investigation of many aspects of individual differences and synesthesia that should inform and inspire valuable research.

References


Reviewed by: Michelle R. Greene, Minerva Schools at KGI, San Francisco, CA

It’s been 40 years since Molly Potter’s first published observation that human observers viewing a series of photographs being shown at a rate of 8 per second can detect a particular photograph merely from a brief description of its contents, such as “picnic” (Potter, 1975). Twenty years later, Simon Thorpe et al. demonstrated that observers could detect the presence of an animal within a complex scene with only 20 ms of exposure to the image. Event-related potentials (ERPs) recorded during this task indicate that the upper bound for visual recognition is 150 ms after stimulus onset (Thorpe, Fize, & Marlot, 1996). These remarkable observations raise the question of how we make such rapid sense of the “stimulating scenery mess” (in Bar’s words) of our visual world. These studies, along with other pioneering works of Biederman in the 1970s, paved the road for scene understanding to be a respectable topic of study within high-level vision. Since then, there have been a number of remarkable demonstrations of human scene understanding abilities in memory (Konkle, Brady, Alvarez, & Oliva, 2010), visual search (Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011), and other domains. Therefore, this edited volume on scene understanding by Kveraga and Bar is especially timely and will be of interest to the readership of *Perception*.

The book is organized into 14 chapters and spans much of the entire range of scene understanding research. Although each chapter can provide a good stand-alone introduction to a particular perspective, experimental technique, or theoretical approach, there is a narrative flow to the chapter ordering that marches not only through different experimental fields (from experimental psychology to neuroscience to computation) but also provides complementary perspectives to similar questions, such as how we stitch single views into a coherent whole place (Chapter 1 by Intraub and Chapter 3 by Park and Chun), the role of context in perception (Chapter 7 by Aminoff and Chapter 8 by Fabre-Thorpe), or how the content of images interfaces with our emotions (Chapter 12 by Joshi et al., Chapter 13 by Bradley et al., and Chapter 14 by Kveraga).

Chapter ordering also highlights significant differences of opinion, especially with regard to the function of the parahippocampal place area (PPA), a brain region that shows selective activation to scene images. While Park and Chun (Chapter 3) and Epstein (Chapter 6) view the PPA as participating in the processing of scene geometry and layout, Rajimehr et al. (Chapter 4) note that PPA activity is biased toward certain low-level visual features, such as rectilinearity, that are shared with scenes, calling into question the scene-specificity of PPA activity. Aminoff (Chapter 7) views the PPA as performing associative and contextual
processing rather than visual processing, while Epstein, anticipating this view, questions the robustness and reliability of these findings. Given the heterogeneity of results and opinions, I am inclined to agree with Epstein’s final analysis that PPA activity cannot be explained by a single cognitive mechanism and may reflect the presence of subareas that are only beginning to be isolated using functional magnetic resonance imaging (fMRI).

The book is also well balanced in history, providing both discussions of earlier pioneering works as well as state-of-the-art techniques. In particular, Chapter 9 by Potter beautifully synthesizes her nearly 50-year career examining the memory representations of briefly viewed scenes. Thorpe (Chapter 10) as well as Stansbury and Gallant (Chapter 11) bring us up to date on bleeding edge computational modeling, including the newly rediscovered convolutional neural networks (Thorpe) and linear encoding models (Stansbury and Gallant).

A recurring theme of this work is the broader question of what is a scene? As Brooks et al. (Chapter 5) note, “Everyone knows what a scene is, to twist a famous quote by William James, but the concept is nonetheless surprisingly hard to define” (p. 85). The authors within this volume provide a number of different answers. Park and Chun note that there is a fuzzy boundary between objects and scenes, and this sentiment is echoed in Thorpe’s chapter. This question becomes a bit thorny when we consider the PPA, often defined as the voxels that have more activity to scenes than to objects. Are very large entities such as airplanes objects or scenes? Viewed from the outside, an airplane may very well be considered an object. However, we can be inside of an airplane, making it also a scene. Consider too buildings that can also serve as landmarks for navigation and are large enough to create the shape of a street scene’s layout. Are these objects or part of the structure of a scene? Are these questions merely a matter of scale, as suggested by Thorpe or are scenes and objects different types of natural kinds?

A related question involves what types of primitives a scene can be decomposed into and a number of different suggestions are made including spatial frequencies (Malcolm and Schyns, Chapter 2), objects (Aminoff and Fabre-Thorpe), and spatial layout (Park and Chun, Epstein). As scenes are complex constructs containing information at all of these levels, Rajimehr et al. (Chapter 4) note the difficulties in attributing neural selectivity to a particular level of analysis.

A serious question for scene understanding is in understanding how photographs depicting a place relate to the place itself, although few authors take up this issue directly, with Intraub being a notable exception. This is particularly relevant for the chapters on affect, as one might question whether a threatening photograph has the same salience as a threatening real-world situation.

The book has a very light editorial touch, with only a brief introduction from Bar at the beginning of the volume. This can be an advantage or disadvantage depending on your perspective. In my view, I would appreciate a bit of editorial analysis and synthesis, especially as this book raises so many meaty questions. However, most chapters strike the right balance between review and analysis so that I found myself discussing and arguing along with the authors as I read, and I feel that this is the mark of a successful edited volume.

Scenes are complex entities depicting events, agents, and objects and containing myriad textures, colors, and motion. As such, their study has spanned several academic fields, including experimental psychology, cognitive and systems neuroscience, computer vision and graphics, as well as esthetics. Kveraga and Bar have done a wonderful job bringing together all of these perspectives into one accessible volume, and I believe that researchers in all corners of high-level vision will find this volume of value.
References


