Acta Psychologica 137 (2011) 138–150

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Looking, language, and memory: Bridging research from the visual world and visual search paradigms

Falk Huettig [a,b,*], Christian N.L. Olivers [c], Robert J. Hartsuiker [d]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, The Netherlands
[c] VU University Amsterdam, The Netherlands
[d] Ghent University, Ghent, Belgium

ARTICLE INFO

Article history:
Received 3 March 2010
Received in revised form 22 July 2010
Accepted 23 July 2010
Available online 3 September 2010

PsycINFO classification:
2323
2326
2340
2346
2720

Keywords:
Visual search
Visual world paradigm
Eye movements
Attention
Language processing

ABSTRACT

In the visual world paradigm as used in psycholinguistics, eye gaze (i.e. visual orienting) is measured in order to draw conclusions about linguistic processing. However, current theories are underspecified with respect to how visual attention is guided on the basis of linguistic representations. In the visual search paradigm as used within the area of visual attention research, investigators have become more and more interested in how visual orienting is affected by higher order representations, such as those involved in memory and language. Within this area more specific models of orienting on the basis of visual information exist, but they need to be extended with mechanisms that allow for language-mediated orienting. In the present paper we review the evidence from these two different – but highly related – research areas. We arrive at a model in which working memory serves as the nexus in which long-term visual as well as linguistic representations (i.e. types) are bound to specific locations (i.e. tokens or indices). The model predicts that the interaction between language and visual attention is subject to a number of conditions, such as the presence of the guiding representation in working memory, capacity limitations, and cognitive control mechanisms.

Human cognition is remarkable in its ability to integrate sensory input with increasingly abstract, high level mental representations involved in memory and language. In the visual domain this has been illustrated for example by the relative ease with which human observers can detect a complex target image from a rapidly presented series of pictures, on the basis of rather scarce written descriptions such as “road scene”, “animal”, or “not furniture” (Intraub, 1981; Potter, 1976). The question is how such higher level representations interact with the sensory information. How does our mental world interact with input from the visual environment?

In this paper we review work from two popular paradigms that originate from two rather different fields in cognitive psychology, but that start to approach each other more and more. In doing so, they increasingly bear on the issue of how language and memory interact with the visual input. The *visual world* paradigm has been developed within the area of psycholinguistics. Using eye movements as a measure, it studies the exploration of a particular visual array, as a function of the visual stimulus properties and concurrent spoken input. As such it provides an on-line measurement of how linguistic processing interacts with sensory processing of the environment. The *visual search* paradigm has been developed within the area of visual attention. It investigates the exploration of a visual array as a function of the interaction between stimulus factors on the one hand, and the observer's goal (i.e. the specific target object he or she is looking for) on the other. As such it provides a measurement of how memory (for the target) interacts with sensory processing of the environment. Eye movements are often used as a dependent measure also in this paradigm. The current review has the goal to bridge these two cornerstones of cognitive psychology, point out interesting theoretical caveats and generate exciting new questions. We will incorporate the findings from both paradigms to propose a new general framework of how linguistic input affects visual orienting.

1. Two paradigms

1.1. The visual world paradigm

Research within the visual world paradigm in psycholinguistics (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995)

[*] Corresponding author. Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Tel.: +31 24 3521374.
E-mail address: falk.huettig@mpi.nl (F. Huettig).

0001-6918/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2010.07.013

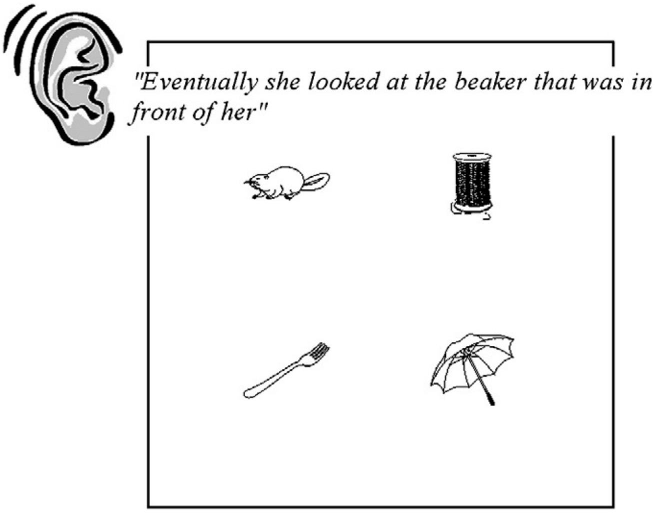
"Eventually she looked at the beaker that was in front of her"

Fig. 1. Typical visual display in the visual world paradigm. This example is based on Huettig, and McQueen (2007), but it is very similar to the displays used in other studies. Participants are presented first with the visual display. One second later the spoken sentence starts to acoustically unfold. The display remains on the screen until the end of the sentence. Of critical interest is where participants look during the acoustic duration of a critical word in the speech stream (e.g., "beaker"). In this example, the critical word is not depicted (i.e. there is no beaker) but phonological (e.g., a beaver — same word onset phonology), shape (e.g., a bobbin — similar global shape), and semantic (e.g., a fork — also a kitchen utensil) competitors of beaker are displayed.

measures eye movements to visual stimuli (e.g., semi-realistic scenes; non-scene displays of visual objects; or an array of printed words) in response to those stimuli and to concurrent spoken language. Fig. 1 shows an example display, together with an example sentence.

The onset of the presentation of the visual stimuli and the spoken language occurs either at the same time or when participants are presented with a short (about one second) preview of the display. There are two main types of tasks: direct action tasks and 'look and listen' tasks. In direct action tasks participants are required to perform an action such as clicking on an object with a computer mouse (e.g., Allopenna, Magnuson, & Tanenhaus, 1998). In 'look and listen' tasks participants are simply told that they should listen carefully to the spoken sentences they hear, that they can look at whatever they want to, but not to take their eyes off the screen throughout the experiment (e.g., Huettig, & Altmann, 2005). In both types of tasks (i.e. direct action or 'look and listen') researchers are particularly interested what happens during the acoustic unfolding of certain manipulated critical spoken words which are embedded in carrier sentences. For example, researchers are interested in how eye gaze to a display containing a beaker and a beaver changes during the acoustic duration of "beaker" when participants hear the instruction "click on the beaker", as compared to gaze directions before and after the (phonologically) ambiguous sequence in the critical word. Eye gaze is thus a function of the integration of (typically pre-activated) visually-derived representations with language-derived representations activated by the speech input.

Although the dependent measure therefore does not straightforwardly reflect linguistic processing, it is language processing that visual world researchers are traditionally most interested in. Much of the research has focused on syntactic ambiguity resolution (e.g., Tanenhaus et al., 1995; Trueswell, Sekerina, Hill, & Logrip, 1999), the nature of prediction during sentence processing (Altmann, & Kamide, 1999, 2009; Kamide, Altmann, & Haywood, 2003), speech perception (e.g., Allopenna et al., 1998; Salverda, Dahan, & McQueen, 2003), lexical ambiguity (Huettig, & Altmann, 2007), bilingual word recognition (e.g., Spivey, & Marian, 1999; Weber, & Cutler, 2004), disfluency (e.g., Arnold, Fagnano, & Tanenhaus, 2003), and many other psycholinguistic issues (see Huettig, Rommers, & Meyer, this issue, for a comprehensive review). In doing so it has advantages over other methods. For example, a great benefit of the visual world method is its ecological validity. Participants are not required to perform an unnatural meta-linguistic task (such as making an overt decision about whether a target stimulus is a word or a non-word) as in many psycholinguistic methods. The tasks in visual world studies such as simply scanning a visual display (e.g., Altmann, & Kamide, 1999; Huettig, & Altmann, 2005), clicking with a computer mouse on the word's visual referent (e.g., Allopenna et al., 1998), or touching the visual referent on a touch-screen computer display (e.g., Yee, & Sedivy, 2006) may be regarded as more realistic language-mediated behaviors than making a lexical decision by pressing a button.

Another real boon of the method is that it provides fine-grained estimates of *ongoing* cognitive processing in the form of fixations to different positions in the visual display over time. The study by Allopenna et al. (1998) nicely illustrates this point. Their participants were presented with a display of four pictures of real objects and four geometric shapes. They were instructed to click on a target picture with a computer mouse (e.g., "Pick up the beaker") and then move it next to, above, or below one of the four geometrical figures. Allopenna et al. found that a depicted object whose name begins in the same way as a word that is being heard (e.g., beetle) competed for eye fixation more strongly, and for a longer period of time, than an object whose name overlaps with the word being heard at word offset (e.g., speaker). Although participants looked at the pictures of both types of competitors more than at completely phonologically-mismatching distractors, they looked more often at onset-matching referents (e.g., the picture of the beetle) than at offset-matching referents (e.g., the picture of the speaker). Thus onset phonological overlap is more important for spoken word recognition than non-onset phonological overlap between target and competitor. This all happens while the acoustic signal unfolds. Thus the time-sensitivity of the visual world paradigm is particularly well-suited to draw inferences about on-line cognitive processing.

