

Nonaccidental Properties Underlie Shape Recognition in Mammalian and Nonmammalian Vision

Brett M. Gibson,^{1,*} Olga F. Lazareva,²
Frédéric Gosselin,³ Philippe G. Schyns,⁴
and Edward A. Wasserman²

¹Department of Psychology
University of New Hampshire
Conant Hall
10 Library Way

Durham, New Hampshire 03824

²Department of Psychology
University of Iowa

Iowa City, Iowa 52242

³Département de Psychologie

Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal, Québec H3C 3J7
Canada

⁴Centre for Cognitive Neuroimaging

Department of Psychology
University of Glasgow
Room 612

58 Hillhead Street
Glasgow G12 8QB
Scotland

Summary

An infinite number of 2D patterns on the retina can correspond to a single 3D object. How do visual systems resolve this ill-posed problem [1] and recognize objects from only a few 2D retinal projections in varied exposure conditions? Theories of object recognition rely on the nonaccidental statistics of edge properties [2–7], mainly symmetry, collinearity, curvilinearity, and cotermination. These statistics are determined by the image-formation process (i.e., the 2D retinal projection of a 3D object [4]); their existence under a range of viewpoints enables viewpoint-invariant recognition. An important question in behavioral biology is whether the visual systems of nonmammalian animals have also evolved biases to utilize nonaccidental statistics [8, 9]. Here, we trained humans and pigeons to recognize four shapes. With the Bubbles [10] technique, we determined which stimulus properties both species used to recognize the shapes. Both humans and pigeons used cotermination, the most diagnostic nonaccidental property of real-world objects, despite evidence from a model computer observer that cotermination was not the most diagnostic pictorial information in this particular task. This result reveals that a nonmammalian visual system that is different anatomically from the human visual system [11–13] is also biased to recognize objects from nonaccidental statistics.

Results and Discussion

Comparative research is vital for our understanding of vision. When members of different species respond similarly to the same visual information, we gain confidence in the prominence of this information (e.g., nonaccidental statistics), irrespective of cultural or genetic influences. Birds represent an important group to compare with mammals, the other major class of warm-blooded, highly mobile, visually oriented animals [11–13]. Because of the unique demands of flight, for the last 200 million years birds have been under strong evolutionary pressures to keep their overall size to a minimum. Although a very large portion of the avian central nervous system is devoted to visual processing [14], the bird brain is still just a fraction of the size of our own. It is this extraordinary mixture of visual competence and small size that makes the study of birds critical to our understanding of the general mechanisms of visual cognition. Thus, three pigeons and six humans participated in our two-phase investigation into the role of nonaccidental statistics in the recognition of simple objects.

In the first phase, pigeon and human observers were subjected to a four-choice recognition task in which they learned to discriminate grayscale images of four objects (see Figure 1). Upon learning to recognize the shapes to criterion, subjects entered a second phase in which Bubbles determined the information that both species used to identify the shapes. On each trial of Bubbles testing, a shape was randomly selected, and its information was partially revealed via a number of randomly located Gaussian apertures. We then used the observer's response to ascertain the image properties underlying identification of each of the four shapes. To provide a benchmark for the information used in testing the two species, we also included a performance-matched model computer observer that knew perfectly the images of the objects and the location of the apertures (see Experimental Procedures). The human participants were divided into a "no-noise" group and a "noise" group (see Experimental Procedures). For the participants in the no-noise group, the number of bubbles sampling the images was adjusted on a trial-by-trial basis to maintain the same performance level as that of the pigeons. For the participants in the noise group and the model observer, the number of bubbles was the average number that the pigeons were administered, and noise was added to the images on a trial-by-trial basis to maintain the same performance level as that of the pigeons. Thus, the noise group provided an additional comparison of human performance with the model observer when both observers encountered noise.

