Spoken language and the decision to move the eyes - To what extent are language-mediated eye movements automatic?

Ramesh K. Mishra[a], Christian N. L. Olivers[b] & Falk Huettig[c,d],

[a]Centre of Behavioral and Cognitive Sciences, University of Allahabad, India
[b] Cognitive Psychology, VU University Amsterdam, The Netherlands
[c]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[d]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands

---- in press, *Progress in Brain Research*, Vol. 202, Decision Making: Neural and Behavioural Approaches ---

Correspondence:
Ramesh K. Mishra
Centre of Behavioral and Cognitive Sciences (CBCS)
University of Allahabad,
Allahabad 211002
UP (India)
E-mail: rkmishra@Cbcs.ac.in
phone: +91-532-2460738

Abstract

Recent eye-tracking research has revealed that spoken language can guide eye gaze very rapidly (and closely time-locked to the unfolding speech) towards referents in the visual world. We discuss whether, and to what extent, such language-mediated eye movements are automatic rather than subject to conscious and controlled decision-making. We consider whether language-mediated eye movements adhere to four main criteria of automatic behavior, namely whether they are fast and efficient, unintentional, unconscious, and overlearned (i.e. arrived at through extensive practice). Current evidence indicates that language-driven oculomotor behavior is fast but not necessarily always efficient. It seems largely unintentional though there is also some evidence that participants can actively use the information in working memory to avoid distraction in search. Language-mediated eye movements appear to be for the most part unconscious and have all the hallmarks of an overlearned behavior. These data are suggestive of automatic mechanisms linking language to potentially referred-to visual objects, but more comprehensive and rigorous testing of this hypothesis is needed.

Key words: attention, automaticity, control, eye-movements, decision -making

Introduction

A little observation of our daily eye movements will probably lead to the conclusion that they are often deliberately planned and preceded by a conscious decision. Such would be the case for example when you decide to look for the soup ladle in the kitchen drawer. On the other hand, we may also notice that many of our shifts in eye gaze appear to be simple motor acts which are not accompanied by an experience of intention, but unfold because some noteworthy aspect of the environment captures attention. We may also observe that both types of eye-movement (deliberately planned or seemingly unintentional) are often guided by spoken language. For example, we may tell a child to mind the step or to look at the beautiful flower, a visitor may ask for a glass of water, or over the phone we may be given directions to find an unknown location. The question we discuss here is whether, and to what extent, language-mediated shifts in eye gaze are *automatic* to the extent that they circumvent *controlled decision-making*.

When Hockett developed his list of features of human language (Hockett & Altmann, 1968), he included *displacement* – the fact that concepts need not refer to an object that is physically present. Interestingly, recent eye tracking research suggests that when the object of spoken language *is* actually physically present, the cognitive system expresses a strong tendency to refer to it. It does this by orienting its visual sensory apparatus towards the object, thus linking a linguistically activated concept (or type representation) to a specific perceptual instance in the outside world (or token representation). For example, in an early study, Cooper (1974) simultaneously presented participants with spoken fictional stories (e.g., about a safari in Africa) and a visual array containing nine line drawings of concrete objects (e.g., a lion, a zebra, a tree, a camera, etc.). Observers were asked to just listen to the stories. Cooper found that listeners' fixations of the objects were very closely time-locked to the unfolding speech input: whenever a spoken word referred to an object, participants rapidly shifted their overt attention to it, even though this was not required for any task.

Studies within the *visual world paradigm* (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) confirmed that participants tend to very rapidly fixate objects whose names are consistent with the unfolding speech signal. In this paradigm, as in Cooper's study, participants typically see a number of pictures of objects on the screen while listening to spoken language. For example, on hearing the first syllable of 'beaker' during the instruction 'Pick up the beaker' participants rapidly shifted their eye gaze to either a picture of a beaker or a beetle (a phonological competitor because of the same word-initial phonology) in a display of four objects when the unfolding word-initial acoustic information matched both beaker and beetle (Allopenna, Magnuson, & Tanenhaus, 1998). When the acoustic information then started to mismatch beetle (on hearing the second syllable of 'beaker') participants moved their eyes quickly away from the beetle while looks to the beaker continued to increase.

