



## Editorial

## Visual search and visual world: Interactions among visual attention, language, and working memory (introduction to the special issue)

The world is full of visual stimuli and so observers need selection mechanisms to attend to the things that are relevant given the situation at hand and ignore other things. As opposed to other animals, humans can also talk about the things they see around them, and look at (or look for) objects that have been mentioned to them using natural language. Interactions between vision, attention, and higher cognitive functions like language have received considerable attention in two subdomains of cognitive psychology. Researchers primarily interested in visual attention have often used the *visual search* task (Wolfe, 1998), in which subjects have to find a certain target object that occurs along with a number of distractor objects in a visual array. It is often assumed that a representation of the target object has to be held in working memory. Researchers primarily interested in language processing have often used the *visual world* paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Here, subjects listen to a spoken utterance and simultaneously view a number of objects while eye-movements are monitored. The spoken utterance usually refers or is related to one or more objects in the scene and the question is whether and when people look at these objects.

In both paradigms, the participants view a set of objects while their attention is steered by information in memory. The type of displays used is often very similar (e.g., compare Fig. 1 in Huettig, Rommers, & Meyer, 2011-this issue and Fig. 1 in Iordanescu, Grabowecy and Suzuki, 2011-this issue). Both types of paradigms have counterparts in the real world (e.g., looking for someone in a crowd or attending to an object that somebody just mentioned). One might thus expect that visual search researchers have a great deal to tell their visual world colleagues and vice versa. On the contrary: we have noticed that there is hardly any interaction between these fields. While it is starting to become clear that visual attention can be driven from the contents of working memory, visual search researchers so far have not addressed the question of how a representation in working memory is constructed and maintained: the “search template” is taken to be a given. And while it is also becoming clear that observers look at objects that (partially) correspond to the speech they hear, visual world researchers do not tend to ask how and why language affects visual attention: the speech–eye link is taken for granted.

There thus seems to be “theoretical no man’s land” with respect to how language, memory, vision, and attention interact (Huettig, Olivers, & Hartsuiker, 2011-this issue). A reason for this state of affairs may be related to the widely held view that vision and language are separate, informationally encapsulated modules (Anderson, Chiu, Huette, & Spivey, 2011-this issue) so that there is little reason for vision and language scientists to read each other’s work and bother about each other’s findings, let alone collaborate with each other. This special issue aims to lay the first stones for a bridge across this divide. We note that a special issue in the *Journal of Memory and Language* (Ferreira &

Tanenhaus, 2007) already addressed language–vision interactions, as do several of the papers here. However, all of the papers in the JML issue were written by language scientists. The current issue takes the study of vision and higher cognition interactions a step further, by bringing together papers from both language and vision scientists.

Specifically, Huettig, Olivers, and Hartsuiker (2011-this issue) set the stage by introducing the visual search and visual world paradigms, pointing out some of the similarities and differences. They sketch a theory of interactions between language and visual attention that is based on findings from both traditions. They claim that working memory is the nexus that binds visual, linguistic, and spatial information. Thus, when observers view a typical visual world display, they bind information about the objects’ visual form, names, and meanings to spatial indices. When the observer then processes a spoken utterance, a match at any of these levels will activate the corresponding spatial location and drive visual attention to that location. This account is compatible with the finding (reported in this issue and elsewhere) that working memory can control visual attention in the visual search task and that such control is based upon relatively abstract matches between memory content and search item.

Then, six papers focus on the visual world paradigm. Huettig, Rommers, and Meyer (2011-this issue) review the literature that has used the visual world paradigm as a tool to investigate the mental architecture of language processing. It is argued that this paradigm has many advantages over more traditional research tools, which often require subjects to make metalinguistic responses (e.g. the lexical decision and syllable monitoring tasks) or even guesses (e.g. the gating task). They stress however that it is important to keep in mind that fixation behavior in the visual world paradigm does not reflect linguistic processing directly but the *outcome* of a process which integrates visual processing with linguistic processing, so that the findings do not necessarily generalize to linguistic processing in the absence of a relevant visual context.

