

On the perception of natural scenes: an introduction to the special issue

Benjamin W. Tatler

School of Psychology, University of Dundee

In this special issue we consider a range of current approaches to understanding aspects of how we visually inspect and encode scenes. This issue follows up a symposium on natural scene perception that was held at the 14th European Conference on Eye Movements in Potsdam in 2007. In this period of extensive and diverse study of scene perception it is important to consider how the variety of approaches and topics studied relate to one another and to the overall aim of this area of visual Psychology. Ultimately our aim must be to understand how vision serves our operation during natural behaviour in real environments. Yet, due to obvious technological and methodological limitations, much of the scene perception work to date has used simplified stimuli and experimental paradigms that fall short of truly natural settings. To achieve the necessary experimental rigour it has been necessary to tackle issues of scene perception under conditions that allow greater experimental control. The balance between naturalness and control in recent years has often come from using photographic images of natural scenes or computer rendered scenes. It is these two-dimensional representations of real scenes that are those studied in most of the papers that comprise this special issue. Dynamic movie sequences offer a step between static 2D scenes and real 3D environments and are beginning to be used by a number of research groups working in this area. The penultimate paper of this special issue shows one aspect of how dynamic scenes can be employed to further our knowledge of the perceptual processes operating when we view scenes.

How we visually sample and encode information from scenes has been at the heart of eye movement research since Dodge, Judd and Stratton first noted the discontinuous sampling by the eye when viewing patterns and simple line illusions. Stratton (1902, 1906) and Judd (1905a, b) both noted that the oculomotor behaviour as people viewed line illusions did not map clearly onto the perceptual experiences of those illusions – for example, there was no evidence that particular patterns of looking either

promoted or denied the experience of seeing the illusion. A landmark study of scene perception was that of Buswell (1935) who recorded eye movement behaviour as people viewed paintings and photographs of natural scenes. Buswell's work revealed a number of key insights that still underlie questions in contemporary scene perception research.

First, he noted that, consistently across observers, there are regions of the scene that receive little or no inspection by the eyes, and others (which he called 'centers of interest') that are fixated frequently and by most observers. This observation suggests that there may be aspects of the visual information present at the 'centers of interest' that 'attract' the viewers' attention. Whether fixations are attracted to particular visual features has become a prominent question in eye movement research and remains controversial. In recent years the dominant quantitative model of eye movement behaviour when viewing complex scenes has been Itti and Koch's (2000) saliency model. This model proposes that low-level information extracted from the visual scene is a key factor in determining where observers will fixate. Whether models based on visual saliency or conspicuity can offer a good account of where humans fixate has been the focus of a large volume of recent literature. Certainly correlations between where people fixate and the presence of visual features in scenes have been demonstrated (e.g., Itti and Koch, 2000; Parkhurst et al., 2002), but these correlations are weak (e.g., Tatler, 2007; Tatler, Baddeley and Gilchrist, 2005a) and, of course, caution must be exercised when interpreting correlations as causal (Henderson, Brockmole, Castelano & Mack, 2007). In this special issue, **Nyström and Holmqvist** present a new technique for evaluating whether the correlation between features and fixation reflects a role for low-level features in eye guidance or emerges from correlations between higher-level factors and low level image features. They find that if they reduce the visual feature information, or saliency, at semantically informative re-

gions, such as a person's face, in a scene, observers still look at these locations. Their work provides an elegant case for the involvement of low-level image features being very small when viewing semantically-rich scenes.