There has been little systematic research with regard to the extent to which these kind of eye movements are automatic or controlled, with most of the visual world studies focusing on linguistic aspects of the unfolding speech, while using eye movements as the dependent variable, and thus mainly as a tool. Yet it is also important to know how symbolic representation systems such as those underlying language interact with perceptual systems that directly ground experience in an outside world (cf. Barsalou, 2008; Glenberg, 2007). Recently researchers have begun to investigate more systematically the interaction of language, vision, attention, and memory during language-mediated eye gaze (see Huettig, Olivers, & Hartsuiker, 2011; Huettig, Mishra, & Olivers, 2012, for review). Are such links automatically forged or do they require conscious and effortful cognitive acts? Here we review which answers may be available in the current literature, and what gaps in knowledge will need to be filled. In doing so we will heavily borrow from the visual attention literature, which has a longer tradition of dealing with issues of automaticity versus control (e.g. Shiffrin & Schneider, 1977).

Automaticity

For our purpose we first need a definition of automatic behavior. As reviewed by Moors and De Houwer (2006), the definition of automaticity has a long history in psychology (e.g. James, 1890; Wundt, 1896/1897; Shiffrin & Schneider, 1977; Logan, 1988), and commonly consists of a list of criteria. Here we adopt the general taxonomy developed by Moors and De Houwer (2006) and distinguish the following features of an automatic process:

1. It is fast and efficient. This means that an automatic program requires relatively few steps and little to no cognitive effort or capacity. It may thus run in parallel to other processes.
2. It runs unintentionally. This means that an automatic process is not subject to control by the observer, nor does it progress because of current behavioral goals. Note that an automatic process may be *triggered* by a current, intentional goal state, but that after

that it runs uncontrollably (e.g. Bargh, 1992; Folk, Remington, and Johnston, 1992). This is often referred to as *contingent* automaticity. For example, a skilled piano player may intentionally decide to play a particular piece, but then may often find his or her fingers unstoppable. Similarly, when you decide to look for the tomatoes in the supermarket, the strawberries may involuntarily trigger your attention because they happen to partly match what you are looking for (e.g. color, Huettig & Altmann, 2011). Similarly, when you are deliberately looking for a picture of a lock, the semantically related key may capture your gaze (Moores et al., 2003).

3. It is unconscious. This means that one may be conscious of the states that started the process (such as the intention to start playing a piano piece), or of the end result (i.e. hearing the tune), but that one has little awareness of the intermediate steps, i.e. how one state led to the other.

Together with Logan (1988), we add to this that automatic behavior is often overlearned, arrived at through extensive practice. In the remainder of the paper, we will review to what extent and under which conditions language-mediated eye movements adhere to these criteria. Note that we do not aim to expel the possibility of volitional decision-making in this behavior, i.e. conscious and controlled language-mediated eye movements. In fact, these are most likely to occur, for example when someone asks you to pass the salt. You may deliberately direct your eye gaze to its most likely location. Here we are interested in the other side of the coin: May the salt shaker automatically capture your attention when you hear “salt”, because of its strong relationship to the linguistic input?

1. Is the interaction between language and eye movements fast and efficient?

Cooper (1974), in his early study, discovered that more than 90% of the fixations to objects mentioned in the prose passages were initiated during the acoustic duration or within 200ms after the offset of the word. Participants in visual world experiments typically start to shift their eyes towards the target objects 200ms after word onset (Allopenna et al., 1998, and numerous other studies, see Huettig, Rommers, & Meyer, 2011, for review). Given that it takes about 200ms to initiate and program a saccadic eye movement (Saslow, 1967; but see Altmann, 2011, for an even shorter estimate) this suggests that participants start to shift their eyes towards the target objects as soon as the acoustic signal from the spoken target is processed. Thus, the speed with which target objects and phonological competitors are fixated during the acoustic unfolding of the words in these studies is consistent with the notion of the automaticity of language-mediated eye movements. Another indication that language-mediated visual attention may be automatic comes from findings that objects, which are only partially related in semantics or visual form to objects that unfolding spoken words refer to, are also very rapidly fixated (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005, 2007, 2011; Yee & Sedivy, 2006).