The next two papers further discuss the linkage between language processing and eye-movements. Salverda, Brown, and Tanenhaus ask to what extent linguistic effects on visual attention are driven automatically (by activating a shared conceptual stratum) and to what extent such influences are modulated by top-down influences. They base their argument on visual perception work that used relatively naturalistic displays; such studies have argued for the importance of top-down guidance of attention by the observer’s structure of goals and subgoals. The claim is that the linkage between language and visual attention in the visual world paradigm is similarly modulated by top-down effects such as the task structure. Anderson, Chiu, Huette, and Spivey (2011-this issue) also address the nature of the language–vision link. They argue that not only does language processing affect visual perception, but visual perception affects language processing as well. In their view, language and vision

are not discrete modules, in the sense of Fodor (1983), and both modalities are characterized by highly continuous processing that is best studied with dense-sampling approaches like eye-tracking.

The next two papers ask concrete questions about two important aspects of the language–vision link. Altmann (2011–this issue) asks whether the earliest moment at which a linguistic processing event can influence eye-movements is really 200–300 ms as is typically assumed. To disentangle saccadic eye-movements towards a referent (e.g., “man”) in response to linguistic processing of the word *man* from coincidental saccades that happened to occur while *man* acoustically unfolds, Altmann (2011–this issue) introduces a new method of analysis and applies this to the data of two earlier studies. The analyses lead to the conclusion that linguistic processing can already influence the location of the next saccade after 100 ms. Although this study cannot fully account for the discrepancy with previous estimates, a fascinating possibility is that these times are not only determined by the decision to move the eyes to a particular referent when incoming linguistic information is consistent with that referent’s name, but also by the decision to quickly *cancel* a planned saccade when incoming information is not consistent with that landing site. Barr, Gann, and Pierce (2011–this issue) note that the link between language and vision is further complicated by the fact that eye-movements to objects in the visual world can be driven by the observer’s expectations as based on the non-linguistic preprocessing of the scene. Although in some studies such predictive processing is actually the point (e.g., Kamide, Altmann, & Haywood, 2003), in other studies such “anticipatory baseline effects” might lead to spurious conclusions. The paper tests a number of ways out for this problem and provides a novel statistical solution.

In the last visual world paper, Andersson, Ferreira, and Henderson (2011) ask whether the standard effects of fixating the objects that are mentioned in concurrent speech tasks (Cooper, 1974; Tanenhaus et al., 1995) generalize to situations that more closely approximate the real visual world. These authors therefore report an experiment with speech that is more complex than most previous studies (i.e., mini-monologues containing several referring expressions, sometimes produced at a fast speech rate) and with visual displays that are more complex than typical visual world studies (i.e., photographs of realistic scenes, containing numerous objects in a cluttered arrangement). Even in conditions with fast speech and a fast succession of object mentions, people look at what they hear, demonstrating the robustness of the vision–language link.

Next, four papers focused on the control of visual attention from working memory in the visual search task. Carlisle and Woodman (2011) ask whether this type of control is driven only by an automatic component, or also by top-down influences (i.e., similar to the question posed by Salverda et al. in the context of visual world). To do so, they asked subjects to memorize a cue object and then search for a target (e.g., a square with a gap in a particular location). Trials were neutral (memory object did not re-appear in the search display), valid (memory object appeared as the target in the search display), or invalid (memory object appeared, but not as the target). The crucial manipulation was the proportion of valid trials. There was an effect of cue validity even with a very low proportion of valid trials, but these effects increased with a larger proportion of valid trials. This finding suggests a role for both an automatic and a strategic component in the memory–guidance of visual attention.

