

EVIDENCE FOR COVERT ATTENTION SWITCHING FROM EYE-MOVEMENTS.  
REPLY TO COMMENTARIES ON LIECHTY ET AL., 2003<sup>1</sup>

MICHEL WEDEL

UNIVERSITY OF MICHIGAN

RIK PIETERS

UNIVERSITY OF TILBURG

JOHN LIECHTY

PENNSYLVANIA STATE UNIVERSITY

We argue that our research objectives in Liechty, Pieters, and Wedel (2003) are to provide generalizable insights into covert visual attention to complex, multimodal stimuli in their natural context, through inverse inference from eye-movement data. We discuss the most important issues raised by Feng (2003) and Reichle and Nelson (2003), in particular the task definition, inclusion of ad features, object-based versus space-based attention and the evidence for the where and what streams.

Key words: covert visual attention, space based attention, object based attention, where and what stream, scan paths, hidden Markov model, print advertising.

1. Inverse Inference on Visual Attention

Our research objectives in Liechty, Pieters, and Wedel (2003) are to provide generalizable insights into covert visual attention to complex, multimodal stimuli in their natural context, through inverse inference from eye-movement data. We find that patterns of saccades are indicative of such covert attention states, and our study yields evidence that people exhibit a sequence of rapid switches between local and global covert states of visual attention. During a typical stimulus exposure, there tend to be about three of those switches on average, while subjects almost always start in the local and typically end in the global state; where the local state lasts around 1.1 seconds and the global state lasts around 0.2 seconds. Our study was inspired by the belief that we share with Kingstone, Smilek, Ristic, Kelland Friesen, and Eastwood (2003) that “theories of attention, too often generated from artificial laboratory experiments, have limited validity if attention in the natural world is considered”. This is in keeping with Broadbent’s (1971) view that cognitive theory should aim at solving questions about day-to-day attention of people in real life settings. The complexity of such everyday problems calls for integrative approaches to uncover processes and to produce generalizable findings. Feng (2003) points out that our research relates to, as yet, disparate streams of psychological research each with their different objects, theories, methodologies and data. In particular, it is located at the crossroads of vision research, attention research, and eye-tracking research.

The case for our research approach was presented elegantly in the first paragraphs of Feng (2003). Our goal is to understand real-life attention processes to print advertisements. Narrowing down the problem to one that lends itself to experimental, controlled testing in the lab is likely to render the conditions too contrived to inform us about the real-life process. As a direct

<sup>1</sup> Author names are in reverse alphabetical order.

The authors would like to thank Arnaud De Bruyn for his help in preparing the second figure.

Requests for reprints should be sent to Michel Wedel, Dwight F. Benton Professor of Marketing, University of Michigan Business School, 701 Tappan Street, Ann Arbor MI 48109-1234. E-Mail: wedel@umich.edu

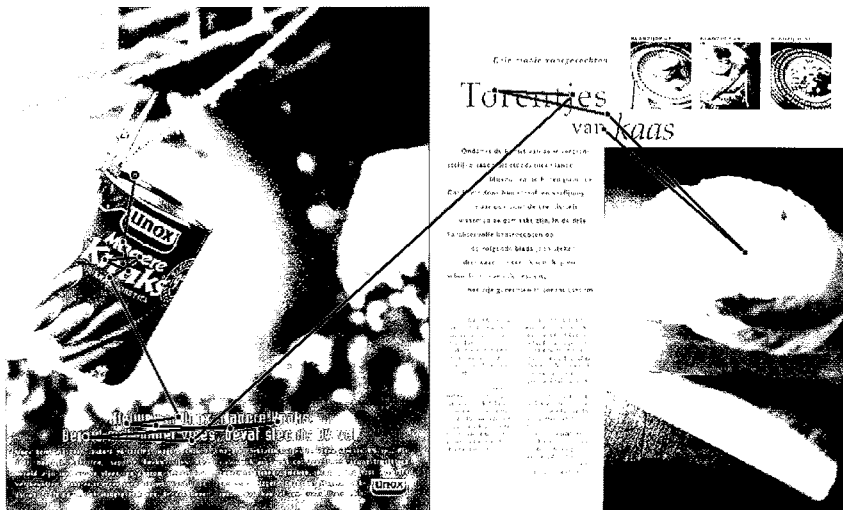


FIGURE 1.  
Scan-path of a subject on an ad (left) and editorial page (right).

consequence of using real stimuli under natural conditions, statistical approaches are needed to provide insights in the underlying processes from data post-hoc. We agree with Feng (2003) that even richer models may allow us to consider other and perhaps more challenging problems, and we set out to develop and test these types of models in future research.