A second contribution of Buswell was to note that viewing behaviour changes over time when we view a scene: the consistency between observers in where is fixated is far higher in the first few seconds of viewing than it is in the last few seconds. The question of how viewing behaviour changes over time is explored in several of the papers in this special edition. That several of these papers consider how eye movement behaviour changes over time reflects a current change in the focus of scene perception research from treating each fixation as an isolated event (which is computationally appealing and simple) to recognising (as Buswell did) that there may be crucial information about eye movement behaviour that can be gleaned from understanding the sequential nature of fixation selection. **Nyström and Holmqvist** contribute to the ongoing debate about whether the first few fixations are more dominantly guided by image features than later fixations (Parkhurst et al., 2002) or not (Tatler et al., 2005a). Their work suggests that the involvement of features in selecting the first fixation is no different to that in selecting later locations to fixate. **Humphrey and Underwood** consider the similarity of sequences of fixations when viewing and later recognising or imaging the same scenes. These authors show that similar sequences of fixations are produced when viewing a scene as when later recognising the same scene. Similarly, when mentally imaging a scene soon after viewing, or 2 days after viewing the sequence of eye movements is similar. **Humphrey and Underwood's** work offers an interesting return to ideas first explored by Noton and Stark (1971), which have for a long time been controversial. A slightly different approach to the same question of how viewing behaviour changes over time is found in **Pannasch, Helmert, Roth, Herbold and Walter's** contribution to this special issue. These authors explore the possibility that has been previously suggested that there may be at least two modes of viewing a natural scene: a global or ambient mode of looking, where the general layout of the scene is extracted, and a more focal mode of viewing particular aspects of the scene. In particular, these authors explore the previous suggestion by their group that ambient scanning dominates the first few seconds of viewing a scene, whereas focal scanning dominates later viewing (e.g., Unema, Pannasch, Joos, &

Velichkovsky, 2005). Their work systematically explores whether this apparent difference in viewing modes early and late in viewing a scene is robust across a range of viewing situations, including repeated viewing of scenes, the density of objects in a scene, the emotional valence of the scene contents, and the mood of the observer. That the early/late differences in viewing style (quantified by the relationship between saccade amplitude and direction) is stable across all these conditions is used to support the notion that this change from ambient to focal viewing may be a fundamental feature of how we view natural scenes. **Tatler and Vincent** further explore the sequential dependencies between successive saccades and fixations. These authors find support for there being periods of focal scanning, characterised by sequences of small amplitude saccades, interspersed with large relocations to new scene regions. They also find support for the existence of corrective saccades in natural scene viewing: small amplitude saccades that follow (and are in the same direction as) large amplitude saccades, and are preceded by short duration fixations. Their findings also speak to the question of inhibition of return in scene viewing, showing an increased latency before saccades launched in the opposite direction to the previous (therefore returning in the direction from which a saccade had just been launched), but no decrease in the frequency of such return saccades. **Tatler and Vincent's** work clearly demonstrates that we cannot treat saccades and fixations as isolated events in any attempt to understand eye guidance.

While how we inspect a scene is a fundamental question in scene perception research, it is not the only focus of contemporary work in this area. As the early eye movement researchers – Dodge, Judd, Stratton and Buswell – all noted, there is a large discrepancy between the disjointed sampling by the eye and the smooth and complete experience that we have of our visual surroundings. A prominent recent direction of scene perception research has been to consider how information from fixations is retained and integrated into representations and memories of scenes. Interest in this question intensified when it was realised that we are often unaware of large changes that occur in scenes we are viewing, provided the changes coincide with brief interruptions to viewing (e.g., Grimes, 1996; Rensink, O'Regan and Clark, 1997). This phenomenon is known as change blindness and is a striking demonstration that the representations we encode from the visual information we sample may be rather sparse. **Smith and Henderson** extend the wealth of pre-

vious research on change blindness to the relatively under-studied but very familiar medium of film. In film we experience frequent cuts within and between scenes, and these can introduce very large changes in the visual information presented to the viewer. Despite this, we are able to watch and make sense of film exceptionally well and we are often unaware of these edits. **Smith and Henderson** offer a first systematic exploration of the extent to which we are blind to these edits in film and find that cuts often go unnoticed by the viewer even when they are specifically asked to look out for them. Furthermore, different types of cuts are detected with different frequencies: those that adhere to rules laid down by film makers are more often missed than those that violate these editing 'rules'.

The phenomenon of change blindness and **Smith and Henderson's** demonstration of 'edit blindness' both clearly illustrate that visual representation is less veridical and comprehensive than previously thought. However, these results do not preclude the possibility that some information survives and is integrated into long-term memories of objects. Previous authors have shown that object recall may be quite good and that information about objects accumulates over multiple fixations (e.g., Hollingworth & Henderson, 2002; Melcher, 2006; Tatler, Gilchrist & Land, 2005b). **Võ, Schneider and Matthias** offer a first extension of this study of object memory into considerations of individual differences in visual attention capabilities. Vo and Schneider show that the storage capacity of VSTM – assessed using Bundesen's (1990) Theory of Visual Attention parameters – dramatically influences object memory. Those with higher VSTM capacities were better able to distinguish new from previously-seen objects at test than those with lower VSTM capacities. This work demonstrates that individual differences in visual processing abilities must be considered in any comprehensive understanding of how we encode and remember information from natural scenes.