Soto, Mannan, Malhotra, Rzeskiewicz, and Humphreys (2011–this issue) demonstrate that attentional control from working memory can ameliorate the attentional deficit in patients with visual extinction; although such patients can perceive stimuli in the contralesional (“bad”) visual field when these are presented by themselves, they often fail to do so in the presence of another stimulus in the ipsilesional (“good”) visual field. Soto and Humphreys (2006) previously demonstrated that the maintenance of a visual cue (e.g., a red color patch) in working memory facilitated visual search for

contralesional targets that match this cue in color; the current study shows that this is also true when the memory cue is presented in verbal form (e.g., the word “red”). Similarly, these authors show that such cuing helps to deal with the problem of failing to disengage attention from the ipsilesional field. Most importantly, the results are evidence that the match between memory and extinguished visual objects can be at a rather abstract level (i.e., the semantic representation of a written word).

Geyers, Gokce, and Müller (2011–this issue) aim to further elucidate the nature of memory representations driving visual attention. In their study subjects also held an item in working memory and did a visual search task while memory was loaded. Crucially, in one condition the memory item had the same overall shape as the *configuration of items* in the search display (i.e., triangular). In the visual search task, the target could either appear at the location previously occupied by a target, by a distractor, or a previously unoccupied location. Search task reaction times were shorter when the target item on trial  $n$  occurred in the same position as trial  $n - 1$  (compared to an unoccupied position) and longer when it occurred in the same position as a distractor on trial  $n - 1$ . Importantly, this positional inhibition was boosted when the shape of the memory item was identical to that of the configuration of the search items (i.e., a triangle). These results suggest that observers in a visual search task learn configuration information and store it in spatial memory, and that such spatial memories can guide visual attention.

One might argue that in Geyers, Gokce, and Müller (2011–this issue) observers learned something about the task itself. In contrast, Olivers demonstrates effects of long-term memories that have already been established. Olivers’ observers searched a display of well-known traffic signs for a specific target sign on the basis of a verbal instruction (e.g., “the yield sign”). The target sign and all irrelevant filler signs were always in gray, but there was also a colored distractor sign. The distractor’s color could either be related or unrelated to the color that the target sign *would have* in the real world (but not in the task itself, as observers were aware – i.e., in the experiment, the target sign was always gray, but in the real world it is red-rimmed). Although color was thus task-irrelevant, color-relatedness slowed down visual search. Thus, when observers construct a target template for search, they do not only access information that is strictly task-relevant, but also further visual information in long-term memory that is associated with the object. Such effects obviously have similarities with the effects of semantic long-term memory shown in the visual world task, so that for instance people tend to fixate a trumpet when hearing the word piano (Huettig & Altmann, 2005; see Huettig, Rommers et al., 2011–this issue).

The four papers mentioned above address the control of attention from the contents of working memory. Theeuwes, Kramer, and Irwin (in press) turn this around and address the control of working memory from attention. Previous studies have shown the importance of attention for storing information in visual working memory; this study shows that attention is also used for *retrieving* information from working memory. Thus, when observers were retrieving a piece of information (whether a given color had appeared in a briefly flashed display with four colors) they were faster to respond in a secondary dot-probe task when the dot appeared in the location at which the correct color stimulus *used to be*. It thus seems that observers spontaneously attend to locations in space that used to contain task-relevant information. That finding has obvious resemblances with the blank-screen variant of the visual world paradigm, where observers attend to a location on a blank screen where a referent mentioned in speech used to be (see Huettig, Rommer, et al., 2011–this issue). It also nicely fits the ideas in Huettig, Olivers et al., (2011–this issue), according to which observers make bindings in working memory between a spatial location with visual and linguistic information, so that the binding persists even when the stimulus has disappeared.

