The role of symmetry in scene categorization by human observers

John D. Wilder[1], Morteza Rezanejad[2], Sven Dickinson[1], Allan Jepson[1], Kaleem Siddiqi[2], and Dirk B. Walther[3]

[1]Department of Computer Science, University of Toronto
[2]Centre for Intelligent Machines and School of Computer Science, McGill University

[3]Department of Psychology, University of Toronto

Humans accurately classify images of real-world scenes following very brief exposure durations (Potter & Levy, 1969; Thorpe, Fize, Marlot, et al., 1996; VanRullen & Thorpe, 2001). However, the mechanisms and features used by the brain for this task are not well understood. Walther et al. (2011) showed that line drawings of such real world scenes capture the essential structural information required for rapid scene classification. Gestalt psychologists have long argued for symmetry as a grouping principle for object detection and recognition. Here we demonstrate its importance for scene categorization in human vision.

We made use of a novel method for measuring local symmetry in a line drawing of a natural scene in order to assign each contour pixel a symmetry rating. For each scene in our data set, all contour pixels were ranked according to their symmetry rating and the top half were used to create a symmetric binary image, while the remaining ones were used to create an asymmetric binary image. These two images contained no overlap, and when combined resulted in the original intact line drawing. On each trial one line drawing (either intact, symmetric, or asymmetric) was briefly presented (53ms) and was followed by a perceptual mask. Observers then categorized the scene into one of six scene categories (beaches, forests, mountains, city streets, highways, offices). The categorization accuracy of observers was significantly higher for the 50% more symmetric contours (49.7%) than for the 50% less symmetric ones (38.2%). Our results demonstrate that in addition to its role in object detection and recognition, symmetry appears to be crucial for the perception of complex real-world scenes by humans.

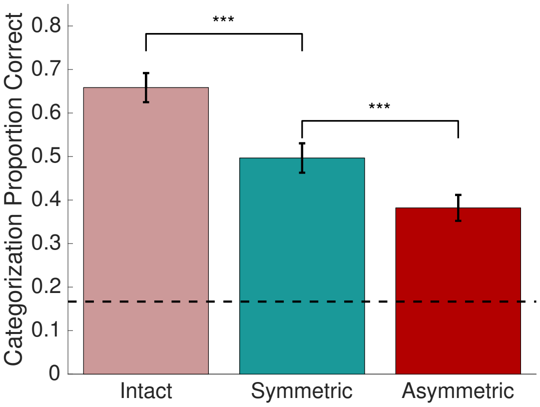
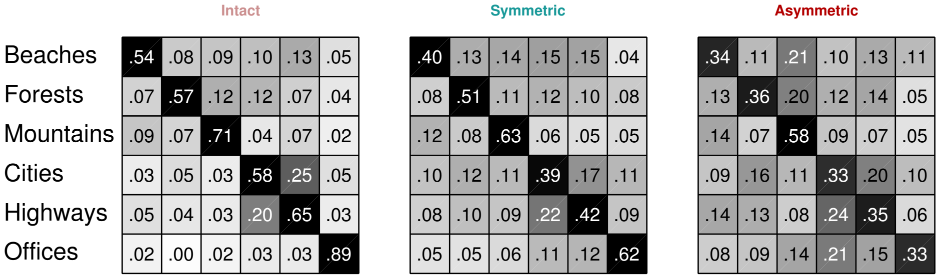


Figure 1. Categorization performance. Participants performed best with intact scenes (65.8%). Participants had moderate performance with symmetric scenes (49.7%), and worst performance with asymmetric scenes (38.2%). Performance was significantly better in the symmetric condition than asymmetric condition ($t = 5.2, df = 11, p = 0.0003$).



Intact

Beaches	.54	.08	.09	.10	.13	.05
Forests	.07	.57	.12	.12	.07	.04
Mountains	.09	.07	.71	.04	.07	.02
Cities	.03	.05	.03	.58	.25	.05
Highways	.05	.04	.03	.20	.65	.03
Offices	.02	.00	.02	.03	.03	.89

Symmetric

Beaches	.40	.13	.14	.15	.15	.04
Forests	.08	.51	.11	.12	.10	.08
Mountains	.12	.08	.63	.06	.05	.05
Cities	.10	.12	.11	.39	.17	.11
Highways	.08	.10	.09	.22	.42	.09
Offices	.05	.05	.06	.11	.12	.62

Asymmetric

Beaches	.34	.11	.21	.10	.13	.11
Forests	.13	.36	.20	.12	.14	.05
Mountains	.14	.07	.58	.09	.07	.05
Cities	.09	.16	.11	.33	.20	.10
Highways	.14	.13	.08	.24	.35	.06
Offices	.08	.09	.14	.21	.15	.33

Figure 2. Confusion matrices (averaged across observers). The rows represent the true scene category, and the columns represent the human response. Good performance results in large values on the diagonal; errors result in larger values off the diagonal. For example, intact scenes have a strong diagonal as performance was highest in that condition. In order to see how similar the errors are between conditions we compute the correlation between the off diagonal elements of the confusion matrix. The intact and symmetric errors have the largest correlation ($r = 0.76$). The correlation between the intact and asymmetric was slightly smaller ($r = 0.6$).