Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 2515–2520
Vision Research
www.elsevier.com/locate/visres

Internal surface representations approximated by reverse correlation

Frédéric Gosselin [a,*], Benoit A. Bacon [b], Pascal Mamassian [c]

[a] Département de Psychologie, Université de Montréal, C.P. 6128 Succ. Centre-ville, Montréal (Qué), Canada H3C 3J7
[b] Laboratoire de Vision et de Perception, Ecole d'Optométrie, Université de Montréal C.P. 6128 Succ. Centre-ville, Montréal (Qué), Canada H3C 3J7
[c] Psychology Department, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK

Received 10 October 2002; received in revised form 3 November 2003

Abstract

We presented two naïve observers with 20,000 random-dot stereograms. On each trial, the observers had to indicate the presence or absence of a complex 3D pattern (a large '+' sign in relief). However, unbeknownst to them, the stereograms *did not contain any signal, but only disparity noise*. Responses and verbal reports indicate that the observers 'saw' the suggested 3D surface configuration in roughly half the trials even though structured local low-level signal was never presented. Using reverse correlation, we derived an approximation of the internal surface-based representations, or templates, that best accounted for the observers' responses. These templates were shown to be spatially well defined and temporally stable. We propose that the 3D surface-based representations that we derived are the first approximations and depictions of the intermediary process that allows the visual system to successfully link degraded, bottom–up signal and high-level, top–down object recognition.

Keywords: Binocular disparity; Reverse correlation; Stereopsis; Superstitious perception; Surface

1. Introduction

The process through which the visual system recovers complex, three-dimensional (3D) visual scenes from inverted, degraded, bi-dimensional retinal images is still poorly understood. It is however generally agreed that an intermediate process must occur between early information pickup and more complex visual processes such as object recognition. It has been suggested that this intermediary stage involves the construction of a surface-based representation of the visual scene (Marr, 1982; Nakayama, He, & Shimojo, 1995; Pylyshyn, 1999).

The concept of a surface representation stage parsimoniously accounts for an array of visual phenomena. These include modal and amodal completion in which missing information has to be inferred to successfully interpret the geometry of the visual scene (Bacon & Mamassian, 2002; Kanisza & Gerbino, 1982; Kellman & Shipley, 1991; Nakayama et al., 1995; Nakayama, Shimojo, & Silverman, 1989; Tse, 1999, 2002; but also see Rubin, 1921). It has also been proposed that such a surface-based intermediary stage plays a role in more basic visual functions such as depth perception, motion perception and texture segmentation (see He & Nakayama, 1992, 1994a, 1994b; Nakayama & Shimojo, 1990).

To this day, however, this surface representation stage largely remains a theoretical construct inserted between low-level information and high-level vision as part of a serial process. Indeed, 'isolating' this stage for study is made difficult by the fact that a surface representation is inherently anchored in retinal low-level information. Neri, Parker, and Blakemore (1999), for example, used reverse correlation to approximate the surface template that subjects used to solve a stereoscopic task. Their stimuli, however, always contained some disparity signal

* Corresponding author. Tel.: +1-514-343-7550; fax: +1-514-343-2285.
E-mail address: frederic.gosselin@umontreal.ca (F. Gosselin).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2004.05.016

so that the template they inferred could have been directly modulated by the form of the low-level information.

We present here an attempt to extract a surface representation purely determined by the task, and not by any residual signal present in the stimulus. Our demonstration follows the paradigm of 'superstitious perceptions' first used by Gosselin and Schyns (2003) for two-dimensional patterns.

We asked observers to indicate the presence or absence of a complex 3D pattern (a large '+' sign in relief) in random-dot stereograms which, unbeknownst to them, only contained disparity noise. Although the stimuli contained no structured signal, the observers' responses and verbal reports indicated that they 'saw' the suggested configurations.

The absence of structured signal therefore allowed us to bypass the early, low-level stage of visual processing without compromising object recognition or the high-level processes on which it is dependent.

