

Running Head: Non-accidental properties underlying shape recognition . . .

Non-accidental properties underlie shape recognition in mammalian and non-mammalian vision

Brett M. Gibson<sup>1</sup>, Olga F. Lazareva<sup>2</sup>, Frédéric Gosselin<sup>3</sup>, Philippe G. Schyns<sup>4</sup>,  
Edward A. Wasserman<sup>2</sup>

<sup>1</sup>Department of Psychology, University of New Hampshire,

<sup>2</sup>Department of Psychology, University of Iowa, <sup>3</sup> Département de psychologie, Université de

Montréal, <sup>4</sup>Centre for Cognitive Neuroimaging, Department of Psychology,

University of Glasgow

Address for correspondence:

Brett Gibson, Ph.D.  
The University of New Hampshire  
Department of Psychology  
Conant Hall  
10 Library Way  
Durham, NH 03824  
Ph: 603.862.1569  
FAX: 603.862.4986  
Email: [bgibson@cisunix.unh.edu](mailto:bgibson@cisunix.unh.edu)

## Summary

An infinite number of 2D patterns on the retina can correspond to a single 3D object from the outside world. How do visual systems resolve this essentially ill-posed problem [1] and recognize objects from only a few 2D retinal projections in varied conditions of exposure? Theories of object recognition rely on the non-accidental statistics of edge properties [2-7]: mainly, symmetry, collinearity, curvilinearity, and cotermination. These statistics are entirely 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 non-mammalian animals have also evolved biases to utilize non-accidental statistics [8, 9]. Here, we trained humans and pigeons to recognize four shapes. With *Bubbles* [10], we determined which stimulus properties both species used to recognize the shapes. We discovered that both humans and pigeons used cotermination, the most diagnostic non-accidental 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 non-mammalian visual system that is dramatically different anatomically from the human visual system [11-13] is also biased to recognize objects from non-accidental 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. non-accidental 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]. Due to the unique demands

of flight, birds have been under strong evolutionary pressures for the last 200 million years 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 four humans participated in our two-phase investigation into the role of non-accidental statistics in the recognition of simple objects.

In the first phase, pigeon and human observers were subjected to a 4-choice recognition task in which they learned to discriminate greyscale images of four objects (see Figure 1). Upon learning to recognize the shapes to criterion, a second phase began 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. We also included a performance-matched model computer observer who knew the images of the objects and the location of the apertures to provide a benchmark for the information used in testing the two species (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 per 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. 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 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 (ROI) representing: (a) *Cotermination information*, the most informative non-accidental property in the real world [3], (b) *Edge information*, another non-accidental 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*  $p < .05$ ;  $Z_{crit} = 2.13$ ). As can be seen in Figure 3, 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;  $Z = 13.42$ ,  $Z = 8.52$ , 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 ( $Z = 15.38$ ). 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, as it did lead to new object regions being used in some instances (e.g., barrel). The results suggest that both pigeons and people utilized non-accidental 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 even ROI normalized for size is used as alternative measure (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 and 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] using similar stimuli as 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. The 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 non-accidental edge properties). Also, from a computational 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, non-accidental properties help humans to resolve the ill-posed problem of object recognition [2-8]. In nonhuman primates, inferior temporal cortex neurons have also been shown to represent objects [18, 19], including the encoding of non-accidental 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 non-mammalian visual systems confirms the ubiquitous nature of non-accidental 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 using 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 colored report areas that were located to the NW, NE, SW, and SE of the display area, respectively. Six adult human participants (3 females and 3 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 x 128 pixel (for humans, spanning 3.67 x 3.67 degree 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 Figure 1) was randomly selected and the response was recorded. For pigeons, the four colored report areas appeared and the response was recorded; food was delivered following 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 (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-grey mask punctured by several Gaussian punch holes of 8 pixels (0.23 degrees of visual angle, for humans) of standard deviation (called “bubbles,” see Figure 1, Row 2). For the participants in the no noise group, the number of bubbles sampling the images was adjusted on a trial per trial basis using 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 bubbles (std = 3.23) over two blocks of 500 testing trials. 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, varying 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 shape images partially revealed with the same bubble mask; the highest correlation determined the response. The three human participants completed two blocks of testing with the bubbled images; each block comprised 500 trials, for a grand total of 3,000 trials. The model observer completed a total of 9,600 trials like the pigeons.