It is clear from the range of papers included and the topics covered that there remain many unanswered questions in how we perceive natural scenes: from how we view scene to what we represent in long term memory. Yet, progress is being made. Certainly, we have refined the questions we need to ask and the techniques that we can use to ask them. Perhaps the biggest remaining challenge in this field, and one that is sadly not represented in this special issue, is to extend the wealth of current understanding and research effort into truly natural envi-

ronments. Our goal must be to understand how vision serves behaviour in natural settings yet the vast majority of research into natural scene viewing uses images displayed on computer monitors. Such simplified experimental settings of course fall well short of the demands and requirements of operating within a real environment and we must seriously ask the question whether the insights we have gained from these laboratory-based settings really tell us how perception operates in natural environments. In the past technological limitations of eye tracking equipment have limited what has been possible realistically to study in real world settings, but this limitation is no longer as strong as it was. Hopefully, over the next few meetings of the European Conference on Eye Movements we will see an increasing shift of natural scene perception research into dynamic real world environments. Only when this shift is made will we be able to evaluate what we have learnt about viewing images of natural scenes. The artificial situation of sudden onsets and offsets of scenes and the physical constraints of the monitor may themselves have a strong influence on how we view scenes (e.g., Tatler, 2007), and if these constraints do not operate in the real world, we may find that our understanding of natural scene perception undergoes some interesting developments in the next few years...

References

- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97, 523-547.
- Buswell, G. T. (1935). *How People Look at Pictures: A Study of the Psychology of Perception in Art*. Chicago: University of Chicago Press.
- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Atkins (Ed.), *Perception: Vancouver Studies in Cognitive Science* (Vol. 2, pp. 89-110). New York: Oxford University Press.
- Henderson, J. M., Brockmole, J. R., Castelano, M. S., & Mack, M. L. (2007). Visual saliency does not account for eye movements during search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Oxford: Elsevier.

- Hollingworth, A., & Henderson, J. M. (2002). Accurate Visual Memory for Previously Attended Objects in Natural Scenes. *Journal of Experimental Psychology-Human Perception and Performance*, 28(1), 113-136.
- Humphrey, K. & Underwood, G. (2008). Fixation sequences in imagery and in recognition during the processing of pictures of real-world scenes. *Journal of Eye Movement Research*, 2(2): 3, 1-15.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12), 1489-1506.
- Judd, C. H. (1905). The Müller-Lyer illusion. *Psychological Monographs*, 7(1), 55-81.
- Judd, C. H. (1905). Movement and consciousness. *Psychological Monographs*, 7(1), 199-226.
- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision*, 6, 8-17.
- Noton, D., & Stark, L. (1971). Scanpaths in Eye Movements During Pattern Perception. *Science*, 171(3968), 308-&.
- Nyström, M. & Homlqvist, K. (2008) Semantic Override of Low-level Features in Image Viewing – Both Initially and Overall. *Journal of Eye Movement Research*, 2(2): 2, 1-11.
- Pannasch, S., Helmert, J. R., Roth, K., Herbold, A-K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2): 4, 1-19.
- Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1), 107-123.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368-373.
- Smith, T. J. & Henderson, J. M. (2008). Edit blindness: the relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, 2(2): 6, 1-17.
- Stratton, G. M. (1902). Eye-movements and the aesthetics of visual form. *Philosophische Studien*, 20, 336-359.
- Stratton, G. M. (1906). Symmetry, linear illusions, and the movements of the eye. *Psychological Review*, 13, 82-96.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14), 4, 1-17.
- Tatler, B. W. & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2): 5, 1-18.
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005a). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, 45(5), 643-659.
- Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005b). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, 58(5), 931-960.
- Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473-494.
- Võ, M. L-H., Schneider, W. X., & Matthias, E. (2008) Transsaccadic Scene Memory Revisited: A ,Theory of Visual Attention (TVA)' Based Approach to Recognition memory and Confidence for Objects in Naturalistic Scenes. *Journal of Eye Movement Research*, 2(2): 7, 1-13.