We then used reverse correlation (Ahumada & Lovell, 1971; Beard & Ahumada, 1998; Gold, Murray, Bennett, & Sekuler, 2000; Gosselin & Schyns, 2003; Neri et al., 1999) to give form to the intermediary process that allowed the observers to relate the encoded disparity noise to the detection of the target. In other words, we have revealed the internal representations, or templates, that best accounted for the observer's responses. As our stimuli contained no signal, the internal surface representations that we depict can be said to be purely top–down, in the sense that they are uncontaminated by low-level signal.

2. Methods

One 22 year old female (MB) and one 23 year old male (AF), both undergraduate students at the University of Glasgow, were paid to take part in the experiment. They were experienced psychophysical observers with normal stereoscopic vision but were naive as to the rationale and aims of the experiment.

The participants were asked to complete 40 blocks of 500 trials (approximately 10 h overall) within two weeks. On each trial, a new random-dot stereogram appeared and remained on screen until the observer responded. The participants were instructed to indicate whether a large plus sign (i.e., '+') was present or not by pressing the appropriate key (yes–no paradigm). They were told that the plus sign covered the full length and width of the stimulus area and that it would appear nearer than the background (i.e., in relief). They were also told that the plus sign would be present in 50% of the trials, but would be difficult to perceive due to a large amount of noise. No additional details were provided about the experiment. In particular, it is important to note that the observers were never shown an image of the plus sign without noise before running the experiment.

The observers sat 1 m away from the monitor, placed their chin on a chin-rest and viewed the stimulus pairs through a modified Wheatstone stereoscope. The stimuli were created using the PsychToolbox (Brainard, 1997; Pelli, 1997) for MATLAB and were presented on a 21'' monitor connected to a Macintosh G4 computer. The mid-gray background luminance was set to 18.5 cd/m^2. Both halves of the stereograms subtended 2.470°×2.470° of visual angle (154×154 pixels). They were composed of a white background filled with 700 black texture elements spanning 0.048°×0.048° of visual angle (3×3 pixels). The average density of each half-image of the stereogram (black to white pixel ratio) was 0.232. Each texture element was randomly positioned in the left-eye and was shifted in the other eye equiprobably by either 0.963 (one pixel to the right) or −0.963 (one pixel to the left) arcmin (see Fig. 1). These disparities placed each dot approximately 9 mm in front or behind the screen.

All this information can be summarized in a 'disparity map': the disparities of the 700 texture elements are put at the locations of their center in a 150×150 matrix; the locations corresponding to the absence of texture element are assigned the value of zero.

The number of texture elements and their disparity were chosen such that a single noisy surface was the dominant percept (pyknostereopsis; Tyler, 1983). Crucial to our demonstration is the fact that the random-dot stereograms never contained any consistent signal. Both the position of the dots and the depth at which they appeared were determined in a fully random manner. Of course, the degree of correlation between individual stimuli and the hypothetical cross slightly varies across trials. It is these slight variations that allow the use of reverse correlation. It must be understood, however, that these correlations are very small. Also, each stimulus is much more correlated with a variety of other hypothetical visual objects than they are with a cross. Thus, the main factor in subject responses is clearly their expectation of what is to be seen in the noise. One of us

Fig. 1. Sample random-dot stereogram. Both the location and the depth of the individual texture elements were determined in a fully random manner. The stereograms therefore contained no signal.

has shown this directly in the context of the detection of letters defined by contrast. Observers saw different letters (i.e., 'H' and 'Y') in the *same* sequence of contrast noise fields and approximations of the internal representations that best accounted for the observers' perceptions were derived (Gosselin & Schyns, 2002).