For pigeons, bubble numbers were adjusted every 10 days of testing (20, 40, 50, 50, 30, 30, 40, 40, 40, and 40 bubbles) to maintain performance between chance and ceiling levels (mean = 58% correct, std = 14%). 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 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 regions of interest [25]. Tests on the difference of percentages with Bonferroni corrections for multiple comparisons assessed the reliability of the results.

### Regions of Interest (ROIs)

Prior to *Bubbles* testing, we precisely defined the ROIs for the non-accidental/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 coterminations 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 not included in the contours (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. 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 non-accidental cotermination information or non-accidental 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 either position, cotermination, or edge information; or 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 non-accidental 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.

## 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 Cognit Neurosci*, 13, 444-453.

8. Lowe, D. (1984). *Perceptual organization and visual recognition*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
9. Mysore, S. G., Vogels, R., Raiguel, S. E., and Orban, G. A. (2006). Processing of kinetic boundaries in Macaque V4. *J of 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. et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nat 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 H. P. Zeigler & H.-J. Bischof (Eds.), *Vision, Brain, and Behavior in Birds* (pp. 265-283). Cambridge, MA: MIT Press.
14. Shimizu, T. & Bowers, A. N. (1999). Visual circuits of the avian telencephalon: Evolutionary implications. *Behavioural Brain Research*, 98, 183-191.

15. Tjan B. S., and Legge G. E. (1998). The viewpoint complexity of an object recognition task. *Vis Res*, 38, 2335-50.13.
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. Kayeart, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *J Neurosci*, 23, 3016-3027.
21. Watson, A. B. & Pelli, D. G. (1983) QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys*, 33 (2), 113-20.
22. Floreano, D., Zufferey, J. C., & Nicoud, J. D. (2005). From wheels to wings with evolutionary spiking neurons. *Artif Life*, 11, 121-138.

23. Muratet, L., Doncieux, S., Briere, Y., & Meyer, J.-A. (2005). A contribution to vision-based autonomous helicopter flight. *Robotics and Autonomous Systems*, 50, 195-205.
24. Eckstein, M. P., and Ahumada, A. J., Jr. (2002). Classification images: A tool to analyze visual strategies. *J Vis*, 2, i-i, <http://journalofvision.org/2/1/i/>, doi:10.1167/2.1.i.
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, <http://journalofvision.org/5/9/1/>, doi:10.1167/5.9.1.

## Figure Captions

*Figure 1.* 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 of the individual classification images that reached statistical significance.

*Figure 2.* Each of the four columns corresponds to one of the four objects. Each of the three rows corresponds to one of the three types of object information (cotermination, edges, and shading). Overlaid to 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.* Percentage of ROIs (cotermination, edge, shading) containing significant pixels for each of the classes of observers.