Visual world research thus far has produced little data concerning the *efficiency* of language-mediated eye gaze as there have been little to no manipulations of set size (i.e. number of objects in the display) or cognitive load (memory set size). Here visual world studies could borrow from the visual search paradigm. In this paradigm it is standard practice to vary set size (the number of objects in the display) to assess how efficient selection of particular visual information is – that is, to what extent processing is parallel, and immune to interference from other visual information present in the scene (see Wolfe, 1998, for a review). For example, if visual world-type interactions between language and eye movements rapidly diminish with increasing set sizes, then this would suggest a strong attentional component: Visual objects need to be (at least partly) attended in order for their phonological or semantic codes to be available for interaction. Another promising approach may be dual task studies. Voluntary eye movements appear to be very susceptible to dual task demands (Jonides, 1981; see also Hunt & Kingston, 2003).

So far studies looking at visual search by category have been scarce, and have yielded mixed results.[1]. Egeth, Jonides, and Wall (1972) found that observers can efficiently search for “any digit” amongst random letters. Other work has revealed categorical effects on search for basic visual properties such as orientation (“steep” versus “shallow”) and color (Wolfe, Friedman-Hill, & O’Connell, 1992; Daouties, Pilling, & Davies, 2006). On a higher level, using pictures comparable to those used in visual world studies, Levin, Takarae, Miner, and Keil (2001) found rather efficient search for animals amongst artifacts and vice versa (with search slopes in the range of 5 to 15 ms/item). The possibility remains however that such search was driven by perceptual rather than conceptual differences between semantic categories

---

[1] There is a large literature on searching natural scenes for specific categories of stimuli, but these typically address a different type of semantic interaction than we are interested here, namely how the semantics induced by the entire visual scene drive search towards specific objects or locations. Here we are concerned with whether the linguistic aspects of multiple independent visual objects in an array are rapidly available, regardless of the scene or visual context

(Krueger, 1984; see Lupyan, 2008 for a nice control of visual similarity, albeit with simple letter stimuli rather than pictures of objects). Indeed other visual studies using semantic target categories have shown rather inefficient search, with search slopes exceeding 30 ms/item (e.g. Farid & Bravo, 2009; Yang & Zelinsky, 2009; Schmidt & Zelinsky, 2009; Wolfe et al, 2004). Wolfe et al. (2004), for example, found that search for a categorically defined target (e.g. "fruit") was quite inefficient, as was search for a more specific category. The only efficient search condition was when an exact picture of the target was provided. Taken together this suggests that the categorical labels for pictures are not readily available, and in that sense do not automatically guide attention.

This does not preclude the possibility that *when* these labels are available for the presented pictures, they will guide attention. Such contingencies may occur when the number of objects in a display is sufficiently small so that they fit in working memory. Huettig, Olivers, and Hartsuiker (2011) have recently proposed that working memory may play a central role during language-mediated eye movements (cf. Knoeferle & Crocker, 2007; Spivey et al., 2004). They argue that phonological, semantic, and general visual representations are automatically linked on the basis of long term memory, but that the specific binding to currently presented stimuli (in space and time) is a function of working memory. As a consequence, the number of bindings available is limited, probably to around four (Trick & Pylyshyn, 1994; Cowan, 2001).

2. Is the interaction between language and eye movements unintentional?

On hearing a spoken word such as 'tomato' participants show an increased likelihood of fixating a ball (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2004, 2007; a visual competitor because tomatoes and balls have a similar global shape), a red sweater (Huettig & Altmann, 2011, a visual competitor because tomatoes are typically red), and a cucumber (a semantic competitor because tomatoes and cucumbers are both vegetables, Huettig & Altmann, 2005; Yee & Sedivy, 2006). These data suggest that even when we know what it is in the concurrent visual environment that is being referred to (participants had plenty of preview of the visual display in these studies), we often cannot help but look at other objects which are only partially related. These findings suggest that during natural conversation there are pressures which drive our attention spuriously to objects which are not the ones the speaker intended.