3. Results and discussion

In spite of no signal ever being presented, the observers detected a large plus ('+') sign in 48.8% (MB) and 50.6% (AF) of the trials. When asked about their response strategy, both MB and AF crossed their index fingers together to form a '+' and respectively said "I was looking for two perpendicular bars crossing in the middle of the stimulus" and "I was looking at the intersection of the lines, waiting for the plus to jump out". Recent experiments by Goffaux, Corentin, Schyns, Gosselin, and Rossion (2003) demonstrate that detection in such superstitious perception experiments is accompanied by gamma activation in the infero-temporal region whereas rejection is not. Gamma activation has been linked to object perception (Tallon-Baudry & Bertrand, 1999). This strongly corroborates the verbal reports of our observers.

When asked about the aim of the experiment, both said it was about "detection thresholds". MB spontaneously reported that she thought the independant variable was the amount of noise. Neither suspected that the stimuli might not be present. In other words, the observers genuinely saw large plus signs, which is consistent with other reports of 'superstitious' perceptions (i.e., 'S's and smiles in contrast bit noise; Gosselin & Schyns, 2003).

We used reverse correlation to identify and depict the information that the observers used while they were experiencing 'superstitious' perceptions. For each observer, a 'detection image' and a 'rejection image' were computed by adding all the disparity maps of the stimuli leading to detection and rejection, respectively. We subtracted the rejection image from the detection image to produce a classification image. This classification image is proportional to the best least-square linear fit to the detection data [1]. Fig. 2(a) shows the classification image for each observer with the convention that dark correspond to negative disparities and bright to positive disparities. Dark '+'s are revealed for both observers, consistent with a '+' sign protruding in relief. These classification images are linear approximations of the templates that the observers used to match against the noise. In other words, they are depictions of the internal surface representation that allowed the observer to link low-level information (the noise) and high-level information (the large '+' sign they looked for).