For each kind of observer (pigeon, human, and model), those image pixels that significantly correlated with the performance of one or more observers ($S_r = 4,791$ pixels; $FWHM = 18.84$; $Z_{crit} = 3.24$; $p < .05$) are shown in color overlaying grayscale images of the objects in Figure 1. The three basic colors (key in the center

*Correspondence: bgibson@cisunix.unh.edu

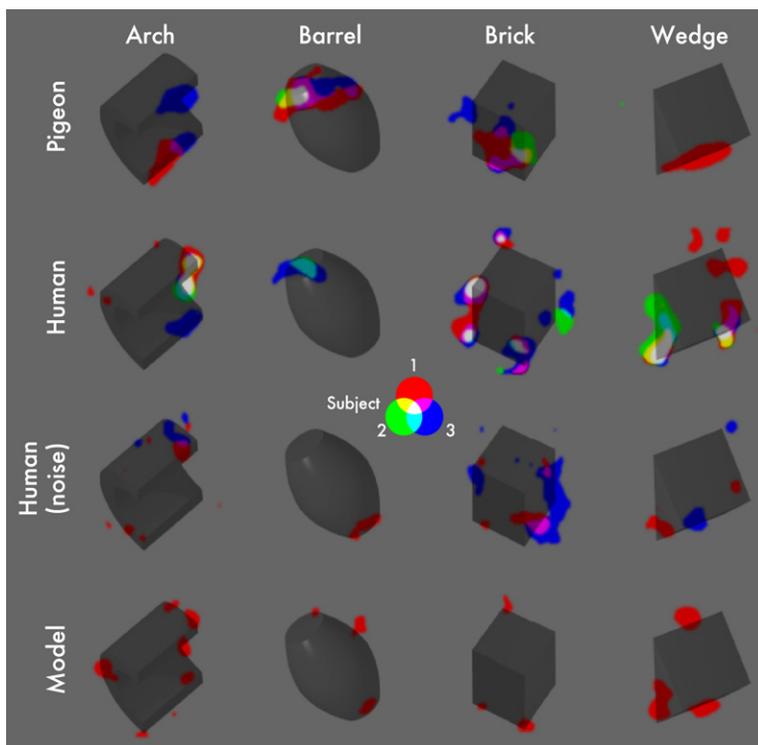


Figure 1. Statistically Thresholded Classification Images for Each Species with Each of the Four Objects Taken into Consideration Each of the four columns corresponds to one of the four objects, whereas each of the four rows corresponds to one of the four observer groups. The three basic colors (key in the center of the figure) indicate the pixels that were used by individual observers; combinations of different basic colors indicate overlap in the use of pixels by two or three observers. The color pixels overlay grayscale images of the objects and indicate the regions that reached statistical significance for the individual classification images.

of the figure) indicate the pixels used by individual observers; combinations of different basic colors indicate overlap in the use of pixels by two or three observers. To formally determine the correspondence between significant image pixels and possible object properties, we precisely defined three regions of interest (ROIs) representing: (a) cotermination information, the most informative nonaccidental property in the real world [3], (b) edge information, another nonaccidental property [2], and (c) shading information, an accidental property of the chosen shapes [2, 15] (see Figure 2 and Experimental Procedures). For both species and

the model observer, we computed the percentage of ROIs ($N_{cotermination} = 6,423$, $N_{edge} = 4,872$, and $N_{shading} = 7,792$ pixels) containing significant pixels; this is a good measure of information use because it factors out the size of the ROIs. Bonferroni-corrected tests were applied within species on all pairwise differences between the percentages (*family-wise* $p < .05$; $Z_{crit} = 2.13$).

Pigeons and humans in both the no noise and noise groups used coterminations (18.7%, 14.4%, and 2.2% of ROIs containing significant pixels, respectively) more than edges (10.3%, 9.5%, and 1.6% of ROIs containing significant pixels, respectively; $Z = 13.42$,