Finally, Iordanescu, Grabowecky and Suzuki (2011-this issue) address the multimodal character of visual search in the real world. They demonstrate that searching for a given object (e.g., a cat) is speeded up when the object name is simultaneously presented acoustically (i.e., consistent with visual world studies) but also when the characteristic sound of the object (meowing) is played. Furthermore, when the objects in the search matrix are replaced by their written names, there is still facilitation from the spoken name but no longer from the characteristic sound, a finding the authors interpret as the result of the lack of an association between written words and the sounds the referents of these words typically make. This study thus puts constraints on what information observers bind to a location in visual search/world type tasks: information that is not directly associated seems to be not bound. It is interesting to note that written-word studies with the visual world paradigm find strong phonological (looking at the written word beetle when hearing the word beaker) but no semantic (looking at another piece of kitchen utensil) or visual (looking at another cylindrical object) effects (Huettig & McQueen, 2007). Finally, Iordanescu et al.'s (2011-this issue) final experiment shows some promise that auditory facilitation of visual search might actually be helpful in relevant real-world situations, such as searching for guns among luggage at an airport.

Although the field may still be a long way off from understanding the complex relations between vision, attention, working memory, and language, the papers reported in this special issue help to populate the theoretical no-man's land. In doing so, it seems that the different research traditions have much more in common than mere similarities in the visual displays they use. In fact, many of the issues are similar, including the question of whether visual attention is modulated by strategic factors and task factors (e.g., Carlisle & Woodman, 2011; Salverda, Brown, & Tanenhaus, 2011-this issue), and the question of whether and how attention is affected by associations in long-term memory (e.g., Huettig, Rommers, et al., 2011-this issue; Iordanescu, Grabowecky & Suzuki, 2011-this issue; Olivers, 2011-this issue). As editors of this special issue, we think we have achieved our goal of collecting a number of very interesting papers that help us to build the bridge between the two fields. What we now hope is that researchers from both research traditions will be inspired by all of these papers, especially the ones from the other tradition.

## References

- Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eye to. *Acta Psychologica*, 137, 190–200 (this issue).
- Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, 137, 181–189 (this issue).
- Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I See What You're Saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, 137, 208–216.
- Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, 137, 201–207 (this issue).
- Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta Psychologica*, 137, 217–225.

- Cooper, R. M. (1974). Control of eye fixation by meaning of spoken language: New methodology for real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Geyers, T., Gokce, A., & Müller, H. J. (2011). Reinforcement of inhibitory positional priming by spatial working memory contents. *Acta Psychologica*, 137, 235–242 (this issue).
- Ferreira, F., & Tanenhaus, M. K. (2007). Introduction to the special issue on language–vision interactions. *Journal of Memory and Language*, 57, 455–459.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96, B23–B32.
- Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57, 460–482.
- Huettig, F., Olivers, N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, 137, 138–150 (this issue).
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137, 151–171 (this issue).
- Iordanescu, L., Grabowecky, M., & Suzuki, S. (2011). Object-based auditory facilitation of visual search for pictures and words with frequent and rare targets. *Acta Psychologica*, 137, 252–259 (this issue).
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Olivers, C. N. L. (2011). Long-term visual associations affect attentional guidance. *Acta Psychologica*, 137, 243–247 (this issue).
- Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, 137, 172–180 (this issue).
- Soto, D., & Humphreys, G. W. (2006). Seeing the content of the mind: Enhanced awareness through working memory in patients with visual extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 4789–4792.
- Soto, D., Mannan, S. K., Malhotra, P., Rzeskiewicz, A., & Humphreys, G. W. (2011). Distinguishing non-spatial from spatial biases in visual selection: Neuropsychological evidence. *Acta Psychologica*, 137, 226–234 (this issue).
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9, 33–39.

Robert J. Hartsuiker

Department of Experimental Psychology, Ghent University,  
Ghent, Belgium

Corresponding author. Department of Experimental Psychology,  
Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium.

Tel.: +32 9 264 6436; fax: +32 9 264 6496.

E-mail address: Robert.Hartsuiker@ugent.be.

Falk Huettig

Max Planck Institute for Psycholinguistics, Nijmegen,  
The Netherlands and Donders Institute for Brain,  
Cognition, and Behaviour, Radboud University,  
Nijmegen, The Netherlands

Christian N.L. Olivers

VU University Amsterdam, Amsterdam, The Netherlands

7 January 2011