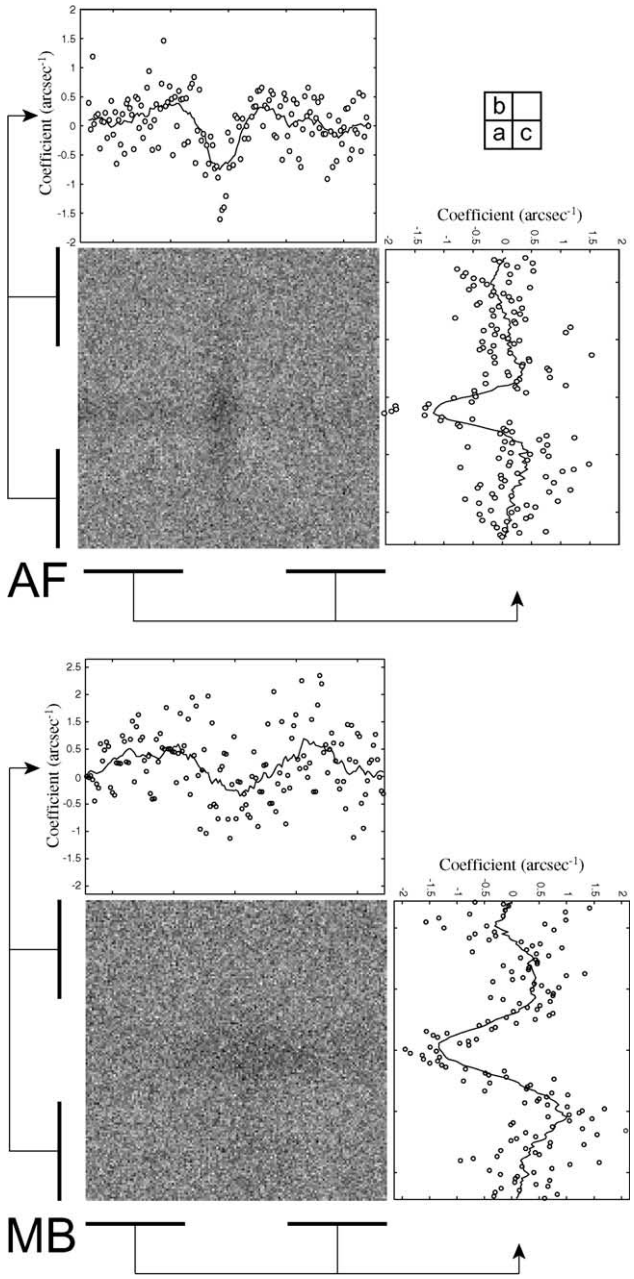

Fig. 2. Results as revealed by reverse correlation. (a) The lower-left panels are the raw classification images. In order to segregate horizontal and vertical information, we averaged the first and last third of the rows (see the two segments joined by a line underneath the classification images) and the first and last third of the columns (see the two segments joined by a line left of the classification images) in the raw classification images in order to obtain the open circles of the (b) top and (c) right scatterplots, respectively. Solid lines are travelling averages of 16 successive data points. In all images, the negative disparity peaks indicate the central bar in depth toward the observers.

The regular, symmetric geometry of the classification images allowed us to segregate their horizontal and

[1] Here, the scaling factor is equal to 150^2 (i.e., the area of a 'disparity map' in pixels)/700 (i.e., the number of texture elements)/20,000 (i.e., the number of trials) = 0.0016.

vertical components. We averaged the first and last third of the horizontal lines (see the two segments joined by a line underneath the classification images) and the first and last third of the vertical lines (see the two segments joined by a line left of the classification images) and obtained the cross-section of the templates shown in Fig. 2(b) and (c) for both observers. The open circles represent the raw averages and the solid lines, the travelling averages of 16 successive data points. This analysis reveals that the horizontal bars were more clearly represented than the vertical bars. More importantly, the cross-section of the templates shows that the information immediately adjacent to the bars, both horizontally and vertically, was also taken into account in interpolating the surface configuration. Indeed, it can be clearly seen that positive disparity lobes stand on both sides of the negative peak that represents the bar itself.

The lobes on either side of the central bars in the classification images suggest an inhibitory mechanism to enhance the representation of the '+' sign against the surrounding regions. In other words, the observers were sensitive not only to the '+' itself but to the full surface configuration in which it was embedded. Such figure-ground segregation is an important aspect of the surface representation stage (Nakayama et al., 1995). Furthermore, the structures of our templates are consistent with what Neri et al. (1999) have found using reverse correlation on disparity signal plus noise.

Disparity 'bit' noise has equal energy at all spatial frequencies. Since it is not biased for any spatial frequency, the expected energy of a randomly produced classification image should therefore be uniform across the whole spectrum. Any bias in the spectral analysis of the classification image would therefore be indicative of structured information of the kind that might underlie superstitious perceptions.

We found such biases in the spectral analysis of the classification images of our two observers, in the range 1–3.236 cycles per image (cpi) for MB and 1–3.945 cpi for AF (Fig. 3(a)). This range is far lower in scale than the local information contained in individual texture elements (indicated by the arrows in Fig. 3(a) for both observers) and is therefore indicative of a more global representation. The peak energy in the classification images can be estimated by fitting a Gaussian function to the energy histogram. The best fit is shown as solid lines in Fig. 3(a) and peaked at 1.167 cpi for MB (std = 2.069; $R^2 = 0.999$) and at 2.303 cpi for AF (std = 1.642; $R^2 = 0.951$). To determine the observer-specific bias range, we included all spatial frequencies within 1.96 standard deviation away for the mean of the bestfit.