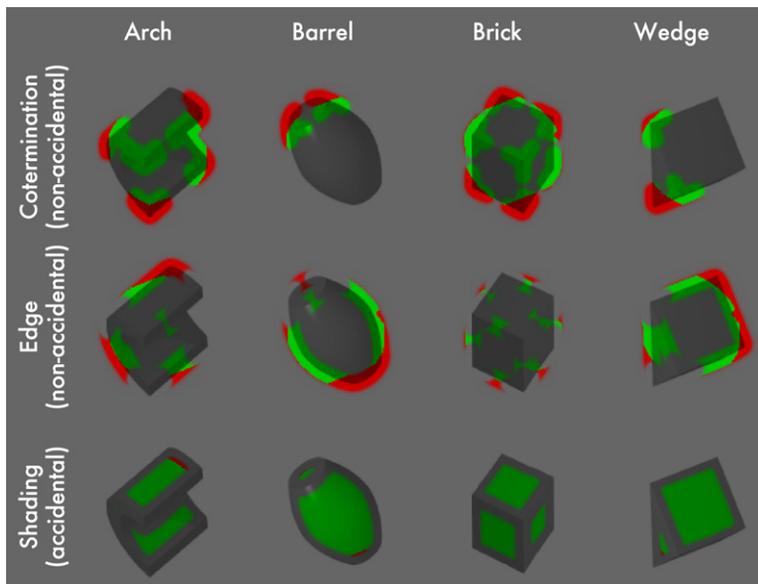


Figure 2. Cotermination, Edge, and Shading Regions for Each of the Four Objects

Each of the four columns corresponds to one of the four objects. Each of the three rows corresponds to one of the three regions of interest (ROIs) (coterminations, edges, and shading). Overlaid on the grayscale objects, the color pixels indicate the object information inside (green) and outside (red) the intersection of the area occupied by the four objects.

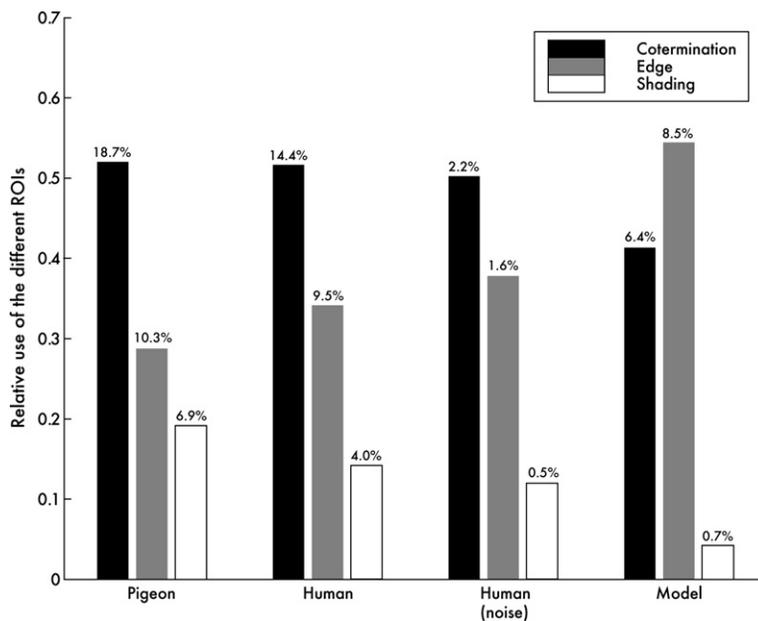


Figure 3. The Relative Use of Cotermination, Edge, and Shading Regions for Each of the Observers

Standardizing the percentage of ROIs (cotermination, edge, and shading) containing significant pixels (indicated above each bar) to sum to 1 within each subject group facilitated the comparison of the histogram bars.

$Z = 8.52$, and $Z = 2.24$, respectively) and more than shading (6.9%, 4.0%, and 0.5% of ROIs containing significant pixels, respectively; $Z = 17.43$, $Z = 17.80$, and $Z = 7.10$, respectively). Also, pigeons and humans in both the no-noise and noise groups used edges more than shading ($Z = 7.67$, $Z = 13.79$, and $Z = 6.76$, respectively). In contrast, the model observer used edges (8.5% of ROIs containing significant pixels) more than coterminations and shading (6.4% and 0.7% of ROIs containing significant pixels, respectively; $Z = 4.40$ and $Z = 23.34$, respectively), and it used coterminations more than shading ($Z = 15.38$). Figure 3 illustrates the relative use of the different ROIs for each subject group. The raw percentages of ROIs containing significant pixels (indicated above each bar) sum to 1 within each subject group to facilitate comparison.