Importantly, the focus of at least some recent visual world research has shifted more towards exploring the interaction between language and vision rather than core issues of linguistic processing (e.g., Altmann, & Kamide, 2007; Gleitman, January, Nappa, & Trueswell, 2007; Huettig, & Altmann, 2007; Huettig, & McQueen, 2007; Knoeferle, & Crocker, 2006; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). Huettig and Altmann (2005) (see also Yee, Overton, & Thompson-Schill, 2009), for instance, investigated whether semantic properties of individual lexical items can direct eye movements towards objects in the visual field. Participants who were presented with a visual display containing four pictures of common objects directed overt attention immediately towards a picture of an object such as trumpet when a semantically related but non-associated target word (e.g., '*piano*') was heard. Different measures of semantic relatedness (semantic feature norms, Cree, & McRae, 2003; Latent Semantic Analysis, Landauer, & Dumais, 1997; and McDonald's Contextual Similarity measure, McDonald, 2000) each separately correlated with fixation behavior (Huettig, & Altmann, 2005; Huettig, Quinlan, McDonald, & Altmann, 2006). These data show that language-mediated eye movements are a sensitive measure of overlap between the conceptual information conveyed by individual spoken words and the conceptual knowledge associated with visual objects in the concurrent visual environment.

1.2. The visual search paradigm

As in the visual world paradigm, in visual search studies the participant is presented with a display of multiple objects. The participant has the explicit task of finding a pre-specified target defined by a certain feature, for example a red object among green objects. He or she is usually notified of what to look for through either spoken or written instructions, followed by a few example trials. The task is usually to respond as fast as possible by either determining the presence of the target (present/absent response), the identity of the target (e.g., is the red object a square or a triangle), or by making an eye movement towards it.

Unlike in visual world studies, in visual search an important manipulation has been the set size, which is the total number of objects in the display. This is because the rate at which response times (RTs) increase with increasing set size (the "search slope") provides a measure of the efficiency with which the display can be searched. The flatter the search slope, the more efficiently the target is found. This efficiency is taken to reflect the ease with which attention can select relevant information from a display of competitors. The search efficiency is known to depend on stimulus factors as well as on the goals of the observer. For example, all else being equal, a target that carries a distinctive and salient feature relative to the distractors (e.g., a red object among green objects) is found more easily (Treisman, & Gelade, 1980), especially if the observer also knows what this feature is (Wolfe, 1994). Similarly, salient objects tend to attract eye movements (e.g., Theeuwes, Kramer, Hahn, & Irwin, 1998). A substantial part of the visual search literature has focused on what the basic visual features are that observers can find efficiently (e.g., color, orientation, motion), whether or not observers can easily find conjunctions of basic features, and whether the search process – when it happens to be inefficient – is serial in nature or rather a limited-capacity parallel process. This vast literature has been extensively reviewed elsewhere (Wolfe, 1998, 2003; Palmer, Verghese, & Pavel, 2000), but important for present purposes is the general consensus that search for salient perceptual features tends to be fast and efficient, whereas search for visually complex or semantically defined objects tends to be slow and inefficient (Wolfe, 1998). The latter finding means that the semantic codes are not readily available for search, at least not within the time frame of a typical trial.

More recently, researchers have become interested in the nature of the target representation. When observers are looking for an object, they must have some mental description of that object that eventually guides their attention in the right direction. This description is often referred to as the *target template*, but another frequently used term is *attentional set*. In fact, most visual search studies regard the target template as simply given: it is assumed that when the participant is instructed to look for a specific feature at the beginning of the experiment (e.g., the written instruction "look for the red object"), he or she will set up some sort of veridical description of the target for the remainder of the task and that is that. The existence of such a target-specific representation and that it indeed affects visual search have been amply demonstrated by manipulating the instructions on what to look for. For example, Folk and Remington (1998) asked observers to look for red targets in one condition, and for green targets in another. Prior to the visual search display, irrelevant cues were presented that could coincide with the subsequent target location (valid cue condition), but more often did not (invalid cue condition). Cues could also be red or green, independent of the target color. Although irrelevant to the task, the cues had a clear effect on target search, such that valid cues led to speeded RTs. However, this only occurred for cues that carried the looked-for color (e.g., red cues when the target was also red, as opposed to when the target was green). In other words, attentional guidance towards the irrelevant cues was determined by the attentional set of the observer.

Exactly to what extent selection in visual search is determined by the salience of the visual features on the one hand, and the attentional set of the observer on the other, has been a matter of extensive and heated debate over the past two decades, involving those advocating pure salience-driven selection on one side (e.g. Theeuwes, 1991, 1992, Schreij et al., 2008, 2010; Yantis & Jonides, 1984), those advocating pure voluntary, top–down driven selection on the other side (Folk et al., 1992; Folk, & Remington, 1998; Folk, Remington, & Wu, 2009), and those taking an intermediate position (e.g. Olivers, & Humphreys, 2003a; Wolfe, 1994).

1.3. A theoretical no-man's land?

It is obvious that the visual world and visual search paradigms have much in common. Both paradigms measure visual attention in a spatial array of multiple objects, and in both paradigms researchers are interested in studying the interaction between the visual stimuli and higher order cognitive biases as induced by task goals, or language. However, they approach these issues from rather opposite ends of the information processing chain. While visual search investigators are mainly interested in the mechanisms of visual orienting, and usually treat top–down biases as simply given (i.e. pre-existing or induced only once at the start of an experimental block), visual world investigators are usually interested in the dynamics of linguistic processing, and treat the visual attention system as simply given (that is, they use it as a tool to answer their linguistic questions). This has created a theoretical no-man's land in which the exact interaction between higher order cognitive representations (conceptual and linguistic) and visual attention has been left unspecified.

Many detailed models of visual search exist, and they are virtually all concerned with the interaction between the bottom–up salience of the stimulus and the top–down goals of the observer (e.g., Cave, 1999; Humphreys, & Müller, 1993; Itti, & Koch, 2000; Palmer et al., 2000; Treisman, & Sato, 1990; Wolfe, 1994). Most of these theories assume some form of topographically organized spatial map, in which bottom–up stimulus-related activation is combined with top–down activation stemming from the search goals of the observer. The location of the object with the highest activation will then draw attention (i.e. be selected), followed by the object with the next highest activation, and so on. However, the nature of "what one is looking for" and in which type of memory it is kept are typically less well described, if described at all. Where in the cognitive system is the target template kept? And what is the nature of its representation? Is it linguistic in nature (in line with the verbal instructions given prior to the task), or is it more visual (in line with the goal of finding a visual object)? Moreover, top–down guidance in these models is usually limited to guidance on the basis of basic features like color and form, rather than, for instance, semantics.

Similarly, within the field of psycholinguistics, there are currently few explicit models on exactly how language affects visual orienting. An early linking hypothesis was proposed by Tanenhaus, Magnuson, Dahan, and Chambers (2000) to explain the effect of phonological overlap between spoken word and visual referent. They proposed that "informally, we have automated behavioral routines that link a name to its referent; when the referent is visually present and task relevant, then recognizing its name accesses these routines, triggering a saccadic eye movement to fixate the relevant information" (p. 565). This phonological matching hypothesis however cannot account for effects of semantic and/or visual overlap. Later, Dahan and Tanenhaus (2005) argued instead that "word–object matching occurs at the level of visual features and not at the level of pre-activated sound forms" (p. 457). To explain phonological effects within this visual matching hypothesis, Dahan and Tanenhaus (2005) argue that hearing/be/ activates the visual form features of all objects whose names start with these phonemes, including both a beaker and a beetle. These are then matched with the visual form features of the depicted beaker and the depicted beetle, inducing a bias towards those objects. The visual matching hypothesis can also explain visual effects (e.g., looking at a cable on hearing "snake", Dahan, & Tanenhaus, 2005; Huettig, & Altmann, 2004;2007). However, neither of these two hypotheses can explain semantic effects (e.g., looking at a trumpet on hearing "piano", Huettig, & Altmann, 2005). Huettig and Altmann (2005) therefore concluded that eye movements are driven by the degree of match along multiple dimensions (including, but not restricted to simple visual form), between a word and the mental representations of objects in the concurrent visual field. Altmann and Kamide (2007) further specified this as follows: first, seeing the trumpet in the display activates featural representations such as visual form, function, and contextual dependencies (McRae, de Sa, & Seidenberg, 1997; Rogers, & McClelland, 2004). Similarly, hearing "piano" also activates its corresponding featural representations. Second, overlap in featural representations causes overlapping pre-activated representations (i.e. the representation of the trumpet in the display and its associated representations) to increase in activation even further. Third, changes in activation state change the attentional state of the cognitive system. Finally, changes in the attentional state will increase the probability of a saccadic eye movement towards the spatial location associated with the change in attentional state.[1]

Such changes in attentional state have been more explicitly modeled in a neural network developed by Mayberry, Crocker, and Knoeferle (2009). The model describes how linguistic expressions result in higher activation (upscaling) of visual referents, which in turn activate related linguistic constructs. For example, the word "detective" in a linguistic utterance may enhance the representation of a picture of a detective present in a visual display, which in turn may trigger the verb "to spy on", because the detective in the display actually happens to be spying on someone. The activation of the verb in turn evokes a referent that can be spied on, which is then fed back to potential matches in the display, which become more active (i.e. more attended). This way the different attentional states emerge naturally from the language–vision interactions.