Figure 1 in Reichle and Nelson (2003) illustrates the potential problems of generalizing inference from controlled settings to other domains. They present a *hypothetical* sequence of fixations that, as one may tacitly assume, is derived from their extensive experience with participants in reading tasks. Compare this figure to our Figure 1. This figure shows the *observed* scan-paths of an individual for an ad. The difference between the two Figures illustrates that generalization from controlled reading tasks to real life settings, where the stimulus contains both textual and pictorial components, is extremely hazardous (see also Henderson & Hollingworth, 1999). While the E-Z Reader may help in understanding when and where the eyes move when reading text, we believe it is not suited to describe eye-movement patterns to complex, multimodal stimuli, and even less so to obtain insights into the covert visual attention processes (see De Graef & Gormeys, 2003). The mere accuracy with which the model reproduces the marginal distribution of saccade lengths does not testify to its accuracy in describing covert attention (Roberts & Pashler, 2000). In calibrating their model to our data, Reichle and Nelson (2004) specify different probabilities for saccades within the same element, to neighboring elements, and to nonneighboring elements. By making those modifications—without which the EZ-reader model would give an ill account even of the marginal saccade frequencies—they account for differences in local and global eye-movements. We refer to Feng (2004) for a discussion of the benefits that can be gained from using our approach and discuss the most important issues raised in the commentaries next.

## 2. Definition of Exploration Tasks

Our study involved an exploration task—details of which are in the original paper, very similar to the tasks in Yarbus (1967) and Zangemeister, Sherman, and Stark (1995). In previous studies using such exploration tasks, eye-movements showed a significant relationship with subsequent (indirect) memory (Wedel & Pieters, 2000). We argue that while exploration is a common task for people when they see (e.g., ads) in the real world, research has not extensively studied it.

We were primarily interested in bottom-up processes. Yantis (1998) argues that if no explicit goal instruction is provided in the task, people naturally assume a goal in which they are most sensitive to salient bottom-up features. When people explore scenes, attention is more strongly driven by perceptual than by conceptual features (Pashler & Harris, 2001). Our exploration task should thus induce visual exploration with minimal top-down control (Folk, Remington & Johnston, 1992). That people scanned the ads “very casually” (Reichle & Nelson, 2003) was thus precisely what we intended to accomplish. We intended to create a realistic setting, but to induce minimal top-down influence or demand effects (as may occur when asking people to imagine hypothetical situations). We believe that the exploration task with minimal goal-directed task instructions accomplishes that, while producing qualitatively different eye-movement patterns from those during controlled reading of text (Figure 1).

### 3. Incorporation of Features

We agree with Feng (2003) that it would be a logical next step to relate top-down and bottom-up features to the observed eye-movements and the covert attention states, to improve our understanding of shifts in covert attention and help in improving ad design. We have presently not included such features because our focus was on the covert attention states during ad exploration, rather than on specific ad features that may influence the attentional process of prime interest. In view of the many possibilities, including (basic) perceptual features, textual objects, pictorial objects and higher-level conceptual features, it is not clear how to code the ads. And, each of those features could be incorporated in the intensity of the hidden states, or in the transition matrices or parameters for the processes that govern the moves, or both. Different types of ad content (perceptual/conceptual) may affect the attention process quite differently, and more theory is needed to help guide the selection of the relevant features and the specification of the model. Feng (2003) makes the useful suggestion to use informativeness of elements of the ads as a measure of content. We think this is important, but it is as yet not clear how informativeness should be operationalized (Henderson & Hollingworth, 1999). The “informativeness” is defined in terms of the reduction of uncertainty with respect to achieving certain goals and thus depends on subjects’ goals and the task (Mallot, 2000). For example, Wedel and Pieters (2000) showed that the brand element contained most information for subsequent indirect memory tasks