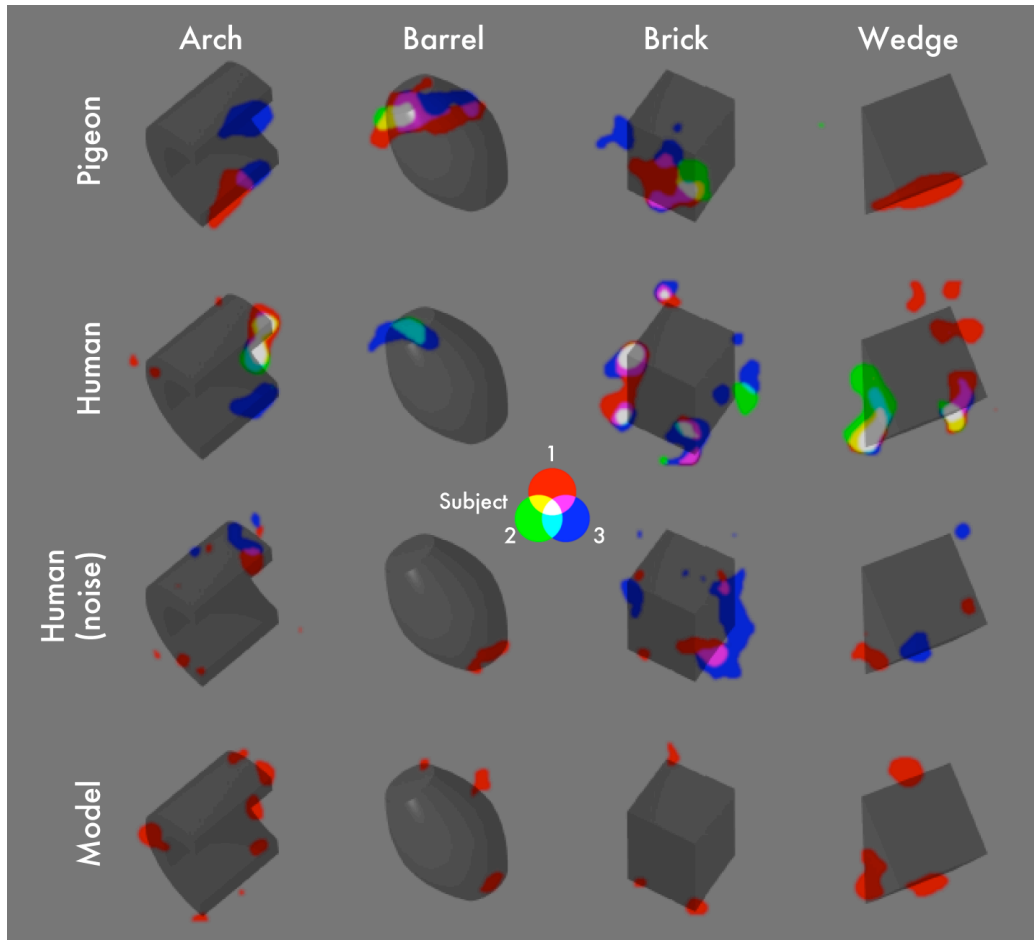


Figure 1

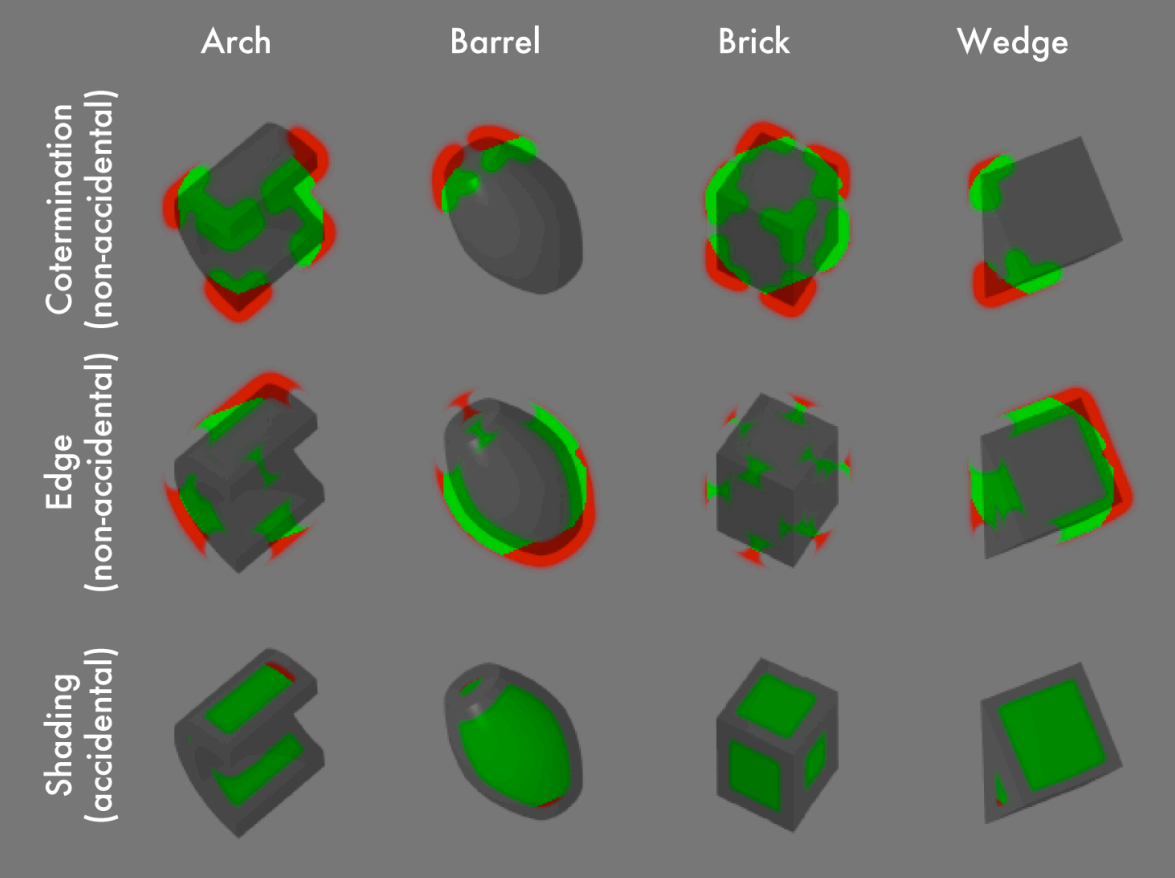