Although intuitively appealing, such accounts leave a number of important issues unanswered. For example, do changes in activation *constitute* or *cause* a shift in attention. Note that in this respect activation of a semantic or visual representation per se is not sufficient to cause a shift in orienting (or eye gaze for that matter). Some sort of spatial pointer appears to be required that binds these representations to a certain location. However, space appears not to be represented in the Mayberry et al. (2009) model. Others have assumed a role for spatial indices (e.g., Altmann, 2004; Altmann, & Kamide, 2007; Ferreira, Apel, & Henderson, 2008; Richardson, & Spivey, 2000), but are unclear about the nature of these indices. That is, they do not specify what type of memory is involved in maintaining these indices, nor under which constraints (though see Spivey, Richardson, & Fitneva, 2004 as discussed later). In the end the activation of a semantic or visual form representation needs to result in the activation of the associated location. For example, Altmann and Kamide (2007), in line with connectionist accounts of language processing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Elman et al., 1996; MacDonald, & Christiansen, 2002), make no distinction between long-term memory and working memory. Altmann and Kamide (2007, p. 515) write about 'episodic traces' but also state that their account does not in fact rely on them. Moreover, they state that "we take this episodic trace to be a temporary record of both the experience of the object, including its location, and the conceptual representations associated with that experience (and henceforth, we use the term 'episodic' to refer to such temporary records or traces)" (p.512). It is clear that conceptual representations must be stored in long-term memory. However, Altmann and Kamide (2007) go further than that and argue that "language-mediated eye movements are little different theoretically (and perhaps no less automatic behaviorally) than priming effects which have elsewhere been explained in terms of spreading activation and/or conceptual overlap (cf. Collins, & Loftus, 1975; Neely, 1977)" (p. 514). Priming however is typically considered as a long-term memory phenomenon (e.g., Jackendoff, 2002; but see Neely, 1991, for the view that participants use the prime to generate a short-term expectancy set that consists of potential targets).

Others have implied or even explicitly argued that short-term memory, or working memory plays an important role (Ferreira et al., 2008; Knoeferle, & Crocker, 2007; Richardson, & Spivey, 2000; Spivey, & Geng, 2001; Spivey et al., 2004), mainly on the basis of evidence showing that participants prefer to orient to the locations of objects that are referred to in the spoken input, even when those objects are no longer there. The visual object representations and their associated locations must thus have resided in some form of memory (see also Theeuwes et al., this issue; but see Altmann & Mirkovic, 2009 for arguments that this is not necessarily working memory). These previous discussions of (working) memory in the visual world literature however have remained somewhat vague. Knoeferle and Crocker (2007) for instance incorporate an explicit working memory component into their account, but they do not discuss the nature of the working memory component in any detail (other than saying that its contents experience some decay). Mayberry et al. (2009), in the description of their connectionist model of situated language comprehension, use the term 'memory' only once (p.488) right at the end of the paper when discussing future directions. Similarly, Ferreira et al. (2008), when explaining linguistic–visual interactions only talk about "memory", without further specification. Occasionally they appear to suggest episodic bindings between different types of representation. For example, they mention "object files" as a possible equivalent of such bindings. Object files are token representations in which object properties are combined and tied into a spatio-temporal file (Kahneman, Treisman, & Gibbs, 1992), and, within the visual attention literature, have been linked to visual working memory (see Pylyshyn, 1989; Cavanagh, & Alvarez, 2005). This idea was made most explicit by Spivey et al. (2004), who proposed that it is the working memory that mediates language–vision interactions, through a visuospatial pointer or indexing system which, following Pylyshyn, temporarily stores the positions of objects.

In the present paper we aim to build a strong case for why it is indeed working memory that is crucial for language–vision interactions. Importantly, we explore and extend this idea with what is known from the visual search literature. Integrating the evidence across the two different research fields may answer important questions, and will most certainly generate new ones.

2. Bridging the gap: working memory

2.1. The role of working memory in visual world studies

So where is the link forged between a linguistic utterance on the one hand, and an attentional orienting response to a visual object on the other? In Fig. 2 we propose a general framework which takes current accounts such as the one proposed by Altmann and Kamide (2007) as a starting point. We propose that the basis for language–vision interactions lies in long-term memory. It has to, of course, since that is where the names and meanings of words and objects, and the

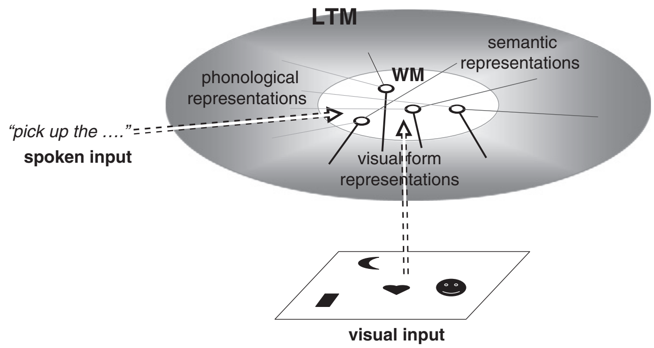

Fig. 2. A general framework for how working memory serves as the central interface between language and visual orienting. When looking at a display, visual form representations become bound to specific spatio-temporal indices within working memory. Given sufficient time, associated semantic and phonological codes will also be bound to the existing visuospatial working memory representation, thus creating a nexus of linguistic and visuospatial activity. The most active location in working memory will eventually determine the most likely direction of the eye movement at a given point in time.

[1] Note that Altmann and Kamide's account is more complex than sketched out here. To account for anticipatory eye movements (for instance reflecting verb-thematic fit, plausibility, tense information, etc.) they invoke affordances about an object (e.g., an full glass affords that it can be drunk from). For our present discussion these points are not essential (see Altmann, & Kamide, 2007, for further detail).

associations between them are stored. Linguistic (phonological, syntactic, and semantic) and non-linguistic (e.g., visual) representations are connected and activation cascades through the system such that, on hearing spoken words,[2] activation of candidate phonological structures in long-term memory typically activate associated semantic structures, which in turn, perhaps to a lesser degree, activate their associated visual representations. Similarly, seeing visual objects activates associated stored visual representations in long-term memory which, given sufficient time, activate semantic structures and eventually phonological forms (i.e. the object names, cf. Meyer, & Damian, 2007; Morsella, & Miozzo, 2002; Huettig, & McQueen, 2007).

Long-term memory thus serves as a knowledge base, providing useful information about stable aspects of our environment. However, in daily life this knowledge often has to be linked to unstable and often rather arbitrary information on the object's current location, when it occurred, and which other objects it related to then. For example, when passing the fields on your way to work, you may notice that the horse may stand to the left of the barn on one day, but to the right on the day after. This means you rely on momentary and rather arbitrary knowledge about that particular situation, rather than on general knowledge of horses and barns. Similarly, people have little trouble understanding an arbitrary situation such as expressed by "Yesterday she found the garlic crusher in the flower pot", even though there is no a priori association between yesterday, finding, garlic crushers or flower pots, nor are people likely to change their world model on the basis of this particular instance. In other words, cognition is often *situated*: it is not only based on general knowledge (e.g., scene schema knowledge), it is based on events that happen here and now, or there and then (Brooks, 1991; Pylyshyn, 2001; Spivey et al., 2004). Importantly, the situation is often no different for the typical visual world task, in which several rather unrelated objects are placed arbitrarily within the array, and people listen to rather arbitrary sentences like "she looked at the trumpet".[3]

Like Knoeferle and Crocker (2007) and Spivey et al. (2004); see also Ferreira et al., 2008 we propose that working memory plays a crucial role in language–vision interactions. Working memory appears to be the ideal mechanism to serve exactly the function of grounding cognition in space and time, allowing for arbitrary short-term connections between objects. It enables us to link knowledge (as provided by long-term memory) to the here and now, or, when planning things, the there and then. We view working memory as the capacity to hold and bind arbitrary pieces of information. This idea is by no means new. In vision for example, theorists have proposed that visual objects are "instantiated", that is their long-term representations (often called "types") are bound to a specific location and moment in time (often called an object file, token, or index, Kahneman et al., 1992; Kanwisher, 1987; Pylyshyn, 2001; see also Hoover & Richardson, 2008). Moreover, this idea connects well with existing views on working memory. For example, Baddeley (2000) has proposed an episodic buffer, which is capable of holding and binding representations from different modalities.

Furthermore, the idea that working memory is what instantiates an object is consistent with 'embodied' cognition views of working memory (Postle, 2006; Spivey et al., 2004; Wilson, 2002). For example, some have argued that a memory of an object's location in the visual array is nothing more than implementing, but not executing, a motor program to either saccade or point towards that location (Theeuwes, Olivers, & Chiszk, 2005), similar to the argument that attending to an object is nothing more than planning to look at that location (Klein, 1980). In these cases the motor code provides the specific spatial instantiation of the object. Conversely, eye movements have been shown to disrupt spatial working memory (Baddeley & Lieberman, 1980; Smyth & Scholey, 1994). Finally, note that our view is also not so different from what Altmann and Kamide (2007) describe as episodic traces, in that working memory might provide the basis for what they refer to as the *experience* of an object, "including its location, and the conceptual representations associated with that experience" (p. 512). The main difference is the claim that working memory is a necessary condition for this episodic experience.

What then is the chain of events when observers view a typical visual world display of a random collection of randomly spaced objects, while listening to a sentence? First, we assume that a restricted number of objects from the display are encoded in what is initially a visuospatial type of working memory (Baddeley, & Hitch, 1974; Logie, 1995; Pylyshyn, 1989; Cavanagh, & Alvarez, 2005). At this stage, specific visual shapes are bound to their respective locations. In the case of unknown shapes, these representations may be based entirely on visual routines, but in the case of visual world, the usually very familiar objects will rapidly trigger perceptual hypotheses based on long-term memory codes. With the activation of these codes comes the cascaded activation of associated semantic, and potentially also phonological codes, all within a few hundreds of milliseconds. In the end, given sufficient time, this results in a nexus of associated knowledge about the object, which, within working memory, is all bound to the object's location. This binding of an entire complex of representations to a location is necessary to explain linguistically-mediated eye movements. Note again that merely activating associated long-term memory nodes is insufficient to achieve this, since these are not associated with arbitrary locations.

