Superstitious Perceptions

Frédéric Gosselin (GOSSELIF@PSY.GLA.AC.UK)
Philippe G. Schyns (PHILIPPE@PSY.GLA.AC.UK)
Lizann Bonnar (LIZANN@PSY.GLA.AC.UK)
Liza Paul (LIZA@PSY.GLA.AC.UK)
Department of Psychology, University of Glasgow, 58 Hillhead Street
Glasgow, Scotland, G12 8QB UK

Abstract

It has long been observed that we sometime perceive complex scenes in blots, rocks, or clouds, but the phenomenon has attracted little scientific attention. We propose that a weak–or superstitious–match between a memory template and a sparse stimulus is responsible for such perceptions. We provide reverse-correlation evidence for this theory.

Introduction

> If you look at walls that are stained or made of different kinds of stones […] you can think you see in them certain picturesque views of mountains, rivers, rocks, trees, plains, broad valleys, and hills of different shapes […] battles and rapidly moving figures, strange faces and costumes, as well as an infinite number of things […]
>
> (Leonardo da Vinci, *Notebooks*)

We have all seen a human face or a landscape in a cloud floating by, in a pebble lying on a beach, or in blots on a wall. Notorious examples of this phenomenon include the Mars channels and the Man on the Moon; Hermann Rorschach has even made it the basis of a projective test. The earliest known reference to the phenomenon reaches back as far as classical antiquity, and thousands of others have been enumerated (Janson, 1973; Gombrich, 1960). Given this human fascination for the phenomenon, it is surprising how little–if any–scientific attention it has received. Here, we provide evidence that these perceptions result from a weak–or superstitious–match between a memory template and a sparse stimulus. Beyond the anecdotes, a rigorous study of superstitious perceptions could reveal important properties of internal object representations. It is one aim of our research to illustrate this point.

We instructed naïve observers to decide whether one particular target (the letter 'S' in Experiment 1 and a smiling face in Experiment 2) was present or not in stimuli. No signal was ever presented in the stimuli. Each stimulus comprised only two-dimensional static bit "white" noise. White noise has several desirable properties: It has equal energy across the entire spatial frequency spectrum and does not correlate across trials. In other words, white noise does not in itself represent coherent structures in the image plane and across trials.

These properties make white noise the perfect basis for reverse correlation (see Appendix), a statistical technique that uses noise to derive the information the observer *uses to respond* in a particular visual task (e.g., Ahumada & Lovell, 1971; Beard & Ahumada, 1998; Neri, Parker & Blakemore, 1999; Gold, Murray, Bennett & Sekuler, 2000). In Experiment 1, we used reverse correlation (supplemented with careful debriefing) to assess the properties of the letter 'S' that the observers superstitiously perceived (remember that they only saw white noise). Experiment 2 replicated the findings in the more realistic case of faces.

Experiment 1: 'S' as in Superstitious

In this experiment, we asked a first subject to detect in white noise the presence of a black 'S' on a white background filling the image. As just explained, only bit noise was presented.

Method

Subject

One 24-year old female student from the University of Glasgow with normal vision was paid £50 to participate in this study. She was an experienced psychophysical observer, but had no knowledge about the goals of the experiment.

Procedure

The experiment ran on a Power PC Macintosh using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). It comprised 20,000 trials equally divided into 40 blocks. The subject took two weeks to complete the experiment. A 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 subject was told, however, that she was participating in a detection experiment. She was instructed to say whether or not a black letter 'S' on a white background filling the image was present. No more detail was provided about the 'S'. We told her that 50% of the trials were positive. The subject was under no time pressure to respond.

When the 20,000 trials were completed, we debriefed the subject. We asked her the following questions: How often did she see the letter? When she saw it, how noisy was it? What strategy did she use to respond?

Results and discussion

On 22.7% of the trials the subject pressed on the ’yes’ key, indicating that an ’S’ was present. During debriefing, she said that she saw an ’S’ each time she responded positively, and she estimated the quantity of noise in the letter ’S’ to vary between 30% and 50%. She summarized her strategy as follows: "I simply waited to see if the S "jumped out at me"."

All the static bit noise images leading to a ’yes’ response were added together and so were those leading to a ’no’ response. The two resulting images, the ’yes’ and the ’no’ images, were normalized. A raw classification image was then computed by subtracting the normalized ’no’ image from the normalized ’yes’ image. This classification image is the linear template that best explains the behavior of the subject in the least square sense of the term (see Appendix).