Crucially, these competitor effects occur with active (direct action, e.g. "Click on the …") and simple look and listen tasks (in which participants are simply asked to listen to the spoken language and not to take their eyes off the visual display) . The data on semantic effects from active and look and listen tasks converge (compare Yee & Sedivy, 2006, with Huettig & Altmann, 2005), as do the data on visual effects (compare Dahan & Tanenhaus, 2005, with Huettig & Altmann, 2007). Thus, even without explicit instruction, participants show an increased likelihood of fixating these related but unintended objects. Hintz and Huettig (submitted) replicated the time-course pattern of phonological, semantic, and visual competitor effects of Huettig and McQueen (Experiment 1, 2007) using an active task with the same visual items and spoken materials (Huettig and McQueen, 2007, had used a look and listen task). The task-independence of these effects support the notion of a substantial automaticity of language-mediated eye movements.

A further set of findings which is consistent with such a view are studies in which participants are shown a blank screen while they hear the unfolding language (Spivey & Geng, 2001; Altmann, 2004). In these studies participants view a visual display of objects or semi-realistic line drawings. They subsequently listen to spoken language while the visual stimuli have been taken away. Importantly, observers tend to re-fixate locations on the empty screen which were previously occupied by objects referred to in the speech, even though this is not required for the task. These re-fixations of empty locations are similarly time-locked to the unfolding speech as are language-mediated eye movements when a visual scene is co-present. This effect is not limited to language-vision interactions but may be a more general property of how episodic memory works, in that memory may use location to retrieve earlier memories (Dell'Acqua, Sessa, Toffanin, Luria, & Jolicoeur, 2009; Eimer & Kiss, 2010; Kuo, Rao, Lepsien, & Nobre, 2009; Theeuwes, Kramer, and Irwin, 2011).

Although the above is suggestive of automatic linking of language to potentially referred-to visual objects, it deserves pointing out that although the visual objects were largely irrelevant to the observers, there was no real incentive to refrain from looking at them. Observers might simply have chosen to look at them because they had nothing else to do or because they suspected a connection. We are aware of only one recent study which investigated the effects of language on visual discrimination when the linguistic-visual match was actually harmful to performance. In their Experiment 2, Salverda and Altmann (2011) presented participants with a display of four objects (e.g. a cat, a sun, a house, and a car). After a few seconds one of the objects turned green (e.g. the cat), and the task was to make an eye movement to the target, in response to the color change. Just prior to the color change, a spoken word was presented which always referred to a distractor object (e.g. the sun). Salverda and Altmann (2011) found that eye movements to the target object were slower when the spoken word referred to a distractor object than when it referred to an object not in the display. Salverda and Altmann interpreted this finding as evidence that the spoken word resulted in an automatic and involuntary shift of covert attention to the distractor object. In their Experiment 3, participants saw two objects and one of the objects shifted up and down. Participants' task was to detect the direction of the target object's shift. Participants were faster when the spoken word referred to the target object than when it referred to an absent object and overall slowest when the word referred to the distractor object. Salverda and Altmann (2011) concluded that task-irrelevant spoken names of objects interfered with performance in the simple visual task even though the spoken words were irrelevant to the task and participants were explicitly instructed to ignore the words, and that thus referents of spoken words capture attention.

The biases towards distractors that match the linguistic content is reminiscent of similar findings in the literature on the interactions between working memory and visual search (see Olivers, 2008, for a review). It has been found that irrelevant distractors can lead to more interference with visual search when they match the current contents of working memory. For example, when observers were asked to remember a red object for a later memory test, red distractors in an otherwise unrelated visual search display attracted more eye movements (Olivers et al., 2006; Soto et al., 2005). Interestingly, this memory-based attentional guidance effect also works for verbally presented material. When Soto and Humphreys (2007) presented a color word like “red” to remember, instead of showing a red object, search was also slowed when a color-matching object was present in the display. Again, because such effects occur despite being harmful to performance this indicates that they are not intentional.