Second, a spoken utterance will activate long-term phonological and semantic codes. It seems that this may be sufficient to explain at least some visual world findings. For example, when viewing a trumpet while hearing "he will play the triangle", the mere activation of *triangle* spreads to *trumpet* on the basis of semantic and phonological (same initial phonemes) similarity. Activation of *play* also activates things that can be played (cf. Altmann, & Kamide, 1999, 2007). In turn, this all strengthens the nexus of the specific trumpet's representations and location within working memory, increasing the probability of triggering a saccadic eye movement. Summarizing, the chain of events is that first the visual display is processed up to a high level, including the creation of conceptual and linguistic representations. At this high level, these representations subsequently match up with those activated by the linguistic input, activation that then feeds back to the linked location.

It can be argued that even the activation of linguistic representations and their associations is not simply a matter of long-term memory. Language for instance is abundant with phonological, syntactic, and semantic ambiguities — ambiguities that need an on-line memory structure (or "situation model") in order to be resolved. For example, Jackendoff (2002;2007) has argued for the necessity of working memory in order to be able to account for (i) the binding of linguistic structure (see also Marcus, 1998, 2001; Knoeferle, & Crocker, 2007), (ii) the possibility of several occurrences of an item activating the same 'node' (working memory may contain more than one token of a type stored in long-term memory), and (iii) the encoding and instantiating of variables (since all combinatorial rules of language require typed variables; see Marcus, 2001, for discussion).

Moreover, it has been shown in visual world studies that world knowledge about who is the plausible agent of an action influences eye gaze (e.g., "the man/girl will ride the motorbike/caroussel"), as anticipatory eye movements (i.e. object fixations) depended on the particular subject/verb combination (Kamide et al., 2003). It is difficult to argue that "man" automatically primes motorbike, let

[2] For ease of exposition we mainly use single words in our examples throughout this paper. Note however that our account is not per se restricted to the interaction of single words with visual information but also holds for the interaction of sentence and discourse-level representations with representations retrieved from the visual environment.

[3] In the research by Altmann and colleagues (e.g., Altmann, & Kamide, 1999, see also Knoeferle, & Crocker, 2006) the spoken sentences are typically related to what is depicted in a semi-realistic scene. However, even in these studies objects are often arranged in quite an arbitrary manner.

alone that "girl" automatically primes "caroussel". Instead, the preferences here appear to be caused by an on-line model constructed on the basis of the linguistic input and the visual scene. Similarly, Knoeferle and Crocker (2007) have shown that participants rely on depicted events over stereotypical thematic role knowledge for incremental thematic interpretation even when the scenes are no longer co-present on a computer display. We argue that working memory is exactly what is needed for building such on-line models, as it allows for arbitrary objects to be linked to times, places, and each other.

2.2. The role of working memory in visual search

At the same time, working memory provides a useful framework for explaining top–down biases in visual search. It provides an answer to the question where the attentional set is maintained. Probably the most influential general framework in this respect is Desimone and Duncan's (1995) *biased competition* model, but more detailed models of visual search have been developed before and after that which also more or less explicitly assume an important role for short-term or working memory (e.g., Bundesen, 1990; Bundesen, Habekost, & Kyllingsbæk, 2005; Duncan, & Humphreys, 1989; Grossberg, Mingolla, & Ross, 1994; Hoffman, 1979). Desimone and Duncan based their framework on monkey physiology data showing that the same neurons were active during active remembering as well as during the actual presence of the search target (Chelazzi, Miller, Duncan, & Desimone, 1993). They argued that which perceptual content is prioritized is directly determined by the contents of working memory. Keeping an object in visual working memory pre-activates perceptual object representations and thus automatically biases visual selection towards those objects when they actually appear in the display. In fact, Desimone and Duncan (1995) concluded that "[v]isual search simply appears to be a variant of a working memory task, in which the distractors are distributed in space rather than in time." (p. 207). As pointed out by Pashler and Shiu (1999), similar ideas can be traced back to Pillsbury (1908), who stated that "searching for anything consists ordinarily of nothing more than walking about …. with the idea of the object prominently in mind, and thereby standing ready to facilitate the entrance of the perception when it offers itself.", and Oswald Külpe (1909), who wrote that "impressions which repeat or resemble ideas already present in consciousness are especially liable to attract the attention" (p. 439).

Note here that in a typical visual search task the events typically occur in reverse order as compared to visual world tasks: Participants usually first receive a spoken or written instruction as to what to look for. This information is presumably processed within working memory, which also retrieves and activates the sought for visual feature. Only then the visual display appears, containing the target. This target then matches with the representation in visual memory and adds the missing spatial information to the nexus of representations describing the target. So here the match presumably occurs at the visual end of working memory rather than the linguistic or semantic end.

3. Experimental evidence

What is the evidence for a role of working memory in attentional guidance, and how does this relate to linguistic influences on attention? In this section we treat findings from the visual attention literature that are directly relevant to the hypothesis that working memory mediates the language-based guidance of attention.

3.1. Does the content of working memory automatically guide attention?

A first prediction from the working memory hypothesis is that a working memory representation should be sufficient to guide attention, and may do so automatically. There is little relevant data from the visual world paradigm testing this prediction, but evidence has recently been provided within the visual search paradigm (Olivers, Meijer, & Theeuwes, 2006; Olivers, 2009; Soto, Heinke, Humphreys, & Blanco, 2005; Soto, Humphreys, & Heinke, 2006; Soto, & Humphreys, 2007; see also Downing, 2000; Farah, 1985; Pashler, & Shiu, 1999, for earlier evidence from spatial cueing, psychophysics, and rapid serial presentation tasks). Fig. 3 illustrates the general procedure, which consists of two interwoven tasks. The observer is first presented with a relatively simple visual object, for example a colored disk (Olivers, Humphreys, & Braithwaite, 2006) or colored canonical shape of some other kind (e.g., a square, circle, or triangle; Soto et al., 2005). The observer is asked to remember the object for a later memory test, at the end of the trial. In the meanwhile, while holding the object in working memory, the participant is asked to switch to a visual search task, in which he or she is asked for a target that bears no relationship to the memory content. Participants in the Olivers et al. study were asked to look for a diamond among disks. The crucial manipulation was that one of the distractors in the visual search task could match the object held in memory, in color and/or shape. The idea was that if working memory content indeed guides visual selection, attention may be diverted away from the target and towards the matching distractor. This should result in RT costs.

Indeed, this is what was found: the presence of a matching distractor resulted in prolonged RTs compared to when no such distractor was present. This was confirmed when eye movements were measured. Both Soto et al. (2005) and Olivers, Humphreys, et al. (2006) found that the first saccade after search display onset was more likely to go in the direction of a distractor when this distractor matched the content of working memory. Note that observers had no incentive of looking at the matching distractor, since it did not match the target description and would only hinder visual search. This provides support for the notion that the guidance of attention from working memory occurred indeed automatically, in accordance with the biased competition framework.

3.2. The guiding content has to be in working memory

A second prediction from the working memory hypothesis is that it is *necessary* for representations to be active in working memory to guide attention. Again, there is little relevant data from the visual world paradigm testing this prediction. Some visual world studies use an explicit task (e.g., "click on the beaker") that forces participants to listen to the spoken input, whereas other studies employ free listening and free viewing ('look and listen' tasks, e.g., Huettig, & Altmann, 2005). One might expect more working memory involvement in the first case, but this has never been investigated thoroughly. Moreover, even in the free viewing and -listening cases, participants are still instructed to look at the display and listen to the sentence.

More direct comparisons have been made within visual search studies. For example, Olivers, Humphreys, et al. (2006); see also Downing, 2000; Soto et al., 2005 found no evidence for a selection bias towards visual information (i.e. color) that merely had to be viewed prior to the visual search display, whereas such a bias existed for information that had to be remembered for a later memory test. In another experiment of the Olivers et al. study, observers had to remember an object, but could then first complete the memory test before they moved on to the search task, rather than the other way around. By then the memorized object was no longer relevant, and indeed it ceased to affect visual search. In a final experiment, observers were asked to remember two objects. A few moments later they were then notified that one of the two objects was no longer relevant. In a subsequent search task, the no longer relevant object did not affect search, whereas the still relevant object did. In all, these results suggest that mere exposure, and the priming of long-term representations that this may cause, is insufficient to induce attentional guidance.

