every visual stimulus, the technique could be applied to a wide range of visual and auditory events. However, there are also serious limitations to the technique, arising from its linearity. For example, the observer needs to be properly instructed that the same target is always presented, that it does not change position across trials, that it always has the same black on white contrast, and so forth. Even though it is theoretically possible to extend the technique to nonlinear problems (Wiener, 1958), it is practically difficult because of (a) the required number of trials and (b) the required time for data analyses. However, Psychology needs new techniques to characterize the properties of memorized information. Here, we developed a technique that approximates these properties from the superstitious perceptions of objects in white noise.

References

Abbey, C. K., Eckstein, M. P., and Bochurd, F. O. (1999). Estimation of human-observer templates in two-alternative forced-choice experiments. In E. A. Krupinski (Ed.), Proceedings of the Society of Photo-optical Instrumentation Engineers, pp. 284-295, San Diego: SPIE.

Ahumada, A. J. & Lovell, J. (1971). Stimulus features in signal detection. Journal of the Acoustical Society of America, 49, 1751-1756.

Bayer, H. M., Schwartz, O. & Pelli, D. (1998). Recognizing facial expressions efficiently. IOVS, 39, S172.

Beard, B. L. & Ahumada, A. J. (1998). "A technique to extract the relevant features for visual tasks" In B. E. Rogowitz and T. N. Pappas (Eds.), Human Vision and Electronic Imaging III, SPIE Proceedings, 3299, 79-85.

Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision, 10, 433-436.

Gold, J., Murray, R. F., Bennett, P. J. & Sekuler, A. B. (2000). Deriving behavioral receptive fields for visually completed contours. Current Biology, 10, 663-666.

Gombrich, E. H. (1960). Art and Illusion: A Study in the Psychology of Pictorial Representation. London: Phaidon.

Janson, H. W. (1973). "Chance images" In Philip P. Wiener (Ed.), Dictionary of the History of Ideas. New York: Charles Scribner's Sons.

Kreiman, G., Koch, C. & Fried, I. (2001). Imagery neurons in the human brain. Nature, 408, 357-361.

Neri, P., Parker, A. J. & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. Nature, 401, 695-698.

Oshawa, I., DeAngelis, G. C. and Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science, 249, 1037-1041.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437-442.

Pelli, D. G., Burns, C. W., Farell, B. & Moore, D. C. (in press). Identifying letters. Vision Research.

Solomon, J. A. & Pelli, D. G. (1994). The visual filter mediating letter identification. Nature, 369, 395-397.

Watson, A. B. (1998). Multi-category classification: template models and classification images. Investigative Opthalmology and Visual Science, 39, S912.

Wiener, N. (1958). Nonlinear Problems in Random Theory. New York: Wiley.

Authors Note

Frédéric Gosselin, Département de psychologie, Université de Montréal; Philippe G. Schyns, Department of Psychology, University of Glasgow.

We thank Liza Paul, Lizann Bonnar, Benoit A. Bacon, and Simon Garrod for proofreading drafts of this article. This research was supported by ESRC Grant R000237901.

Correspondence concerning this article should be addressed to Frédéric Gosselin, Département de psychologie, Université de Montréal, C.P. 6128, Succursale centre-ville, Montréal, Québec, Canada, H3C 3J7; or to Philippe Schyns, Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow, Scotland, United Kingdom, G12 8QB. Electronic mail may be sent via the Internet to frederic.gosselin@umontreal.ca or to philippe@psy.gla.ac.uk.

Footnotes

[1] We suppose that the observer matches two vectors at each trial of the experiment: a stimulus vector of dimensionality k and a template vector **ß** of the same dimensionality representing the memorized pattern to match against the input (e.g., the letter 'S'). We also suppose that the observer's response is a linear function of this match. We can arrange the n stimulus vectors of the experiment in the $n * k$ matrix **X**. A linear equation then describes the behavior of the observer in the experiment: $\mathbf{y} = \mathbf{ß}\mathbf{X}+\varepsilon$, where $\mathbf{y}$ is a n-dimensional response vector, and ε is an n-dimensional vector of "error" random variables with $E(\varepsilon) = \mathbf{0}$ and $V(\varepsilon) = \sigma^2\mathbf{I}$. The *least square* estimate of **ß** is $(\mathbf{X'X})^{-1}\mathbf{X'y}$. If the stimulus vectors are uncorrelated, we have $(\mathbf{X'X})^{-1} = (k\mathbf{I})^{-1} = k^{-1}\mathbf{I}$. Therefore, $\mathbf{ß} = k^{-1}\mathbf{X'y}$. Leaving the constant k aside and assuming that the responses can only have the values 1 or -1, this last equation reduces to summing all the stimulus vectors that led to a response of 1 and subtracting from it the sum of all the stimulus vectors that led to a response of -1.

[2] We call these perceptions 'superstitious' because the correlation between the information template and input noise is extremely weak (r = .026 on average in Experiment 1, and still smaller in Experiment 2), even if we assume that the observer uses the same unique template and detection criterion throughout the experiment.

[3] Andala Mono, Apple Chancery, Arial, Book Antiqua, Bookman Old Style, Capitals, Century Gothic, Ventury Schoolbook, Charcoal, Chicago, Comic Sans MS, Courier, Courier New, Gadget, Geneva, Georgia, Helvetica, Impact, LED, Letter Gothic MT,

Mishawaka, Monaco, New York, Sand, Techno, Teletext, Textile, Times, Times New Roman, Trebuchet MS, and Verdana.