The Salverda and Altmann (2011) and Soto and Humphreys (2007) studies provide evidence that (either auditorily or visually presented) verbal material can lead to an attentional bias. Such effects be may be mediated by semantics. Several studies have found that visual selection is biased towards items that are semantically related to the target item (Moores et al., 2003; Belke, Humphreys, Watson, Meyer, & Telling, 2008; Meyer, Belke, Telling, & Humphreys, 2007; Telling, Kumar, Meyer, and Humphreys, 2010). For example, when observers are asked to look for a picture of a motor bike, a picture of a helmet interferes more with search than an unrelated picture. However, none of these studies provide evidence for or against an automatic component: By nature of the experiments, the semantic competitors were always related to what the participants were actively looking for. This in contrast to the typical visual world setting, in which the spoken utterance often has no relationship to the task.

Despite the evidence suggesting that linguistic material automatically leads to interference from matching visual objects, there is also evidence that such interactions may be under at least some inhibitory control. If language-mediated eye gaze is largely automatic then the question arises why we do not get constantly distracted if even only partially-related objects capture attention. There must be a mechanism which prevents the explosion of spurious shifts in eye gaze during our daily language-vision interactions. Within the visual world paradigm, to our knowledge, there is only one recent study which has reported evidence for inhibition (Hintz and Huettig, submitted). Their participants heard single spoken target words while looking at four objects embedded in visual displays of different complexity and were asked to indicate the presence or absence of the target object. During filler trials the target objects were present, but during experimental trials they were absent and the display contained various competitor objects. For example, given the target word '*beaker*', the display contained a phonological (a beaver, *bever*), a shape (a bobbin, *klos*), a semantic (a fork, *vork*) competitor and an unrelated distractor (an umbrella).When the objects were embedded in semi-realistic line drawings, including human-like cartoon characters which were shown to interact with the objects, inhibition of all competitors was observed at various points in time, i.e. fewer looks to the phonological, semantic, visual shape competitors than to the unrelated distractor. These results provide support for the notion that language-mediated eye movements are under at least some control and that a complete account of language-mediated eye gaze will have to include inhibitory mechanisms. Once again, these findings are similar to visual search findings investigating working memory influences. In contrast to the earlier-mentioned visual search studies showing that memory-matching objects interfere more with search, several studies have found that under some conditions, objects matching the current contents of working memory actually interfere *less* with an ongoing search task than non-matching objects do (Downing & Dodds, 2004; Woodman & Luck, 2007). One possibility is that observers actively use the information in memory to avoid distraction in search. Another possibility is that observers try to shield one task from another, such that when they perform the search task they suppress the memory task, and the associated representations with it. Such strategic mechanisms are expected to take time. Consistent with this, Han and Kim (2009) reported that interference from memory-matching distractors in visual search turns into benefits with increasing delays. Similar mechanisms might account for the inhibition found in the visual world experiments mentioned above.

3. Is the interaction between language and eye movements unconscious?

When asked whether they noticed any relationship between spoken words and visual objects, participants in visual world studies tend to state that they did not note any of the manipulations, or, in a few cases, that they felt they noticed some (e.g., semantic) but not others (e.g., phonological or shape). There are however no data on this other than anecdotal evidence.

There are a few examples from the visual search literature in which experimenters asked observers what they thought they were looking at. For example, Theeuwes, Kramer, Hahn, and Irwin (1998) asked observers to make an eye movement to a colored target. Unpredictably a new but irrelevant object abruptly appeared in the display. The results showed that up to about a third of the initial eye movements went to the abrupt onset rather than to the color target. However, when asked, observers claimed they hardly noticed the onset and certainly never looked at it. Other visual search studies suggest that observers often even have little idea what they have been looking *for*. Observers in a study by Müller, Krummenacher, and Heller (2004) were asked to look for either an orientation-defined or a color-defined target. Target types were randomly mixed in a block of trials. After the last correct trial of the experiment, observers were asked what the target was that they had just responded to – that is whether it was orientation or color-defined. Of the 84 observers, only 37 correctly mentioned the target dimension, and of those, only 11 correctly mentioned the specific feature (e.g. red). Given that in visual search observers may even be unaware of the nature of visual stimuli that directly correspond to their goals, we may suspect that the typically more subtle visual world type interactions also escape observers' awareness – but this obviously remains to be tested.