Given that the ad elements may contain a variety of different types of information, we chose to impose a spatial grid to describe the eye-movements of participants across the ads without recourse to their content. Each of the 48 cells of the grid corresponds to roughly an eight degrees visual angle, which is similar to the midrange visual angle used by Zangemeister et al. (1995) for discrimination between local and global saccades. It is a mistake however, to assume that the definition of the cell size *directly* affects our definition of local and global covert visual attention (Reichle & Nelson, 2004). We do not assume a saccade to a non-neighbor grid to be identical to a global attention shift, but use it in a probabilistic sense as an indicator of a covert attention shift. This is referred to as inverse inference by Feng (2003). Also, we do not treat all cells the same, but specify separate parameters for all transitions between all cells of the grid for every ad. We refrained from presenting this parameter set because we were not interested in overt attention, or the effect of perceptual and conceptual features, but wanted to focus instead on the covert attention process. We agree that inferences may be sensitive to the particular grid that is chosen, and future research would need to determine how robust our results are.

### 4. Space-Based Versus Object-Based Attention

That we aim to identify local and global states of covert attention does not imply that we adhere to the view of space-based attention. Our model does give a spatio-temporal account of observable eye-movements, for which there is ample neurological evidence. Saccade control largely resides in the superior colliculus (SC). The SC contains a topological map of the visual

field and the spatial location of a cell in the SC accurately corresponds to the location of its receptive field in the visual field (Schiller, 1998). The SC is connected to occipital visual areas, specialized in the processing of color, form and motion, amongst others. Those areas display a similar topographical organization (Farah, 2000). This makes it likely that in the absence of strong top-down inhibition, saccades are directed at spatial locations in the visual field that provide neural stimulation from their basic perceptual features. Thus, regardless of whether covert visual attention is space- or object-based, eye-movements are necessarily spatial, driven by a topological motor map of the visual field in the SC. Our model reflects this.

Contrary to Reichle and Nelson's (2004) interpretation of Farah's work, Farah (2000, pp. 177–178) herself concludes that "attention is fundamentally spatial", and that we cannot attend to a stimulus without attending to its location, but that attentional selection may occur from object representations. Without specific (search) goals or expectations on objects in the ad, the perceptual features of the ad may strongly guide attention bottom-up, inducing a dominant spatial aspect in visual attention that does not require a mental representation of objects. In addition, what constitutes an object is often ambiguous for naturally occurring stimuli (Logan, 1996). The multimode, overlapping and hierarchically grouped objects that occur in real-life contexts are in our view less likely to invite pure object-based attention in exploration tasks, and this is precisely what makes it so exciting to examine them.

Whereas overt visual attention is intrinsically spatio-temporal, we believe that covert attention can be simultaneously space-based and object-based (Logan, 1996), depending on top-down and bottom-up factors. The nature and location of objects may interact with the intensity of space-based and object-based attention processes. Our model only supports the conclusion that at a particular point in time, a person explores the stimulus in the global or the local covert attention state. It has nothing to say about whether the focus of attention is object-based or not. That is: we do not actually infer what participants are attending to. Our model does not even preclude covert attention to two objects that are separated spatially. But, we agree that this needs further study. In addition, the concepts of object-based and space-based attention seem to be correlated with those of the focus of visual attention on "what" and "where", as discussed below.

### 5. The "Where" and "What" Streams

We think our model provides evidence for the existence of distinct local and global scan-paths—given the data, the two-state model is over 200 times more likely than the single-state model that for example underlies the EZ-reader model. We might have considered a three-state model, which we did not do given the strong prior theoretical evidence for two-states. It is clear that more research is needed to corroborate our conclusion that the local and global scan-paths are indicative of activity in respectively the "what" and "where" streams, but in view of the available evidence reconfirming the existence of these two streams was not our foremost objective.