The same conclusion was reached by Soto and Humphreys (2007) for memories of verbal, rather than visual, material. Instead of presenting a picture of for example a red square prior to the search

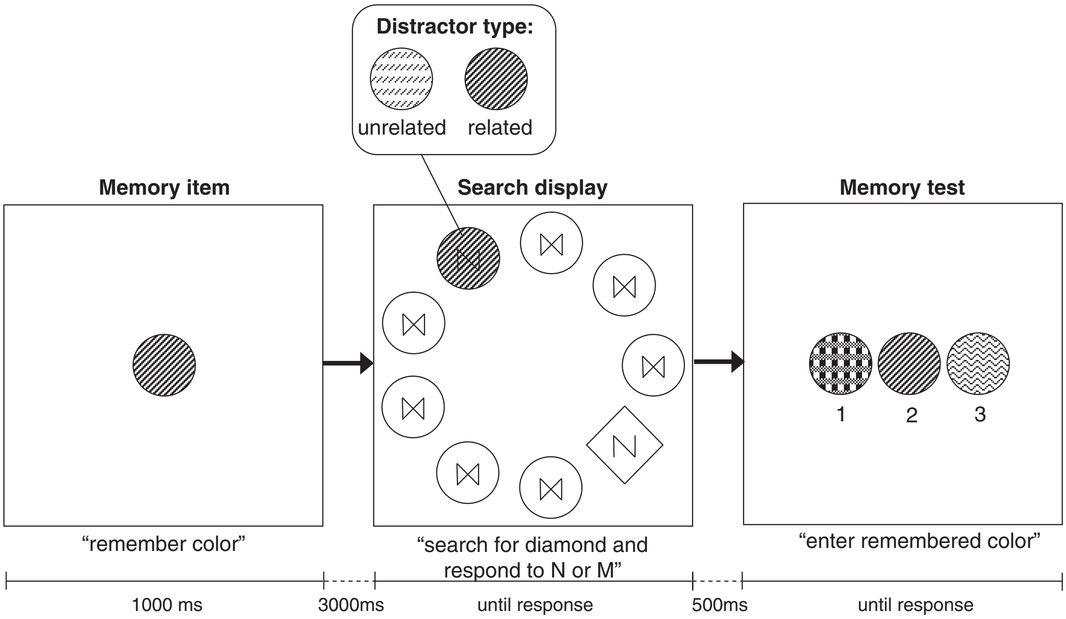

Fig. 3. General procedure of the interwoven working memory and visual search tasks. This example is based on Olivers, Meijer, and Theeuwes (2006), but is very similar to the procedures used in other studies (e.g., Downing, & Dodds, 2004; Houtkamp, & Roelfsema, 2006; Soto et al., 2005; Woodman et al., 2007). In this example, observers would first have to remember a color, and then search for a diamond in the subsequent visual search display (and respond N or M to the letter inside). This display also contained a distractor of a salient color, which could match the color in memory.

display, they presented a printed description of the object (e.g., "red square"). They again found increased interference from distractors matching the description, but only when observers had to remember the description or say it out loud just prior to the display (which Soto and Humphreys argued automatically puts the description in verbal working memory). No effect on search was found when participants had to perform an identification task on the written description but could then forget about it. Thus there was no evidence of priming of long-term memory representations. Instead, the content had to be actively maintained in working memory.

Other important evidence comes from visual search experiments by Wolfe, Klempen, and Dahlen (2000). In these experiments, the visual search display remained identical from trial to trial. In fact, during a block of up to 350 trials, the items (e.g., an array of capital letters arranged around a circle) never disappeared from the screen. All that changed from trial to trial was the instruction as to what was the target letter, as indicated by a visually dissimilar small-case version of the letter inside a central enclosed region. The idea was that if some rich memory representation of the entire visual scene is built up and maintained from trial to trial (including semantic and phonological codes), then visual search should become more and more efficient with each trial. It did not. Search was as inefficient after 350 trials as it was after the first or second trial. Thus it appears that even with an extended preview, the representation of where certain letters were remained rather scarce. Wolfe et al. argued that the momentary awareness of the display during visual search may rapidly disintegrate once the trial is over and attention has disengaged. Only when observers actively attend to the display are object shapes and identities bound to their locations. Indeed, when observers were explicitly asked to *remember* the search array (which was subsequently taken away from the screen), search became more and more efficient with increasing trial number, suggesting that when the display is in working memory, the bindings are indeed available.

3.3. Guidance through long-term memory

Does this mean that attention cannot be influenced by long-term memory at all? The answer is clearly no. For example, we know from visual search studies that when observers repeatedly search for the same set of targets, search becomes highly automated, such that the same objects are later very hard to ignore when they become irrelevant to the task (Schneider, & Shiffrin, 1977; Shiffrin, & Schneider, 1977). Part of this automation fares, among other things, on implicit priming effects from the previous trials (e.g., Maljkovic, & Nakayama, 1994; Müller, Heller, & Ziegler, 1995; Olivers, & Humphreys, 2003a; Olivers, & Meeter, 2006). Such priming effects have been shown back as far as eight trials and have been shown to be rather immune to top–down knowledge about the target identity. Other studies have shown that observers also learn implicitly about the spatial lay-out of displays, leading to more efficient search when this lay-out repeats (even when this repetition goes unnoticed by the participant, Chun, & Jiang, 1998). The automatic and implicit nature of such findings (like priming building up implicitly from trial to trial) argues for the direct involvement of long-term memory. However an important difference here with visual world studies is that aspects of the stimulus repeat: Trial after trial observers see the same target, or the same lay-out, thus enabling learning and the development of some degree of automaticity (Schneider, & Shiffrin, 1977; Shiffrin, & Schneider, 1977). Consistent with this idea, additional working memory load per se (as induced by an extra task) does not affect the efficiency of visual search for a consistent target (Woodman, Vogel, & Luck, 2001), but it does when the target changes from trial to trial (Woodman, Luck, & Schall, 2007). Note that in visual world there either is no target, or the target changes from trial to trial, together with the spoken input and the objects present in the display.

The guidance of attention on the basis of long-term memory is especially evident when searching through real-life scenes such as a picture of a kitchen or a street view. For example, when looking for a toaster, the initial overall recognition of a kitchen scene may generate previously learned biases towards the middle region of the display, towards flat surfaces (evidence for which may be provided by low-level sensory processes), away from the sink and towards electrical sockets. It is beyond the scope of the present review to extensively treat the scene perception literature, and there are excellent reviews elsewhere (e.g., Henderson, & Ferreira, 2004; Rensink, 2000). For the present purpose we again point towards the fact that the numerous occasions one enters a similar scene (such as a kitchen) enables one to learn about the spatial contingencies. In sum, the research reviewed in this section supports the notion that guiding content has to be in working memory but this does not rule out that there are some influences of long-term memory too.

3.4. Capacity limits

Another direct prediction derived from the idea that working memory is central to language-based attentional effects in the visual world paradigm is that such effects are subject to capacity limitations. On the one end, the number of spatial pointers or indices available to the visual system appears to be limited to about four (with the bulk of the estimates ranging between three and five). On average four is the maximum number of objects that can be efficiently prioritized, cued, tracked, counted, and actively remembered (Yantis, & Johnson, 1990; Burkell, & Pylyshyn, 1997; Pylyshyn, & Storm, 1988; Atkinson, Campbell, & Francis, 1976; Mandler, & Shebo, 1982; Trick, & Pylyshyn, 1994; Sperling, 1960; Phillips, 1974; Pashler, 1988; Luck, & Vogel, 1997; see also Cowan, 2001). The prediction then is that visual world type effects should be diminished with more than four objects in the display. In most visual world type studies, however, the number of objects in the display does not exceed four or five objects. Some earlier studies have used more than four objects (e.g. Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Hanna & Tanenhaus, 2004; Keysar, Barr, Balin, & Brauner, 2000; Metzing & Brennan, 2003; Tanenhaus et al., 1995), but none of these have systematically manipulated (or reported the effects of) set size as one would do in standard visual search studies. We know of one exception: Sorensen and Bailey (2007) report some preliminary data in this regard. They replicated the semantic (e.g., "piano"–trumpet) competition finding of Huettig and Altmann (2005) with the typical four-object array. However they found that as array size increased to nine and sixteen objects, semantic competition effects decreased (though semantic competition was still significant in all types of arrays). More specifically, array size had the most dramatic effect on the timing of semantic competition: the larger the array size, the later the effect of semantic competitors occurred. This delay may for example reflect the sequential scanning of groups of up to four objects in the larger arrays, but more research is needed.

A question for further research is whether the limitations, when present, are mainly spatial in nature (i.e. there are only four indices) or whether there is also a limit to the number of semantic and phonological representations that can be tied in. One way of investigating this may be to link multiple meanings to one location, for example by using ambiguous or superimposed stimuli.

3.5. Binding language to space

The working memory account predicts that, given sufficient time, semantic and phonological representations will be tied to a visual object and its location. There is empirical evidence from the visual world paradigm consistent with this prediction. This evidence comes from experiments in which participants first look at a visual display of multiple objects, after which the display is removed and observers listen to linguistic input (Altmann, 2004; Altmann, & Kamide, 2007, Ferreira et al., 2008; Knoeferle, & Crocker, 2007; Richardson, & Spivey, 2000; Richardson, Altmann, Spivey, & Hoover, 2009; Spivey, & Geng, 2001). The linguistic input may be related to visual objects that had been on the screen but are no longer present. The interesting finding is that listeners tend to rapidly fixate empty regions of space that were previously occupied by the items but that are now only alluded to in the spoken input, even though these locations are not relevant to understanding the linguistic expression. Similar evidence has been reported in the visual attention literature (Dell'Acqua, Sessa, Toffanin, Luria, & Jolicoeur, 2009; Eimer, & Kiss, 2010; Kuo, Rao, Lepsien, & Nobre, 2009; Theeuwes, Kramer & Irwin, this issue). Using measures of covert attention (i.e. detection or identification performance at the probed location) rather than overt eye movements, it is shown that observers prioritize the original location of a memorized object, even though this location is not relevant for the response. Furthermore, EEG measures have shown a distinct N2pc-like component (associated with the orienting of selective attention) contralateral to the original location of the memorized object.