There is an objective method to understand the information that drove the illusory perceptions of the ‘S’ in the experiment. As explained earlier, white noise is completely unbiased. If the observer responded randomly (i.e., without having the illusion of the presence of an ‘S’), the classification image would itself be unbiased. From this reasoning, any bias appearing in the spectral analysis of the observer’s classification image should indicate the structures underlying the illusory perceptions. The spectral analysis reveal a bias for information concentrated between 1 and 3 cycles per image, with a peak at 2 cycles per image (see arrow in Figure 1a). This is consistent with Solomon and Pelli's (1994) finding that letter identification is most efficient around 3 cycles per letter.

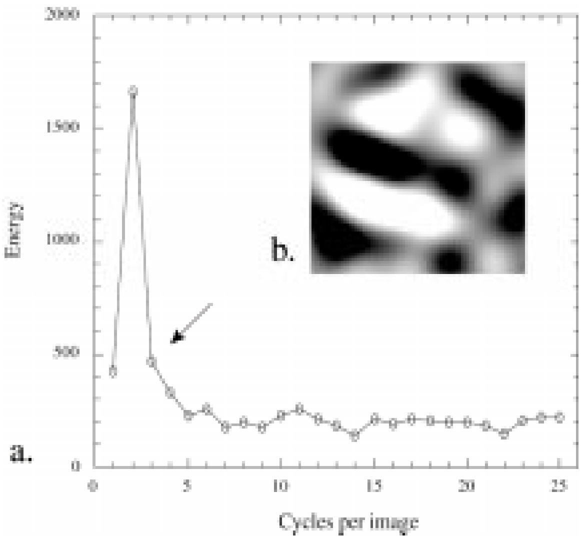

Figure 1. (a) Distribution of energy across the spectrum. (b) Classification image low-passed at 3 cycles per image.

We can visualize the information that drove the illusory detection by filtering the raw classification image with a low-pass Butterworth filter with a cutoff at 3 cycles per image. To provide a better depiction, we further remove all the outlier pixel intensities (two standard deviations away from the mean). The resulting image is a black 'S' on a white background.

To summarize, we have induced illusory perceptions of an ‘S’ by asking one subject to detect this letter in noise. Unknown to her, the stimuli did not comprise the letter, but only white noise. If the subject had been performing only according to the stimulus (i.e., in a bottom-up manner), her classification image should have had the same properties as noise—i.e., having identical energy across all spatial frequencies. However, there was a marked peak of energy between 1 and 3 cycles per degree that could only arise from top-down influences on the interpretation of white noise. Further analyses revealed the precise shape of the letter that the subject thought she saw. Specifically, it is worth pointing out that the best depiction of the information used

Experiment 2: Simile smile

In Experiment 2, we generalized the technique to a more complicated stimulus, using another subject. The task was to discriminate between a smiling and non-smiling face. However, the face presented in noise had no mouth whatsoever.

Method

Subject

One 26-year old female student at the University of Glasgow with normal vision was paid £50 to take part in this study. She was naïve with respect to the goals of the experiment, but was an experienced psychophysics observer.

Procedure

The experiment ran on a Macintosh G4 using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). It consisted in 20,000 trials equally divided in 40 blocks. The subject took three weeks to complete the experiment. 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 black pixels of the contours of a mouthless face (see the white marker in Figure 2b) randomly sampled and, for the remainder, of bit noise with the same density of black pixels. No signal was therefore presented in the mouth area.

The subject was instructed to decide whether the face was smiling or not–no detail was provided regarding the alternative expressions. This ensured that the subject focused on seeking information for "smile". We also told her that the face would be smiling in 50%

of the trials. The subject was under no time pressure to respond. Following the 20,000 trials, we debriefed the subject as in Experiment 1.

Results and discussion

On 7.07% of the trials the subject pressed on the 'yes' key, indicating that the "noisy" face was smiling. During debriefing, she explained that she had been very conservative and that she had only responded 'yes' when she was absolutely certain that the face was indeed smiling. The subject looked for teeth and used the eyes and the nose to localize the mouth.

All the static bit noise images leading to a 'yes' response were added to form a 'yes' image, and all those leading to a 'no' were added to form a 'no' image. A raw classification image was then computed by subtracting the normalized 'no' image from the normalized 'yes' image.

The distribution of energy in the spectrum for the raw classification image is represented in Figure 2a. The energy is concentrated in the bandwidth ranging from 1 to 20 cycles per image (from 0.35 to 12.29 cycles per face–see arrow in Figure 2a). This roughly corresponds to the most efficient bandwidth found by Bayer, Schwartz and Pelli (1998) in an expression identification task (i.e., maximum efficiency centered at 8 cycles per face).