We can visualize the information contained in the classification images by filtering them with a low-pass filter (e.g. Butterworth) with a cutoff at 4 cycles per image. The results can be seen in Fig. 3(b), where negative

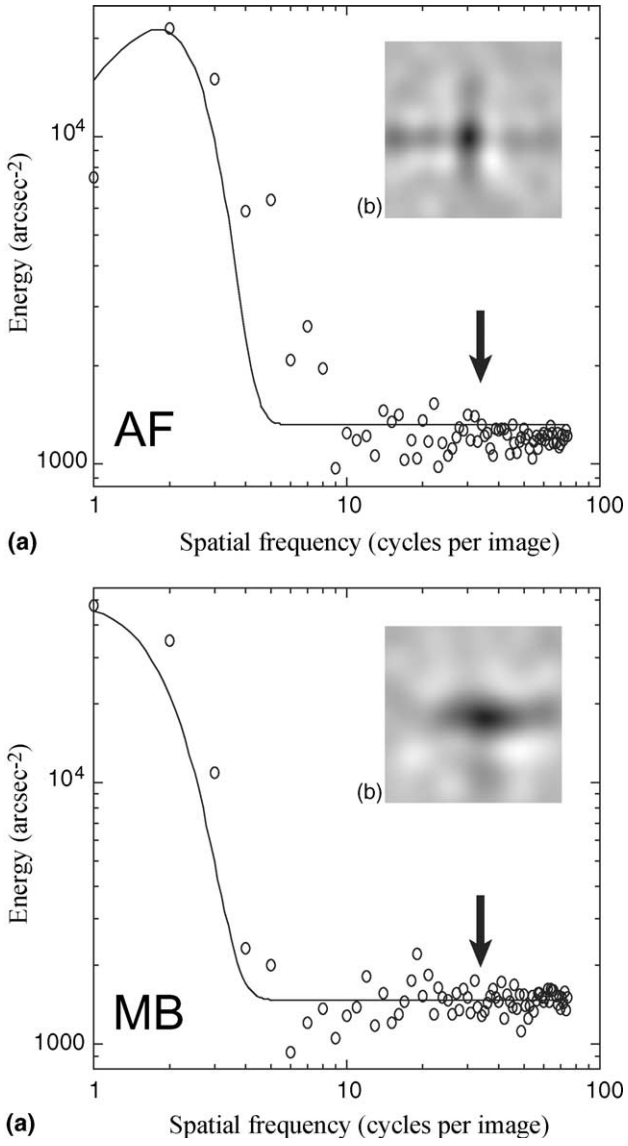

Fig. 3. Spectral analysis of the classification images. (a) The open circles represent the distributions of the average squared amplitude disparity for different spatial frequencies (collapsed across all orientations) of the classification image and the solid lines the best fitted Gaussian functions. Disparity 'bit' noise is by definition unbiased across the spectrum and biases in the spectral analysis therefore reveals structured information. The arrow indicates the spatial frequency range of the individual textured elements, which is considerably higher than the revealed peak. (b) The classification image low-passed by a Butterworth filter with a cutoff at 4 cycles per image. These images depict the templates that best accounted for the observers' behavior in the detection task.

disparities are again depicted in dark and positive disparities in bright. For. both observers, Fig. 3(b) reveals a large dark '+' plus on a bright background.

From this data it is possible to quantify the similarity between our observers' internal surface representations and individual stimulus. As we have said previously it is necessary to have some variance in this similarity to use reverse correlation. However, for our purposes (i.e.