There is preliminary data (Apel, personal communication) that an additional working memory load removes the blank screen effect in language-based orienting. If participants are given a short calculation task after presentation of the visual display (and before the linguistic input) they no longer preferentially fixate related empty locations on the blank screen. This result should be interpreted with care since the load may affect the subsequent speech comprehension rather than the construction and maintenance of a visuospatial representation. Nevertheless, it fits well with our notion that activation between associated linguistic and visual representations cascades within long-term memory, but that the actual guidance of spatial attention depends on such associations being explicitly bound to locations within working memory. More specific predictions may be derived from this. For example, under the assumption that working memory consists of specialized subsystems, the effect of working memory load will depend on the type of information that needs to be remembered. A phonological load might impair the creation of phonological labels for the visual objects in the display, and thus reduce phonologically driven eye movements. Eye movements may then still be driven by visual interactions (e.g., involving color or shape). The converse should also happen: when given an additional perceptual load task, guidance of eye movements by visual target templates may be reduced, but guidance by phonological or semantic input may remain or even be enhanced (as there is less competition from visual codes). Importantly, when given an additional spatial load, all types of guidance (whether visual, phonological, or semantic) should suffer, as all interactions depend on the limited availability of spatial indices. Thus there may be a distinction between certain types of working memory load that prevent bindings of all types of representations and working memory loads that affect only specific codes.

3.6. Timing and level of representation

The framework predicts that the crucial difference between visual world and visual search is one of timing. Differences in timing cause visual and linguistic inputs to match at different levels of representation. Note that this prediction is not specific to working memory, but is consistent with any serial or cascaded activation of representations.

In typical visual search tasks, the linguistic (i.e. the instruction as to what to look for) or mnemonic (i.e. the object to remember) input occurs prior to onset of the visual search display. The idea is that in this way observers have sufficient time to create a perceptual target template that may then guide search. Work by Wolfe, Horowitz, Kenner, Hyle, and Vasan (2004) indicates that observers can set up such a template within 200 ms when it is based on a pictorial cue, and within about 500 ms when it is based on a written single-word instruction (the longer time for linguistic cues presumably reflects the time to read and the time it takes for a visual form representation to be retrieved from long-term memory). In other words, the visual memory representations are assumed to be already in place by the time the search display appears. After display onset, a match of the input signal with this target memory then occurs quite early, at perceptual levels. Participants in visual search studies may therefore routinely rely on matches at the level of visual representations because it is visual codes which are retrieved first from exposure to the search display.

A more specific prediction then is that language will have a stronger effect on visual search the more time there is to extract semantic or linguistic codes from the display. There is some preliminary evidence from visual search that appears to support this prediction. Using very simple displays, Theeuwes and Van der Burg (2007) (see also Theeuwes, Reimann, & Mortier, 2006; Mortier et al., 2010) asked observers to look for an object of either a particular color (e.g., a red circle amongst green circles) or for a particular shape (e.g., a diamond among circles), with target types being randomly mixed within blocks. At the start of each trial, participants received valid information as to which type of target to expect. This information could be presented linguistically (i.e. the instruction "color", or "shape" would appear), or pictorially (i.e. the actual color or shape would appear). Theeuwes and Van der Burg only found performance benefits when a picture of the actual objects had been presented. There was no effect of the linguistic description. This was not because participants failed to read or remember the instructions, because linguistic cues as to *where* the target could be found (rather than what it was) were highly effective. In contrast, linguistic information has been found to be more effective when it concerns complex displays that yield slower search and hence allow for more time to develop additional codes. For example, Wolfe et al. (2004); see also Bravo, & Farid, 2009; Müller, Reimann, & Krummenacher, 2003; Paivio, & Begg, 1974; Vickery, King, & Jiang, 2005 found that although pictorial cues about the identity of the target were still most effective in generating efficient search, linguistic instructions also resulted in benefits — especially when observers were given more time to process them. Consistent with this, Theeuwes and Van der Burg (2008) also found that linguistic cues became more effective when in addition to the target, there was s salient competing distractor present in the display, slowing down search.

Finally, there is one study showing the influence of semantic information on attentional guidance in visual search. Moores, Laiti, and Chelazzi (2003) found that observers were more distracted by non-target objects when those objects were semantically related to the target. For example, when asked to look for a motorcycle, participants looked more often at a helmet than at an unrelated object in the display, as was measured by eye movements and subsequent memory accuracy for objects present in the display. Similarly, Meyer, Belke, Telling, and Humphreys (2007) found interference from objects with the same name (i.e. the picture of the animal "bat" when looking for a baseball "bat"). Again, these displays of multiple everyday objects may have been sufficiently complex to allow sufficient time for a semantic or a phonological match to occur. However, there has been no systematic comparison of different display complexities and how this interacts with linguistic information.

In the visual world paradigm, the visual displays even *precede* the presentation of the crucial spoken word. In many visual world studies participants are given a preview of the display (often 1 s) before the speech input unfolds. Moreover, the acoustic target word (e.g., "beaker") is typically not the first word that is presented in the speech stream, but occurs 1 to 2 s into the message (e.g., "Pick up the *beake*r" in Allopenna et al., 1998; "Eventually she looked at the *beaker*..." in Huettig, & McQueen, 2007). Mental representations of the search display are thus created *before* the target is specified. Crucially, this allows sufficient time for semantic and perhaps phonological representations to be created from the visual objects. Hence, the visual world paradigm enables matches to occur at these levels of representation. The data of Experiment 1 and Experiment 2 of Huettig and McQueen (2007) allow for a more precise time course analysis. In these experiments, participants listened to spoken sentences (which included a critical word) while looking at visual displays containing four spatially-distinct visual items. Even though the spoken sentences and the visual displays were identical across the experiments, eye movement behavior, both in terms of where participants looked and when they looked, was radically different across the two experiments. All that changed was the relative timing of presentation of the linguistic and visual information. When participants had time to look at a display of four visual objects from the onset of the sentences (Experiment 1), attentional shifts to phonological competitors of the critical spoken words (e.g., to a beaver on hearing *beaker*) preceded attentional shifts to shape competitors (e.g., a bobbin) and semantic competitors (e.g., a fork). With only 200 ms of preview of the same picture displays prior to onset of the critical word (Experiment 2), participants did not look preferentially at the phonological competitors, and instead made more fixations to the shape competitors and then the semantic competitors. In other words, when there was ample time to view the display (Experiment 1), it appears that object processing advanced as far as retrieval of the objects' names: There were fixations to all three types of competitor. But when there was only 200 ms of preview before the onset of the critical spoken word (Experiment 2), object processing still involved retrieval of visual and semantic features to a degree sufficient to influence eye movements, but insufficient for retrieval of the objects' names to influence behavior. Huettig and McQueen (2007) suggest that there were no preferential fixations to the phonological competitors under these conditions because, by the time an object's name could have been retrieved, the evidence in the speech signal had already indicated that that phonological competitor was not a part of the sentence.

3.7. Cognitive control and task set

If language–attention interactions are mediated by working memory, then we may expect that, given the executive functions ascribed to working memory, such interactions are subject to a substantial amount of cognitive control. There is preliminary evidence from visual search studies looking at the effects of memory. Some researchers occasionally found that observers responded *faster*, rather than slower, to the visual search target when one of the competing distractors matched the memory item (Downing, & Dodds, 2004; Woodman, & Luck, 2007; Carlisle and Woodman, this issue). Such fast responses provide evidence for the idea that observers can strategically use the information of the memory task to bias their attention away from objects they know to be irrelevant to the search task. The question then is why Soto et al. (2005) and Olivers, Humphreys, et al. (2006) failed to find such biases against the matching distractors. There are strong indications that timing may again be a crucial factor here. Recently, Han and Kim (2009) hypothesized that, by default, working memory content biases visual selection towards matching stimuli, but that given sufficient time during visual search, observers can use the information to exert control over which items can actually be ignored in the search display. In support of this, they found guidance towards memory-matching objects when participants belonged to a population of fast searchers, when search was made easy, and when search started shortly (within 150 ms) after onset of the search display. In contrast, no such guidance or even guidance away from the memory item was found when participants belonged to a population of slow searchers, search was made difficult and slow, or search only started 750 ms after search display onset.

This latter condition of the Han and Kim study is one of the very few in the visual search literature in which the search display was already present for some time before the search started. Others have used partial previews, and have come to similar conclusions. For example, Watson and Humphreys (1997); see also Olivers, Watson, & Humphreys, 1999; Olivers, & Humphreys, 2003b; Olivers, Humphreys, et al., 2006 found that visual search becomes twice as efficient (i.e. search slopes were halved) when half the number of distractors was previewed for about one second. They argued that the preview allows for the top–down inhibitory control of irrelevant search items, so that these items are excluded from search. Note that no such inhibition is expected to occur during the preview in the visual world paradigm, because there is no pre-specified target, and none of the items are a priori less relevant than others. However, it would be interesting to see what happens if the task changes in this respect. Can observers bias away from visual objects referred to by the spoken input if the task so requires? A study by Belke, Humphreys, Watson, Meyer, and Telling (2008) is suggestive here. Following up on Moores et al. (2003), they found that observers tend to orient towards semantic competitors of the target (e.g., the helmet when looking for a motor bike). When observers were given an additional working memory load, the number of saccades towards semantic competitors remained virtually the same, but the fixation times on those competitiors increased substantially. This suggests that observers may have trouble suppressing and moving away from semantic matches, in line with a role for cognitive control.