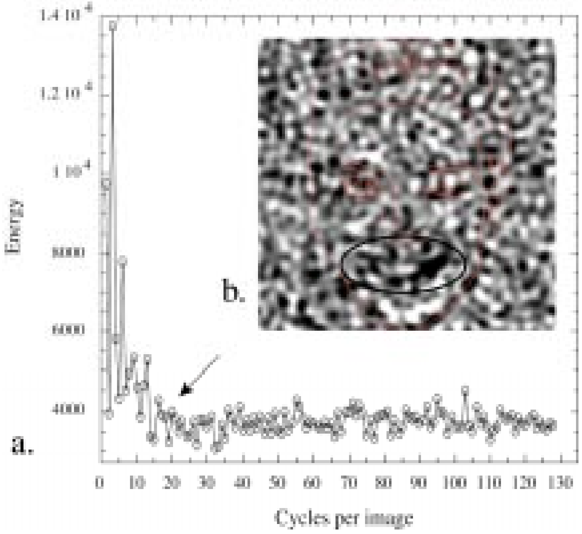

Figure 2. (a) Distribution of energy across the spectrum. (b) Classification image low-passed at 20 cycles per image.

Figure 2b is the raw classification image low-passed at 20 cycles per image with a Butterworth filter–with outlier pixel values removed, followed by a normalization. A white mouthless face marker has been superimposed on filtered classification image. A smile revealing teeth is clearly visible (see circled area in Figure 2b).

Conclusion

The evidence we have gathered in two experiments corroborates the idea that superstitious perceptions result from a weak match between a memory template and a sparse stimulus. We have shown that we could induce superstitious perceptions of a letter ('S', Experiment 1) and part of a face (a mouth expressing a smile, Experiment 2) in bit noise. Reverse correlation demonstrated that observers in these experiments used information from memory ressembling an 'S' and a smile, respectively. It is important to stress that this information did not originate from the signal, by from their memory. It is only because these memory representations are partially correlated with white noise that the superstitious perceptions occur. But then, because white noise is weakly correlated with every visual stimulus, this technique could in principle extend to depicting a wide range of visual representations. In our experiments, these representations had properties expected from what is know in the recognition literature. So, the superstitious perceptions were not random hallucinations, but instead well-constrained perceptions derived from specific knowledge.

Superstitious perceptions could therefore be used to explore the properties of representations in the absence of any bottom-up information.

Acknowledgements

This research was partially funded by ESRC grant R000 237 901.

References

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** (11), 663-666.

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

Janson, H. W. (1973). "Chance images" In Philip P. Wiener (Ed.), *Dictionary of the History of Ideas.* New York: Charles Scribner's Sons.
Neri, P., Parker, A. J. & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, **401** (6754), 695-698.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.
Solomon, J. A. & Pelli, D. G. (1994). The visual filter mediating lettre identification. *Nature*, **369**, 395-397.
Sprent, P. (1969). *Models in Regression and Related Topics.* London: Methuen & Co Ltd.

Appendix

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' or a smiling face).

Suppose further that we arrange the *n* stimuli of the experiment in the $n * k$ matrix **X**. The behavior of the observer for the whole experiment is described by

$$\mathbf{y} = \mathbf{ßX}+\varepsilon,$$

where **y** is an *n*-dimensional vector of decision responses, and ε is an *n*-dimensional vector of "error" random variables with $E(\varepsilon) = \mathbf{0}$ and $V(\varepsilon) = \sigma^2\mathbf{I}$.

For simplicity, the "target present" and "target absent" responses in **y** as well as the white and black pixels in **X** are encoded with values of 1 and –1, respectively.

Given that we know **X** and can observe **y**, we can resolve the linear system of equations by finding **ß**. The *least square* solution requires that we minimize the scalar sum of squares

$$\mathbf{S} = (\mathbf{y}-\mathbf{Xß})'(\mathbf{y}-\mathbf{Xß})$$

for variations in **ß**. Differentiating, we have

$$2\mathbf{X}'(\mathbf{y}-\mathbf{Xß}) = \mathbf{0},$$

which gives, for our least square estimator, the vector

$$\mathbf{ß} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

This is the logic of standard multiple regression (e.g., Sprent, 1969). Because our stimulus vectors are uncorrelated, we have

$$(\mathbf{X}'\mathbf{X})^{-1} = (k\mathbf{I})^{-1} = k^{-1}\mathbf{I},$$

Therefore,

$$\mathbf{ß} = k^{-1}\mathbf{X}'\mathbf{y}.$$

Leaving the constant *k* aside, this equation reduces to summing all stimuli that led to a 'yes' response and subtracting from it the sum of all the stimuli that led to a 'no' responses. This is the essence of reverse correlation.