4. What is the effect of learning on language-mediated eye movements?

Many activities become automatized by learning. Writing for instance is typically preceded by a conscious decision to write but people are usually not aware of their detailed finger movements. Many of us have experienced that when reading a novel our mind had wandered off thinking about something else yet we continued to read for several pages before noticing that we had stopped taking in any meaningful information from the book. Language-mediated eye movements may well be a particularly pertinent example of an overlearned behavior arrived at through extensive practice (cf. Logan, 1988). Some recent research has investigated the effects of years of learning to read and write on spoken language-mediated eye movements.

Using the visual world paradigm, Huettig, Singh, and Mishra (2011) studied language-driven eye movements in high- versus low-literate populations. They replicated previous data (e.g., Allopenna et al., 1998) that high literates (i.e., university students) shift their eye gaze toward phonological competitors as soon as phonological information becomes available and move their eyes away as soon as the acoustic information starts to mismatch. Importantly, however, they observed that low literates only use phonological information when semantic matches between spoken words and visual referents are not present. Moreover, this phonologically-mediated word-object mapping in low literates (if it occurs at all) is (in contrast to high literates) not closely time-locked to the speech input. These data show that low literates' use of phonological information during the mapping of spoken words with visual objects is very different from that of high literates. Learning to read involves many years of practice of phonological decoding. One of the consequences of the strengthening of existing phonological representations in literates appears to be a greater likelihood that these representations are used for other cognitive purposes such as mapping spoken words onto concurrent visual referents. These findings indicate that at least the phonological variety of word-object mapping is an overlearned behavior, modulated by years of practice in reading and writing.

Some studies with small children also point towards some automaticity of language-mediated eye gaze even before literacy acquisition. Johnson et al. (2011) found that 2-year olds who did not yet possess verbal labels for the color attribute that spoken and viewed objects had in common exhibited language-mediated eye movements like those made by older children and adults, i.e. toddlers showed a tendency to look at a red plane when hearing "strawberry". These findings show that that 24-month olds lacking color term knowledge nonetheless recognized the perceptual–conceptual commonality between named and seen objects and that language-mediated visual search need not depend on stored labels for concepts.

Another behavior which is typically modulated by levels of expertise on the task at hand is people's ability to predict and anticipate upcoming events. Eye-tracking in psycholinguistic studies has shown that adult participants can use semantic (Altmann & Kamide, 1999), syntactic (Kamide et al., 2003), and prosodic (Snedeker & Trueswell, 2003) information to anticipate an upcoming visual referent. Similar language-mediated anticipatory eye movements to visual objects have been demonstrated in children aged 10 and 11 years (Nation et al. , 2003), and recently even 24-month olds (Mani & Huettig, in press; see also Borovsky, Fernald, & Elman, in press). Mishra, Singh, and Huettig (2012) found that levels of reading ability attained through formal literacy are related to anticipatory language-mediated eye movements in adults. Indian high literates started to shift their eye gaze to target objects well before target word onset using semantic, associative, and syntactic information from adjectives and particles (preceding the critical noun) for anticipation. In the low literacy group this shift of eye gaze occurred only when the target nouns were heard, more than a second later. Mishra et al. (2012) argued that the steady practice of reading and writing enhances individuals' abilities to generate lexical predictions, abilities that help literates to exploit contextually-relevant predictive information when anticipating which object an interlocutor will refer to next in one's visual environment. A substantial amount of the anticipatory eye movements are likely to rely on automatic associative mechanisms (e.g., the verb eat automatically activating cake; cf. Bar, 2007). Some evidence for anticipatory eye gaze in visual world results however seems to require more 'active' prediction by the participants (e.g, Altmann & Kamide, 2007; Chambers & San Juan, 2008; Knoeferle & Crocker, 2006, 2007), related to, for instance, participants' language production skills (Mani & Huettig, in press; cf. Chang, Dell, & Bock; 2006; Pickering & Garrod, 2007). Both automatic and volitional mechanisms may thus contribute to anticipatory language-mediated eye gaze (cf. Kukona et al., 2011).