Figure 2

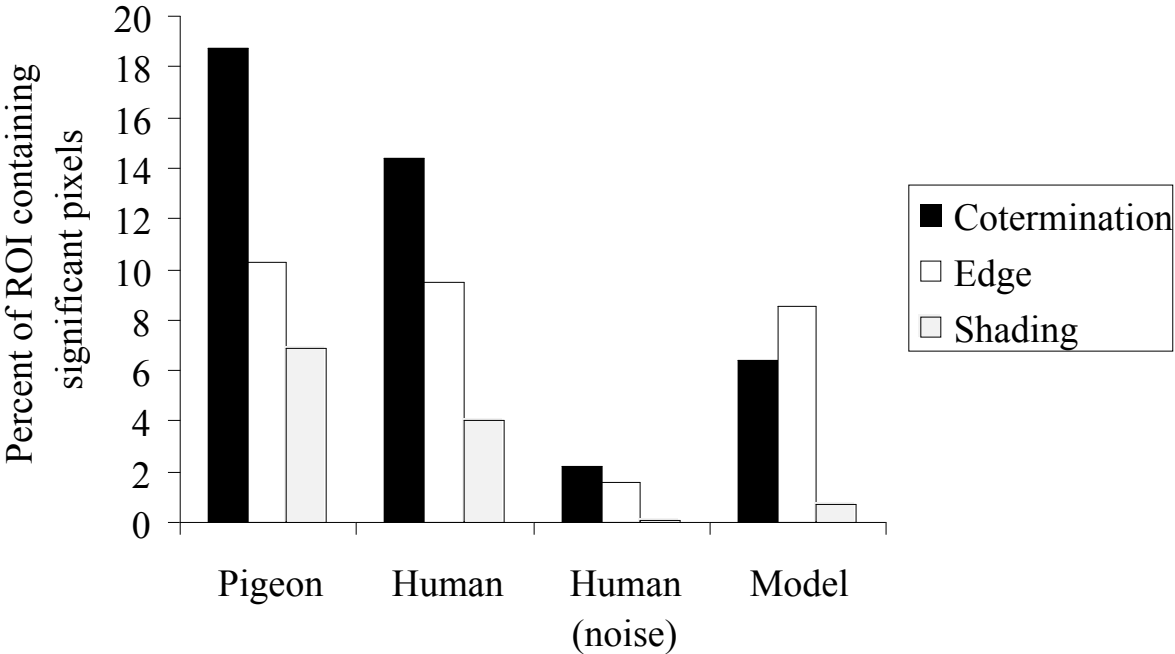


Figure 3