The performance of humans in both the no-noise and noise groups was similar and indicated that the introduction of noise did not make the human performance more similar to the performance of the model. This is not to say that introducing noise did not alter the performance of humans; it did lead to new object regions being used in some instances (e.g., barrel) and reduced the number of significant pixels overall (for one human participant from the noise group, not a single focal region correlated significantly with performance). The results suggest that both pigeons and people utilized nonaccidental cotermination information even though this information is not the most diagnostic for distinguishing among the present pictorial stimuli, as demonstrated by the model observer. Notably, the pattern of information use remains unchanged when the percentage of significant pixels falling in each ROI or ROI normalized for size is used as an alternative measure (data not shown). One concern with the current results is that the people might appear to have been more consistent in their use of information than the pigeons. Of all significant pixels in the biological species' classification images, however, there was 24% overlap among the

human participants who did not have additive noise, 17% overlap among the pigeons, and only 8% overlap among the human participants who did have additive noise. Thus, there does not appear to be a robust difference between biological species. It appeared that one pigeon contributed primarily to the classification image for the wedge stimulus; that bird recognized the wedge 80% of the time, whereas it recognized the arch, barrel, and cube 60%, 69%, and 58% of the time, respectively. Obviously, this bird had an especially effective strategy for recognizing wedges; it consistently used the base of the wedge, a portion also used by the model observer. Why did the other pigeons' wedge-classification images not contain significant pixels? It does not appear that these birds had trouble with the wedge or that they responded randomly; their average correct responses were 63%, 56%, 52%, and 49%, respectively, for the arch, barrel, cube, and wedge. It is either that these birds did not employ a wedge-recognition strategy that was stable over time or that the strategy that they used involved too large an area of the object.

Findings from a recent study [16] involving stimuli similar to those in the current experiment indicated that pigeons may pay more attention to the surface cues of these objects than to the edges. One important difference between the studies is that the tasks for assessing the use of object features are quite different. Pigeons were trained to discriminate multiple views of the four shapes with shading information (to promote learning) before being tested for transfer to line images of the same shapes (without shading) [16]. The pigeons failed to transfer, suggesting that they did not use common information between the shaded and line-drawing versions of the same objects. One advantage of the Bubbles technique is that a transfer task is not required to ascertain the prevalence of one type of information over the other (e.g., shading over nonaccidental edge properties). Also, from a formal standpoint, edges are defined as sharp changes in shading [17], implying an

edge-extraction process that extends beyond the exact location of the edge in the image, one that detects a local transition in global uniformity.

As mentioned earlier, nonaccidental properties help humans to resolve the ill-posed problem of object recognition [2–8]. In nonhuman primates, neurons of the inferior temporal cortex have also been shown to represent objects [18, 19], as well as to encode nonaccidental properties [7, 20]. Our work has disclosed a bias toward cotermination in a phylogenetically distant non-mammalian visual system. The measure of a computational theory is the possibility of multiple, system-specific implementations of a generic set of constraints [21]. Evidence of such generic biases in mammalian and nonmammalian visual systems confirms the ubiquitous nature of nonaccidental properties in the phylogenetic or ontogenetic emergence of object-recognition systems, irrespective of their anatomical structure. Understanding how avian visual systems solve problems that require considerable computational prowess may lead to future technological advances (for example, small visual prosthetics for the visually impaired) in the same way that understanding visual processing in honeybees has led to the development of flying robots and unmanned helicopters [22, 23].

Experimental Procedures

Observers and Experimental Set-Up

Three adult feral pigeons were individually housed and maintained at 85% of their ad lib weights by the use of controlled feedings of mixed grain; the birds had free access to water. The pigeons were studied in operant chambers equipped with a responsive touchscreen and a CRT on the front wall for stimulus display. Pigeons' object-recognition responses were recorded from yellow, blue, red, and green report areas that were located to the NW, NE, SW, and SE of the display area, respectively. Six adult human participants (three females and three males, mean age = 26.6 years, std = 4.9 years) with normal, or corrected-to-normal, vision participated in the experiment. Humans were studied with a Macintosh PowerBook computer; they indicated their recognition response with specific keyboard key presses.