Concluding remarks

Recent empirical results suggest that spoken language can guide visual orienting without volitional control. Language-mediated eye movements tend to be fast, unconscious, and largely overlearned. As such they fit most of the criteria of an of an automatic process (cf.

Logan, 1988; Moors & De Houwer, 2006). The evidence for other criteria of automaticity however appears to be weaker. There is currently very little evidence concerning the efficiency of linguistically-driven visual orienting. Furthermore, although language-mediated eye gaze appears to be unintentional to a large degree, there is some evidence that linguistic information, like visual information, can be used in working memory in order to avoid distraction. Thus, some control appears possible. In this respect, it is probably more fruitful to conceptualize findings as reflecting various degrees of automaticity rather than a clear dichotomy between fully automatic and fully controlled behavior. Note once more that language comprehension is a highly complex, yet excessively trained skill. This may mean that some prior conditions need to be met for language to be able to drive eye movements, such as actively listening to the speech, in combination with a predisposition to make eye movements (i.e. to look around rather than to focus on one location). But once these conditions are met, the integration of language with oculomotor behavior may be unstoppable.

Finally, we believe that a decompositional approach to the investigation of the automaticity of language-mediated eye movements, along the lines of the mentioned criteria (cf. Moors & DeHouwer, 2006) is promising. For ease of comparison, future visual world studies thus could systematically include at least a discussion, if not a test, of these criteria.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419-439.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition. 93*, 79–87.

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *ActaPsychologica, 137*, 190-200.

Altmann, G. T. M., & Kamide, Y. (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

Altmann, G. T. M. and Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502-518.

Bar, M. (2007). The Proactive Brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11(7)*, 280-289.

Bargh, J.A. (1992). Why subliminality does not matter to social psychology: Awareness of the stimulus versus awareness of its effects. In R. Bornstein & T. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 236–255). New York: Guilford.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception & Psychophysics, 70*(8), 1444-1458.

Borovsky, A., Fernald, A., & Elman, J. (in press). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*.

Bravo, M.J. and Farid, H. (2009). The specifity of the search template. *Journal of Vision*, *9(1)*, 34, 1-9.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113, 2, 234-272.

Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, *108*(1), 26-50.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Dahan, D. & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453-459.

Daoutis, C. A., Pilling, M., & Davies, I. R. L. (2006). Categorical effects in visual search for colour. *Visual Cognition, 14*, 217–240.

Dell'Acqua, R., Sessa, P., Toffanin, P., Luria, R., & Jolicœur, P. (2010). Orienting attention to objects in visual short-term memory. *Neuropsychologia*, 48, 419–428.

Downing, P. E., & Dodds, C. M. (2004). Competition in visual working memory for control of search. *Visual Cognition, 11*(6), 689-703.

Egeth, H., Jonides, J. & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology, 3*, 674-698.

Eimer, M., & Kiss, M. (2010). An electrophysiological measure of access to representations in visual working memory. *Psychophysiology*, *47*, 197-200.

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.

Glenberg, A. M. (2007). Language and action: Creating sensible combinations of ideas. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp.361-370). Oxford, UK: Oxford University Press.

Han, S. W., & Kim, M. S. (2009). Do the contents of working memory capture attention? Yes, but cognitive control matters. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1292-1302.

Hintz, F., & Huettig, F. (submitted). Phonological word-object mapping is contingent upon the nature of the visual environment.

Hockett, CF, & Altmann, S. (1968). A note on design features. In TA Sebeok (Ed.). *Animal communication: Techniques of study and results of research*. Bloomington: Indiana University Press. (pp. 61-72).

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., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*, 985-1018.

Huettig, F., & Altmann, G. (2011). Looking at anything that is green when hearing ‘frog’: How object surface colour and stored object colour knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology, 64(1),* 122-145.

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, 54*, 460-482.

Huettig, F., Mishra, R. K., & Olivers, C. N. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology, 2*, 394.