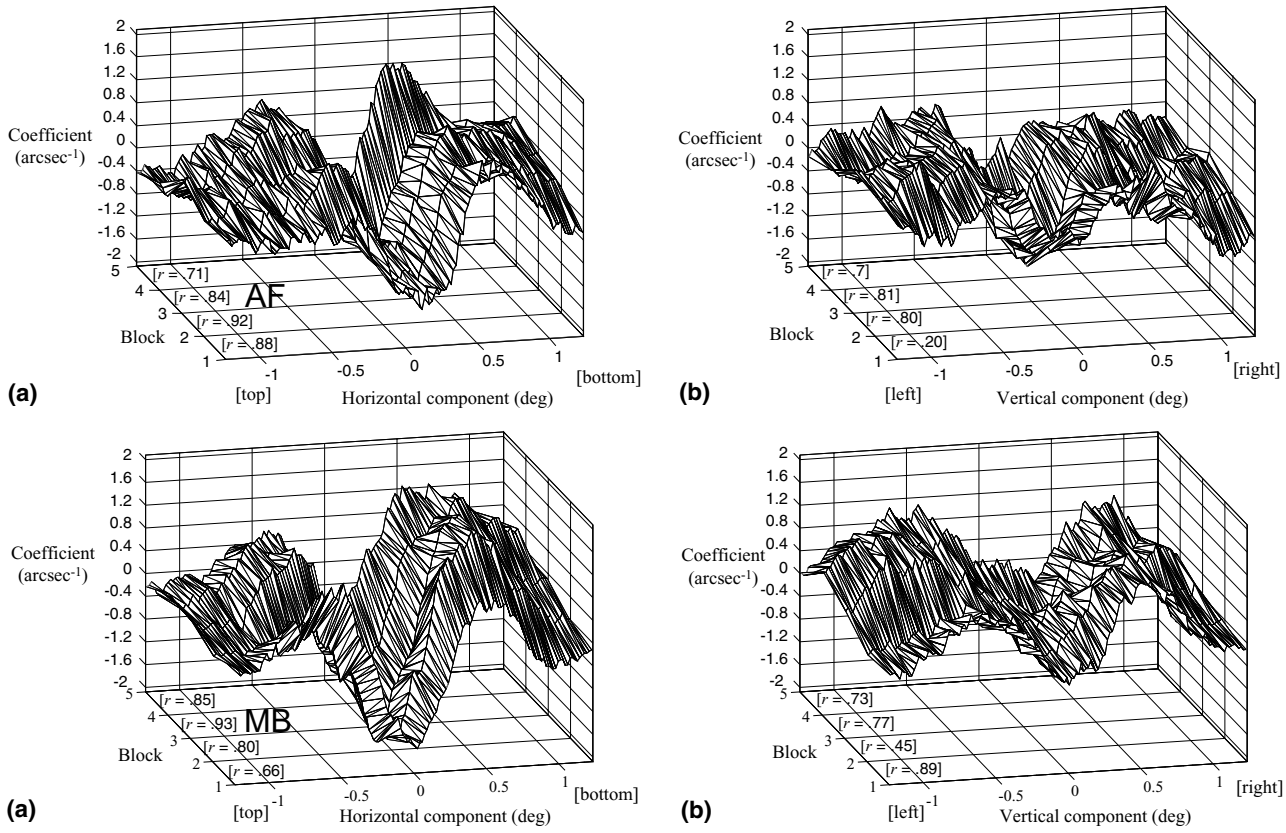

Fig. 4. Stability of the template across time. In order to quantitatively assess the stability of the template, we broke down the information derived from all trials in Fig. 2 into five consecutive blocks of 4000 trials. The solid line for each block corresponds to the travelling average of 30 successive data points. These solid lines were then connected at corresponding vertices by straight lines to form a surface. Pearson correlations are given for all successive blocks. Both the (a) horizontal and the (b) vertical components were remarkably stable over time.

to isolate top–down processes), it is likewise important that it is minimal. The standard deviations of the distributions of Pearson correlations between the subjects' filtered classification images and the 20,000 noise fields are, respectively, $6E-5$ (mean = $1E-4$) and $6E-5$ (mean = $6E-5$) for AF and MB. Certainly, the noise fields were much more correlated with a variety of other hypothetical visual objects than they were with a cross (Gosselin & Schyns, 2002).

Because the classification images were computed from the sum of all trials, it is crucial that the percepts and the templates are stable over time. Verbal reports confirmed that the percepts did not qualitatively vary across blocks and Fig. 4 quantitatively demonstrates the consistency of the templates. The trials were split in 5 blocks of 4000, and the horizontal (Fig. 4(a), both observers) and vertical (Fig. 4(b), both observers) components were again segregated. The high-correlation of the curves across both dimensions (see the Pearson correlations between successive blocks in Fig. 4) indicates consistency and supports the claim that both the templates and the percepts were precise, stable and well defined.