Several early studies in cognitive psychology revealed evidence for independent representation of object identity and location (Eriksen & Rohrbaugh, 1970; Styles & Allport, 1986). Additionally, in the neurological literature there is evidence that information from the visual areas in the occipital lobe separates into two visual streams, one stream receiving input mainly from neurons that have a receptive field located primarily in the periphery, sending its output mainly to the posterior portion of the parietal lobe, while the other receive input mainly from neurons that have receptive fields located near the fovea and sending output to the inferior portion of the temporal lobe (Farah, 2000). Baizer, Ungerleider, and Desimone (1991) presents impressive evidence for the segregation of these two streams and provides strong support for Ungerleider and Mishkin's (1982) original proposal of the where and what streams. This proposal was based on observations that damage to the parietal lobe causes difficulties in visual orientation, while damage to the temporal lobe leads to impaired recognition of objects.

We believe that our findings have ramifications for theories about the "what" and "where" streams of attention processing. Our model identifies from the overt saccade pattern, the activity



top-down processing and the effect of processing goals need to be investigated and incorporated into our model (Farah, 2000; Hollingworth & Henderson, 1999). All of these are interesting challenges, some of which we intend to address in the near future. We are very grateful for the insightful reactions by Reichle and Nelson (2003) and Feng (2003), which have helped us to clarify the motivations for our approach, and reveal the potential of using eye-movements for inverse inference on covert visual attention processes.

#### References

- Baizer, J.S., Ungerleider, L.G., & Desimone, R. (1991). Organization of visual input to the inferior temporal and posterior parietal cortex in Macaques. *Journal of Neuroscience*, *11*, 168–190.
- Broadbent, D.E. (1971). *Decision and stress*. London, U.K.: Academic Press.
- Broadbent, D.E. (1982). Task combination and selective intake of information. *Acta Psychologica*, *50*, 253–290.
- De Graef, P., & F. Germeyns (in press). Reading the scene: Application of E-Z Reader to object and scene perception. *Behavioral and Brain Sciences*.
- Eriksen, C.W., & Rohrbaugh, J.W. (1970). Some factors determining the efficiency of selective attention. *American Journal of Psychology*, *83*, 330–343.
- Farah, M.J. (2000). *The cognitive neuroscience of vision*, Oxford, UK: Blackwell Publishers.
- Feng, G. (2003). From eye movement to cognition: Toward a general framework of inference. *Psychometrika*, *68*, 551–556.
- Folk, C.L., Remington, R.W., & Johnston, J.C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030–1044.
- Henderson, J.M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.
- Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C., & Eastwood, J.D. (in press). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Research*.
- Liechty, J., Pieters, R., & Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika*, *68*, 519–541.
- Logan, G.D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603–649.
- Mallot, H.A. (2000). *Computational vision*. Cambridge, MA: The MIT Press.
- Meyer, D.E., & Kieras, D.E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, *12*, 391–438.
- Pashler, H., & Harris, C.R. (2001). Spontaneous allocation of visual attention: Dominant role of uniqueness. *Psychonomic Bulletin & Review*, *8*(4), 747–752.
- Reichle, E., & Nelson, J.R. (2003). Local vs. global visual attention: Are two states necessary?. *Psychometrika*, *68*, 543–549.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment of theory testing. *Psychological Review*, *107*(2), 358–367.
- Schiller, P. (1998). The neural control of visually guided eye-movements. In J.E. Richards (Ed.), *Cognitive neuroscience of attention* (pp. 3–50). London, U.K.: Lawrence Erlbaum.
- Styles, E.A., & Allport, D.A. (1986). Perceptual integration of identity, location and colour. *Psychological Research*, *48*, 189–200.
- Ungerleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, R.J.W. Mansfield, & M.S. Goodale (Eds.), *The analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Wedel, M., & Pieters, F.G.M. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science* *19*, 297–312.
- Yantis, S. (1998). Control of visual attention. In H. Pashler (Ed.), *Attention* (pp. 223–256). East Sussex, U.K.: Psychology Press.
- Yarbus, A.L. (1967). *Eye movements and vision*. New York, NY: Plenum Press.
- Zangemeister, W.H., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia* *33*, 1009–1025.

*Manuscript received 27 OCT 2003*

*Final version received 18 NOV 2003*