Huettig, F., Olivers, C. 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.

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.

Hunt, A. R., & Kingston, A. (2003). Covert and overt voluntary attention: linked or independent? *Cognitive Brain Research, 18*, 102– 105.

James, W. (1890). *The principles of psychology*, Vol. 1. New York: Dover Publications.

Johnson, E., McQueen, J. M., & Huettig, F. (2011). Toddlers’ language-mediated visual search: They need not have the words for it. *Quarterly Journal of Experimental Psychology, 64*, 1672-1682.

Jonides, J. (1981). Voluntary versus automatic control over the mind’s eye’s movement. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Hillsdale, NJ: Erlbaum.

Kamide, Y., & Scheepers, C., & Altmann, G.T.M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37-55.

Knoeferle, P. & Crocker, M.W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science*, *30*, 481-529.

Knoeferle, P. & Crocker, M.W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye-movements. *Journal of Memory and Language*, *57*, 519-543.

Krueger, L. E. (1984). The category effect in visual search depends on physical rather than conceptual differences. *Perception & Psychophysics*, *35,* 558-564.

Kuo, B. C., Rao, A., Lepsien, J., & Nobre, A. C. (2009). Searching for targets within the spatial layout of visual short-term memory. *Journal of Neuroscience, 29*, 8032–8038.

Levin, D.T., Takarae, Y., Miner, A., & Keil, F.C. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception and Psychophysics* , *63* , 676-697.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Lupyan, G. (2008). The Conceptual Grouping Effect: Categories Matter (and named categories matter more). *Cognition, 108*, 566-577.

Mani, N., & Huettig, F. (in press). Prediction during language processing is a piece of cake - but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance.*

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomic Bulletin & Review, 14*(4), 710-716.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research, 5(1)*, 1-10.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6*, 182-189.

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin, 132*, 297-326.

Müller, H.J., Krummenacher, J. and Heller, D. (2004). Dimension-specific inter-trial facilitation in visual search for pop-out targets: Evidence for a top-down modulable visual short-term memory effect. *Visual Cognition, 11,* 577-602.

Nation, K., Marshall, C.M., & Altmann, G. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86*, 314-329.

Olivers, C. N. L. (2008). Interactions between visual working memory and visual attention. *Frontiers in Bioscience*, 1182-1191.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1243-1265.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Science*s, *11*, 105-110.

Salverda, A.P., & Altmann, G.T.M. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology: Human Perception and Performance, 37(4)*, 1122-1133.

Saslow, M.G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57,*1030–1033.

Schmidt, J., & Zelinsky, G.J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology, 62(10),* 1904-1914.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-881.

Snedeker, J., & Trueswell, J.C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*, 103-130.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 248-261.

Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 730-757.

Spivey, M. & Geng, J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235-241.

Spivey, M. J., Richardson, D. C., & Fitneva, S. A. (2004). Memory outside of the brain: oculomotor indexes to visual and linguistic Information. In J. Henderson and F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

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.

Telling, A. L., Kumar, S., Meyer, A. S., & Humphreys, G. W. (2010). Electrophysiological evidence of semantic interference in visual search. *Journal of Cognitive Neuroscience, 22(10)*, 2212-2225.

Theeuwes, J., Kramer, A.F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: capture of the eyes by new objects. *Psychological Science*, *9*, 379-385.

Theeuwes, J. , Kramer, A.F. & Irwin, D.E. (*2011*). Attention on our mind: the role of spatial attention in visual working memory. *Acta Psychologica, 137*, 248-251

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*, 80-102.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention*. Hove, UK: Psychology Press.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O_Connell, K. M. (1992). The role of categorisation in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 34–49.

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 363-377.

Wolfe, J.M., Horowitz, T.S., Kenner, N., Hyle M., Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44*, 1411-1426.

Wundt, W. (1897). *Outlines of Psychology*. Translated by C.H. Judd. (Reprinted Bristol: Thoemmes, 1999); first published in German as Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Wilhelm Engelmann.

Yang, H., & Zelinsky, G.J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49,* 2095-2103.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1-14.