4. Conclusions

We have elicited 'superstitious' perceptions of a complex 3D surface configuration in pure disparity noise. The absence of structured low-level signal allowed us to study the surface representation stage in isolation from lower processes. Not only are presentations without signal optimal statistically, they are to this day the only tool available to fully isolate top–down mechanisms from bottom–up interference. Using reverse correlation, we have approximated and depicted the internal surface representations, or templates, that the observers "superimposed" on the noise in order to do the task. On the basis of these internal representations, or templates, they did not merely search for cross-like disparities, but actively perceived the cross that they had in mind. Both percepts and templates were shown to be temporally stable and spatially well defined. We have thus revealed internal surface representations uncorrupted by any consistent low-level signal. We have shown that these surface representations could account for the way in which observers performed in the 'superstitious' perception task and, more generally, for the

way the visual system infers complex 3D structure from low-level visual information.

Acknowledgments

This work was made possible by a Human Frontier Science Program grant (RG0109/1999-B) awarded to Pascal Mamassian, and by an NSERC (R0010085) and an NATEQ (R0010287) grant awarded to Frédéric Gosselin. Benoit A. Bacon was supported by NSERC PDF—242082—2001. The authors would like to thank the observers for their patience.

References

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

Bacon, B. A., & Mamassian, P. (2002). Amodal completion and the perception of depth without binocular correspondance. *Perception, 31*, 1037–1045.

Beard, B. L., & Ahumada, A. J. (1998). A technique to extract the relevant features for visual tasks. In Rogowitz, B. E.,, & Pappas, T. N. (Eds.). *Human vision and electronic imaging III, SPIE proceedings* (Vol. 3299) (pp. 79–85). Bellingham, WA: SPIE.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Goffaux, V., Corentin, J., Schyns, P. G., Gosselin, F., & Rossion, B. (2003). Superstitious perceptions of faces revealed by phase-locked gamma oscillations. *Journal of Vision, 3*(9), 94.

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.

Gosselin, F., & Schyns, P. G. (2002). White noise reveals properties of internal representations. *Journal of Vision, 2*(7), 692.

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

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359*, 231–233.

He, Z. J., & Nakayama, K. (1994a). Apparent motion determined by surface layout, not by disparity or three-dimensional distance. *Nature, 367*, 173–175.

He, Z. J., & Nakayama, K. (1994b). Surface shape, not features determines apparent motion correspondence. *Vision Research, 34*, 2125–2136.

Kanisza, G., & Gerbino, W. (1982). Amodal completion: seeing or thinking. In J. Beck (Ed.), *Organization and representation in perception*. Mahwah: Erlbaum.

Kellman, P., & Shipley, T. (1991). A theory of object interpolation in object perception. *Cognitive Psychology, 23*, 141–221.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Company.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: its relation to image segmentation, grouping and the recognition of occluded objects. *Perception, 18*, 55–68.

Nakayama, K., & Shimojo, S. (1990). Da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research, 30*, 1811–1825.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In S. M. Kosslyn, & D. N. Osheron (Eds.), *Invitation to cognitive science*. Cambridge: MIT Press.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 40*(1), 695–698.

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

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341–423.

Rubin, E. (1921). *Visuall wahrgenommene Figuren*. Copenhagen: Gylden Kalske Boghandel.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Science, 3*, 151–162.

Tse, P. U. (1999). Volume completion. *Cognitive Psychology, 39*, 37–68.

Tse, P. U. (2002). A contour propagation account of surface filling-in and volume formation. *Psychological Review, 109*, 91–115.

Tyler, C. W. (1983). Sensory processing of binocular disparity. In C. M. Schor, & K. J. Ciuffreda (Eds.), *Vergence eye movements: basic and clinical aspects* (pp. 199–295). Boston, MA: Butterworth.