Training Phase

On each trial of the training phase, a 128×128 pixel (for humans, spanning 3.67×3.67 degrees of visual angle at a viewing distance of 0.5 m) grayscale image representing one of four geometrical shapes (arch, barrel, brick, and wedge, see Figures 1 and 2) was randomly selected. For pigeons, the four colored report areas appeared, and the response was recorded; food was delivered after a correct response. Humans responded by depressing the appropriate keyboard key and received immediate feedback. The trial was repeated until the correct response was made. The training phase continued until criterion was reached (a minimum of 80% correct responses to each stimulus and an average of 85% correct responses to all four stimuli).

Testing Phase

On each trial of the testing phase, the geometrical shapes were partially revealed by a mid-gray mask punctured by several Gaussian punch holes of 8 pixels (0.23 degrees of visual angle, for humans) of standard deviation (called "bubbles"). For the participants in the no-noise group, the number of bubbles sampling the images was adjusted on a trial-by-trial basis with the QUEST algorithm [21] to maintain the same performance level (58% correct) as that of the pigeons (see below). Humans in this group required on average 5.95 (std = 3.23) bubbles. For the participants in the noise group and the model observer, the number of bubbles was maintained at 38, the average number that the pigeons were administered. We added Gaussian noise to the bubble images and varied the

signal-to-noise ratio with the QUEST algorithm [21] to maintain model performance at 58% correct, the performance level of the other observers. For each trial, the model determined the Pearson correlation between the sparse noisy input and each of the four possible geometric shapes partially revealed with the same bubble mask; the highest correlation determined the response. The six human participants completed two blocks of testing with the bubbled images; each block comprised 500 trials, for grand totals of 3,000 trials in both the no-noise and the noise group. The model observer completed a total of 9,600 trials, like the pigeons did.

For pigeons, bubble numbers were adjusted every 10 days of testing (20, 40, 50, 50, 30, 30, 40, 40, 40, and 40 bubbles) so that performance between chance and ceiling levels (mean = 58% correct, std = 14%) would be maintained. During each daily session, the pigeons were presented 40 bubbled geometric shapes interspersed among 160 unbubbled geometric shapes. The pigeons were tested over 80 days.

Bubbles Analysis

We performed least-square multiple linear regressions on the bubbles and accuracy data [10] to pinpoint the features that different observers used to discriminate the objects. The plane of regression coefficients yielded by this operation is called a classification image [24]. We computed one such classification image per observer per geometric shape. We smoothed all classification images (with a Gaussian kernel with sigma identical to the sigma of the bubbles used in the experiment) and Z scored the resulting images. To estimate the parameters of the distribution of the null hypothesis, we used the area of the classification images that did not contain a signal (i.e., the complement of the intersection of all of the object areas). Next, we applied the Pixel test to each classification image and determined the number of significant pixels in the ROIs [25]. Tests on the difference of percentages with Bonferroni corrections for multiple comparisons allowed us to assess the reliability of the results.

Regions of Interest

Prior to Bubbles testing, we precisely defined the ROIs for the nonaccidental and accidental properties considered in our analyses. Cotermination information was defined in terms of the contours falling within a radius of 15 pixels from the actual coincidences of two or more edges (color pixels in Figure 2, row 1). The contours of the objects were extracted using the Canny method implemented in the Image Processing toolbox for Matlab. These fine contours were convolved with a Gaussian kernel with a sigma of 4 pixels to allow for some spatial uncertainty. The coincidences of two or more contours were annotated by a human observer. Edge information was defined in terms of the contours that were not included in the coterminations (color pixels in Figure 2, row 2). Finally, shading information was defined in terms of the object area that was neither edge information nor cotermination information (color pixels in Figure 2, row 3). Only the edges and surfaces within the region that defined the intersection of the four objects were retained for each of the three classes defined above (green pixels in Figure 2). We discarded the objects' area outside the intersection of the four objects (red pixels in Figure 2) because it always contained a mixture of accidental position information and either nonaccidental cotermination information or nonaccidental edge information. The model observer used accidental *position* information within this region because it used all the available information. We do not know, however, whether the pigeons and the humans used this accidental information because it is confounded with cotermination and edge information. Indeed, 36% of cotermination and 35% of edge pixels—but only 1% of shading pixels—fall outside the intersection of the shape areas. Each biological species might have used position, cotermination, or edge information, or they might have used a combination of position and either cotermination or edge information. Note that this observation reinforces our main argument: humans and pigeons behaved unlike the model observer; they focused (with 63% of all their significant pixels) on the shape areas containing most (65%) of the nonaccidental features, especially coterminations, whereas the model observer focused (with 63%) on the shape areas containing accidental position information.