The role of the task and the control settings that this requires also become clear from some studies that failed to find effects of working memory content on visual attention (and that could not be explained by timing differences, Downing, & Dodds, 2004; Houtkamp, & Roelfsema, 2006). In these studies, observers were actually required to remember *two* objects on each trial: One object for the memory test, the other object was the visual search target, which changed from trial to trial. This means that observers had to actively maintain a search template in working memory, which is directly relevant for the next task, and to remember another object for later. This may have two consequences: 1) observers exert increased control over which of the two items is currently active. If the target template is the more important one, the memory object may be suppressed and no longer interfere with search; and 2) because the target template takes up limited-capacity working memory resources, there may be fewer left for the memory object. Its weaker representation then no longer affects search (see also Houtkamp, & Roelfsema, 2009). In support of the idea that having to remember two objects is the crucial factor, Olivers (2009) found strong attentional guidance from the memory object when it was the only item to be remembered, but no such guidance when a new search target also had to be remembered on each trial. Similarly, in a recent ERP study, Peters, Goebel, and Roelfsema (2009) failed to find effects of a remembered item on visual evoked potentials if this remembered item had to be remembered along with the search target.

There may be important implications for the visual world paradigm here. The fact that a particular task may oust other, less relevant representations from working memory would lead to the prediction that effects of such representations are smaller in visual world studies with a concurrent task, as compared to those without an explicit task (i.e. those characterized by free viewing and listening). The unconstrained nature of the latter version may be an important contributor to finding language-based effects on visual selection: If the observer has no visual target to maintain in working memory, and no explicit visual search task to conduct, there is no a priori incentive to keep the influence of irrelevant linguistic information under control. Moreover, the absence of a constrained task may free the capacity to allow irrelevant objects to become more centrally represented and thus to start affecting behavior (Lavie, Hirst, Fockert, & Viding, 2004). Conversely, in visual search the task constraints (i.e. find the target as quickly as possible) may encourage matching on the basis of early visual representations, although matches at other representational levels are possible and may be used in less time-constrained situations.

The specific task set employed may also depend on the nature of the visual stimuli in the display (e.g., objects or printed words). Huettig and McQueen (2007) showed that when pictures were replaced with *printed words* (i.e. the names of the objects) attentional shifts were made only to the phonological competitors (but not semantic or shape competitors), both when there was only 200 ms of preview (Experiment 3) and when the displays appeared at sentence onset (Experiment 4). Huettig and McQueen suggested that this was because phonological information is the most relevant for a search among printed words. In other words, the search task with printed-word displays led participants to focus attention on the possibility of phonological matches in the situation where the display consists of orthographic representations of the sound forms of words. Recently, Huettig and McQueen (under revision) further investigated this issue. In Experiment 1 the displays consisted of semantic and visual-feature competitors of critical spoken words, and two unrelated distractors. There were significant shifts in eye gaze towards the semantic but not the visual competitors. Thus participants can use semantic knowledge to direct attention on printed-word displays when phonological matches are impossible. In Experiment 2 semantic competitors were replaced with a further set of unrelated distractors but there was still no hint of preferential fixations towards the visual-feature competitors. In Experiment 3, semantically more loaded sentences were presented (so as to encourage deeper semantic processing and visual imagery) but still no shifts in overt attention to visual competitors occurred. It appears that for printed words there is no cascaded processing to visual form representations, but that task demands can sometimes cause this cascade to be switched on (e.g., see evidence that only participants who take part in an explicit perceptual categorization task prior a semantic priming experiment retrieve visual form representations during the subsequent priming task, Pecher, Zeelenberg, & Raaijmakers, 1998). These findings accord well with evidence from other cognitive domains that information processing and behavior varies adaptively depending on the nature of the environment and the goals of the cognitive agent.

Another important aspect may be whether the linguistic input is *directly relevant* to the task. For example, observers might hear “Click on the beaker”, and then have to click on the beaker with the computer mouse (e.g., Allopenna et al., 1998). Since here both the language and the visual objects are directly relevant to the observer, one might expect larger effects on visual selection than in the ‘look and listen’ version (see Salverda et al. this issue, for a similar argument). From an attention perspective however, it would be interesting to compare the language–vision interaction when observers are given a set of instructions in which either the language, either the vision, both the language and the vision, or neither are important for the task.

4. Conclusions and future directions

The working memory model provides a comprehensive framework for explaining interactions between language and visual attention. Working memory has the capabilities to accommodate and integrate different types of representation, by building situation models of linguistic input, and binding objects to space and time. Data on the effects of capacity, complexity, timing, and control are consistent with this model, although they do not provide conclusive evidence for it. It is clear that there are still many unknowns about how language affects vision. We have already addressed a few questions for the future: how does the language–vision interaction behave when displays exceed more than four objects, the supposed limit of visual working memory? How does the language–vision interaction depend on the visual complexity of the display, such as the relative salience of the visual objects compared to their surroundings? How does this interaction fare under different task demands and conditions of cognitive load?

There are many questions we did not even address. For example, what are the implications of our framework for classic models of working memory? Baddeley's model (Baddeley, & Hitch, 1974; Baddeley, 2003), for example, proposes rather distinct slave systems, especially those representing linguistic and visuospatial information. This distinction has always been made on the basis of dissociations, be it in neurological patients, or in the interference caused by different types of task. But the mere finding that language affects and interferes with visuospatial orienting argues against a view of informationally encapsulated subsystems in working memory.

Furthermore, we predict a close link to the limited-capacity spatial indexing system as assumed by others. So far visual search displays have been rather static (see e.g., Horowitz, & Wolfe, 1998; and Pinto, Olivers, & Theeuwes, 2006, Schreij & Olivers, 2009, for exceptions). What does this mean for more dynamic displays in which objects travel around, or interact with each other? Will semantic and phonological codes move along instantaneously with those objects, or at a delay? Some studies suggest that semantic codes that have been arbitrarily assigned to objects indeed do move along with that object (Hoover & Richardson, 2008; see also Richardson, & Spivey, 2000). For example, in the Hoover and Richardson study, the displays consisted of animations of various animals burrowing molehills across the screen. An animal might pop out of the molehill and tell a fact, after which it would disappear. It might continue burrowing to a different location. Observers were then asked a question to which the animal had just provided the answer. Observers tended to look at the animal's current location, suggesting that the index had indeed moved along with it. In this particular experiment, there was a maximum of four animals, and the messages conveyed by them were directly relevant to the observer's task. It remains a question for the future how semantic but also other (e.g., phonological) codes travel along with multiple objects in a display.

On a different level, we did not address how language–attention interactions are instantiated in the brain. We consider this as one of the most challenging questions that cognitive neuroscience still has to solve: How to arrive from linguistic input to perceptually-driven behavior. One interesting route may be to investigate patients suffering from simultanagnosia or Balint's syndrome. Such patients have been shown to have problems with the binding of different visual features of an object, such as color and shape (Friedman-Hill, Robertson, & Treisman, 1995). This deficient binding may extend to other types of code such as associated meaning and phonology. Our framework would also predict that people with severe reductions in visuospatial working memory capacity will show weaker visual world type effects, since the number of pointers that can link visual objects to linguistic codes is even further limited.

Another question that may be best resolved on the neurophysiological level is when a visual stimulus triggers an eye movement, and when it triggers just a covert attention shift. Is this simply a matter of activation strength in specific spatial maps, or is the capture of eye movements functionally special?

Finally, we may ask questions about the routes via which different types of representation activate each other. For example, imagine a visual world array containing a kitchen ladle, a saxophone, and some visually and semantically unrelated objects. Participants then hear a category instruction (e.g., "what is the name of the musical instrument", cf. Huettig, & Hartsuiker, 2008), after which they frequently fixate not only the saxophone, but also the kitchen ladle. A simple unidirectional spread of activation might predict that "musical instrument" activates all sorts of instrument shapes, including saxophone-like shapes, which then match up with not only the saxophone, but also the ladle. However, the working memory framework provides an interesting alternative explanation: The conceptual and visual representations of musical instruments from hearing "musical instrument" first match up with the picture of the saxophone in the display. This in turn leads to increased visual "saxophone-like" shape activation. It is this shape activation that then spreads to the visually similar shape representation of the ladle. In other words, depending on the linguistic context, shape information from one visual object may prime another, semantically unrelated, but visually related object. In psycholinguistics the spread of activation has a long tradition (e.g., Collins, & Loftus, 1975; Collins & Quillian, 1969). Within visual search, a direct spread of activation from target to competitor (or vice versa) is not such a common idea. Visual search scientists would assume a spread of activation from the instruction-based target template to any visual objects more or less matching the template (which could include a distractor), but not from any on-line created object representation to another.

No doubt we can still learn a lot from empirical studies using the visual world and visual search paradigms. In any case, whether or not the theoretical framework we have sketched here is correct, we have now reached the stage where we need sophisticated models on how exactly language affects visual orienting.

Acknowledgments

The order of the first two authors is arbitrary, both contributed equally to this paper. This research was funded by a Vidi grant 452-06-007 from NWO (Netherlands Organization for Scientific Research) to C.O.

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

Altmann, G. T. M., & 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.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye-movements and mental representation. *Cognition*, *111*, 55—71.

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583—609.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*, 25—36.

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4+/−0. *Perception*, *5*, 327—334.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417—423.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829—839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation*, *vol. 8*. (pp. 47—90)New York: Academic Press.

Baddeley, A., & Lieberman, K. (1980). Spatial working memory. In R. Nickerson (Ed.), Attention and performance, VIII, pp. 521—539.

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. *Attention, perception, & psychophysics*, *70*, 1444—1458.

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, *9*, 1—9.

Brooks, R. A. (1991). Intelligence without reason. *Proceedings of the 12th International Conference on Artificial Intelligence*. San Francisco: Morgan Kaufmann.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523—547.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*, 291—328.

Burkell, J. A., & Pylyshyn, Z. W. (1997). Searching through subsets: A test of the visual indexing hypothesis. *Spatial Vision*, *11*, 225—258.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349—354.

Cave, K. R. (1999). The FeatureGate model of visual attention. *Psychological Research*, *62*, 182—194.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345—347.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28—71.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407—428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240—248.

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. *The Behavioral and Brain Sciences*, *24*, 87—185.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*, 163—201.

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.

Dell'Acqua, R., Sessa, P., Toffanin, P., Luria, R., & Jolicoeur, P. (2009). Orienting attention to objects in visual short-term memory. *Neuropsychologia*, *48*, 419−428.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193−222.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467−473.

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

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433−458.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*, 409−436.

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

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Farah, M. J. (1985). Psychophysical evidence for shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General*, *114*(1), 91−103.

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*, 405−410.

Folk, C. L., & Remington, R. W. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 847−858.

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 & Performance*, *18*, 1030−1044.

Folk, C. L., Remington, R. W., & Wu, S. C. (2009). Additivity of abrupt onset effects supports nonspatial distraction, not the capture of spatial attention. *Attention, Perception, & Psychophysics*, *71*, 308−313.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, *269*, 853−855.

Gleitman, L., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544−569.

Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial and object representations. *Psychological Review*, *101*, 470−489.

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.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science*, *28*, 105−115.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action*. New York, NJ: Psychology Press.

Hoffman, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics*, *25*, 319−327.

Hoover, M. A., & Richardson, D. C. (2008). When Facts Go Down the Rabbit Hole: Contrasting Features and Objecthood as Indexes to Memory. *Cognition*, *108*(2), 533−542.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*, 575−577.

Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 426−442.

Houtkamp, R., & Roelfsema, P. R. (2009). Matching of visual input to only one item at any one time. *Psychological Research*, *73*, 317−326.

Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras, & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond* (pp. 187−207). New York, NY: Psychology 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., & 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., & Hartsuiker, R. J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition*, *36*, 341−360.

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., McQueen, J.M. (under revision). Language-mediated visual search is determined by the nature of the information in the visual environment. Memory & Cognition.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. M. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, *121*, 65−80.

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology*, *25*, 43−110.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 604−610.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489−1506.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* : Oxford University Press.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, *1146*, 2−22.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175−219.

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.

Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, *27*, 117−143.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32−38.

Klein, R. (1980). Does oculomotor readiness mediate cognitive control of visual attention? In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale NJ: Lawrence Erlbaum.

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*(4), 519−543.

Külpe, O. (1909). *Outlines of psychology.* London: Swan Sonnenschein.

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

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211−240.

Lavie, N., Hirst, A., Fockert, J. W. D., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339−354.

Logie, R. H. (1995). *Visuo-spatial working memory.* Hove, UK: Lawrence Erlbaum.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279−281.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, *109*, 35−54.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676−703.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657−672.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1−22.

Marcus, G. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, *37*, 243−282.

Marcus, G. (2001). *The algebraic mind.* Cambridge, MA: MIT Press.

Mayberry, M., Crocker, M. W., & Knoeferle, P. (2009). Learning to Attend: A Connectionist Model of the Coordinated Interplay of Utterance, Visual Context, and World Knowledge. *Cognitive Science*, *33*, 449−496.

McDonald, S.A. (2000). *Environmental determinants of lexical processing effort*. Unpublished doctoral dissertation, University of Edinburgh, Scotland. Retrieved December 10, 2004, from http://www.inf.ed.ac.uk/publications/thesis/online/IP000007.pdf

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99−130.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Journal of Memory and Language*, *49*, 201−213.

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

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, *35*, 494−503.

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

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*, 555−563.

Mortier, K., van Zoest, W., Meeter, M., & Theeuwes, J. (2010). Word cues affect detection but not localization responses. *Attention, Perception & Psychophysics*, *72*, 65−75.

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, *57*, 1−17.

Müller, H., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 1021−1035.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General*, *106*, 226−254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theory. In D. Besner, & G. W. Humphreys (Eds.), Basic processes in reading: Visual word recognition (pp. 264–336). Hillsadale NJ: Erlbaum.

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? The effects of memory type, display type, and search type. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1275–1291.

Olivers, C. N. L., & Humphreys, G. W. (2003a). Attentional guidance by salient feature singletons depends on intertrial contingencies. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(3), 650–657.

Olivers, C. N. L., & Humphreys, G. W. (2003b). Visual marking inhibits singleton capture. *Cognitive Psychology*, *47*, 1–42.

Olivers, C. N. L., Humphreys, G. W., & Braithwaite, J. J. (2006). The preview search task: Evidence for visual marking. *Visual Cognition*, *14*, 716–735.

Olivers, C. N. L., & Meeter, M. (2006). On the dissociation between compound and present/absent tasks in visual search: Intertrial priming is ambiguity-driven. *Visual Cognition*, *13*, 202–222.

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.

Olivers, C. N. L., Watson, D. G., & Humphreys, G. W. (1999). Visual marking of locations and feature maps: Evidence from within-dimension defined conjunctions. *The Quarterly Journal of Experimental Psychology*, *52A*, 679–715.

Paivio, A., & Begg, I. (1974). Pictures and words in visual search. *Memory & Cognition*, *2*, 515–521.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*, 1227–1268.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369–378.

Pashler, H., & Shiu, L. P. (1999). Do images involuntarily trigger search? A test of Pillsbury's hypothesis. *Psychonomic Bulletin & Review*, *6*(3), 445–448.

Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *Journal of Memory and Language*, *38*, 401–418.

Peters, J., Goebel, R., & Roelfsema, P. R. (2009). Remembered but unused: The accessory items in working memory that do not guide attention. *Journal of Cognitive Neuroscience*, 1–11.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.

Pillsbury, W. B. (1908). *Attention.* London: Swan, Sonnenschein.

Pinto, Y., Olivers, C. N. L., & Theeuwes, J. (2006). When is search for a static target efficient? *Journal of Experimental Psychology: Human Perception and Performance*, *32* (1), 59–72.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*, 23–38.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509–522.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97.

Pylyshyn, Z. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*, 127–158.

Pylyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.

Richardson, D. C., Altmann, G. T. M., Spivey, M. J., & Hoover, M. A. (2009). Much ado about eye movements to nothing. *Trends in Cognitive Sciences*, *13*, 235–236.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood squares: Looking at things that aren't there anymore. *Cognition*, *76*, 269–295.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Schreij, D., & Olivers, C. N. L. (2009). Object representations maintain attentional control settings across space and time. *Cognition*, *113*, 111–116.

Schreij, D., Owens, C., & Theeuwes, J. (2008). Abrupt onsets capture attention independent of top-down control settings. *Percept & Psychophysics*, *70*(2), 208–218.

Schreij, D., Theeuwes, J., & Olivers, C. N. L. (2010). Abrupt onsets capture attention independent of top-down control settings II: additivity is no evidence for filtering. *Attention, Perception, & Psychophysics*, *72*(3), 672–682.

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.

Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory & Cognition*, *22*(1), 1–13.

Sorensen, D. W., & Bailey, K. G. D. (2007). The world is too much: Effects of array size on the link between language comprehension and eye movements. *Visual Cognition*, *15*, 112–115.

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.

Soto, D., Humphreys, G. W., & Heinke, D. (2006). Working memory can guide pop-out search. *Vision Research*, *46*, 1010–1018.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*(11) whole issue.

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., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281–284.

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, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Spivey, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*, 447–481.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*, 557–580.

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.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, *50*, 184–193.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*, 599–606.

Theeuwes, J., & Burg, E. V. D. (2007). The role of spatial and nonspatial information in visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1335–1351.

Theeuwes, J., Olivers, C. N. L., & Chiszk, C. L. (2005). Remembering a location makes the eyes curve away. *Psychological Science*, *16*(3), 196–199.

Theeuwes, J., Reimann, B., & Mortier, K. (2006). Visual search for featural singletons: no top–down modulation, only bottom–up priming. *Visual Cognition*, *14*, 466–489.

Theeuwes, J., & Van der Burg, E. (2008). The role of cueing in attentional capture. *Visual Cognition*, *16*, 232–247.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459–478.

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.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

Vickery, T. J., King, L. -W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*, 81–92.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top–down attentional inhibition of old objects. *Psychological Review*, *104*, 90–122.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*, 1–25.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.

Wolfe, J. M. (1994). Guided Search 2.0. A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.

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

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual research. *Trends in Cognitive Sciences*, *7*(2).

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*(12), 1411–1426.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Post attentive vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 693–716.

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.

Woodman, G. F., Luck, S. J., & Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cerebral Cortex*, *17*, i118–i124.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science*, *12*, 219–224.

Yantis, S., & Johnson, D. N. (1990). Mechanisms of attentional priority. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 812–825.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 601–621.

Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: Eye movements are sensitive to overlapping semantic features, not association. *Psychonomic Bulletin & Review*, *16*(5), 869–874.

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.