Compliance

The use of human and nonhuman animal participants in this study adhered to the policies of each country and institution.

Received: August 3, 2006

Revised: November 30, 2006

Accepted: December 1, 2006

Published online: February 1, 2007

References

1. Bertero, M., Poggio, T., and Torre, V. (1988). Ill-posed problems in early vision. *Proc. IEEE* 76, 869–889.
2. Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychol. Rev.* 94, 115–147.
3. Binford, T.Q. (1981). Inferring surfaces from images. *Artif. Intell.* 17, 205–244.
4. Lowe, D.G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artif. Intell.* 31, 355–395.
5. Dickinson, S.J., Pentland, A.P., and Rosenfeld, A. (1992). 3-D shape recovery using distributed aspect matching. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 174–198.
6. Hoffman, D.D., and Richards, W.A. (1985). Parts of recognition. *Cognition* 18, 65–96.
7. Vogels, R., Biederman, I., Bar, M., and Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *J. Cogn. Neurosci.* 13, 444–453.
8. Lowe, D. (1985). *Perceptual organization and visual recognition* (Norwell, MA: Kluwer Academic Publishers).
9. Mysore, S.G., Vogels, R., Raiguel, S.E., and Orban, G.A. (2006). Processing of kinetic boundaries in Macaque V4. *J. Neurophysiol.* 95, 1864–1880.
10. Gosselin, F., and Schyns, P.G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vis. Res.* 41, 2261–2271.
11. Jarvis, E.D., Guenther, O., Bruce, L., Csillag, A., Karten, H.J., Keunzel, W., Medina, L., Paxinos, G., Perkel, D.J., Simizu, T., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.* 6, 151–159.
12. Butler, A.B., and Hodos, W. (1996). *Comparative vertebrate neuroanatomy: Evolution and adaptation* (New York: Wiley-Liss).
13. Hodos, W. (1993). The visual capabilities of birds. In *Vision, Brain, and Behavior in Birds*, H.P. Zeigler and H.-J. Bischof, eds. (Cambridge, MA: MIT Press), pp. 265–283.
14. Shimizu, T., and Bowers, A.N. (1999). Visual circuits of the avian telencephalon: Evolutionary implications. *Behav. Brain Res.* 98, 183–191.
15. Tjan, B.S., and Legge, G.E. (1998). The viewpoint complexity of an object recognition task. *Vis. Res.* 38, 2335–2350.
16. Peissig, J.J., Young, M.E., Wasserman, E.A., and Biederman, I. (2006). The role of edges in object recognition by pigeons. *Perception* 34, 1353–1374.
17. Marr, D. (1982). *Vision* (San Francisco, CA: W.H. Freeman and Company).
18. Logothetis, N.K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563.
19. Sigala, N., and Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320.
20. Kaye, G., Biederman, I., and Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *J. Neurosci.* 23, 3016–3027.
21. Watson, A.B., and Pelli, D.G. (1983). QUEST: A Bayesian adaptive psychometric method. *Percept. Psychophys.* 33, 113–120.
22. Floreano, D., Zufferey, J.C., and Nicoud, J.D. (2005). From wheels to wings with evolutionary spiking neurons. *Artif. Life* 11, 121–138.
23. Muratet, L., Doncieux, S., Briere, Y., and Meyer, J.-A. (2005). A contribution to vision-based autonomous helicopter flight. *Rob. Auton. Syst.* 50, 195–205.
24. Eckstein, M.P., and Ahumada, A.J., Jr. (2002). Classification images: A tool to analyze visual strategies. *J. Vis.* 2, 1x.
25. Chauvin, A., Worsley, K.J., Schyns, P.G., Arguin, M., and Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *J. Vis.* 5, 659–667.