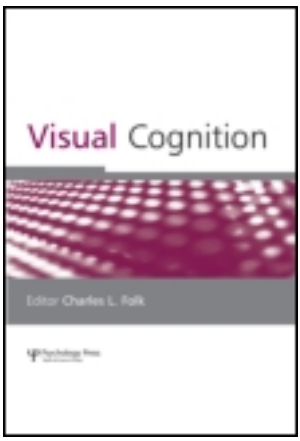
This article was downloaded by: [Adrian Staub]
On: 03 September 2012, At: 08:18
Publisher: Psychology Press
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pvis20

Linguistically guided anticipatory eye movements in scene viewing

Adrian Staub [a], Matthew Abbott [b] & Richard S. Bogartz [a]

[a] University of Massachusetts, Amherst, MA, USA
[b] University of California, San Diego, CA, USA

Version of record first published: 03 Sep 2012

To cite this article: Adrian Staub, Matthew Abbott & Richard S. Bogartz (2012): Linguistically guided anticipatory eye movements in scene viewing, Visual Cognition, 20:8, 922-946

To link to this article: http://dx.doi.org/10.1080/13506285.2012.715599

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages

whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

VISUAL COGNITION, 2012, 20 (8), 922–946

Psychology Press
Taylor & Francis Group

Linguistically guided anticipatory eye movements in scene viewing

Adrian Staub[1], Matthew Abbott[2], and Richard S. Bogartz[1]

[1]University of Massachusetts, Amherst, MA, USA
[2]University of California, San Diego, CA, USA

Downloaded by [Adrian Staub] at 08:18 03 September 2012

The present study replicated the well-known demonstration by Altmann and Kamide (1999) that listeners make linguistically guided anticipatory eye movements, but used photographs of scenes rather than clip-art arrays as the visual stimuli. When listeners heard a verb for which a particular object in a visual scene was the likely theme, they made earlier looks to this object (e.g., looks to a cake upon hearing *The boy will eat* . . .) than when they heard a control verb (*The boy will move* . . .). New data analyses assessed whether these anticipatory effects are due to a linguistic effect on the targeting of saccades (i.e., the *where* parameter of eye movement control), the duration of fixations (i.e., the *when* parameter), or both. Participants made fewer fixations before reaching the target object when the verb was selectionally restricting (e.g., *will eat*). However, verb type had no effect on the duration of individual eye fixations. These results suggest an important constraint on the linkage between spoken language processing and eye movement control: Linguistic input may influence only the decision of where to move the eyes, not the decision of when to move them.

Keywords: Eye movements; Language comprehension; Scene viewing.

A large literature has investigated eye movements in scene viewing, under various experimental conditions (see Henderson, 2003; Rayner, 2009,

Please address all correspondence to Adrian Staub, Department of Psychology, University of Massachusetts, 430 Tobin Hall, Amherst, MA 01003, USA. E-mail: astaub@psych.umass.edu

Thanks to Victoria Neilsen and Dan Petty for assistance with data collection, and to Chuck Clifton for insightful discussion. Part of this work was presented at the 24th CUNY Conference on Human Sentence Processing, Stanford University, the University of Connecticut Psycholinguistics Colloquium, and the 52nd Annual Meeting of the Psychonomic Society, Seattle, WA; thanks to audiences at these venues for helpful comments. Part of this work was carried out in completion of the second author's undergraduate honours thesis at the University of Massachusetts Amherst.

© 2012 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business
http://www.psypress.com/viscog http://dx.doi.org/10.1080/13506285.2012.715599

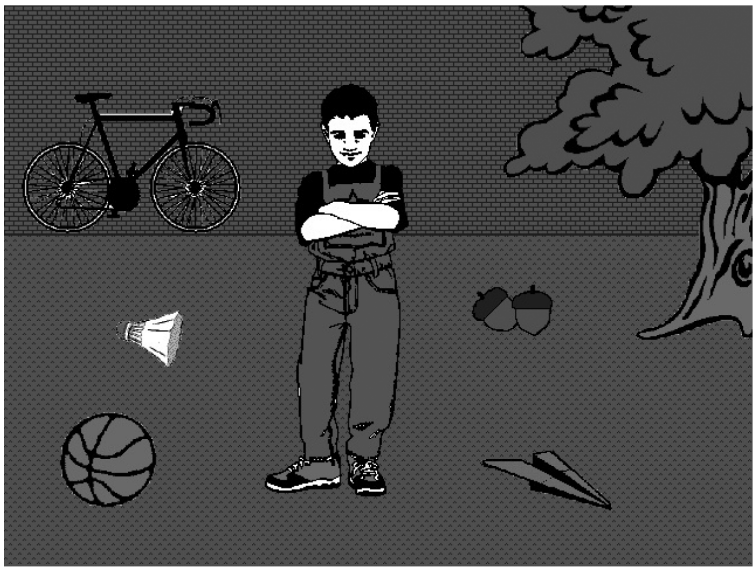
for reviews). Though some models have emphasized the role of visual properties of the scene in determining where fixations occur (e.g., Itti & Koch, 2001), it is not surprising that higher level cognitive factors also play a major role in determining fixation locations (e.g., Torralba, Oliva, Castelhano, & Henderson, 2006). For example, Henderson, Weeks, and Hollingworth (1999) found that fixations are especially likely to land on objects that are anomalous with respect to the scene in which they are located, and Henderson, Pierce, and Schandl (2009) found that an explicit search target was likely to be fixated very early in the viewer's inspection of a scene, regardless of its visual salience. Henderson et al. (2009) proposed a general notion of "cognitive relevance" to explain where in scenes the eyes tend to fixate.

The fact that the eyes very quickly fixate objects that are cognitively relevant has long been known to psycholinguists, who have used eye movement recording to address questions about the sources of information that affect spoken language comprehension, as well as the time course of these influences. Cooper (1974; see also Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) first demonstrated that listeners tend to look at pictured objects as these objects are mentioned in a spoken stimulus. Subsequently, studies using the Visual World Paradigm (VWP) have shown, for example, that listeners' gaze dynamically reflects incremental changes in lexical activation as a word unfolds (Allopenna, Magnuson, & Tanenhaus, 1998), that listeners make pragmatic inferences online about what speakers are likely to mention (Barr, 2008; Sedivy, Tanenhaus, Chambers, & Carlson, 1999), and that listeners use linguistic and real-world knowledge to anticipate how sentences are likely to continue (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003).

Until recently, research on eye movements in scene viewing and research using eye movements to investigate spoken language comprehension have proceeded without meaningful contact. In part, this may be due to the fact that in VWP experiments, the visual displays are line drawings of individual objects, cartoons, or arrays of clip-art images, rather than photographs of scenes. Indeed, Henderson and Ferreira (2004) suggested that the tightness of the linkage between eye movements and spoken language processing may be exaggerated in VWP experiments. They emphasized that these displays almost invariably consist of collections of objects or people that stand in no obvious thematic or event-based relation to each other, or to the background against which they appear. There is nothing that these displays are intrinsically about; a visual world display is not, taken as a whole, a picture *of* anything in particular. (Indeed, it is often the case that VWP displays do not accurately represent the relative size of objects; see, e.g., the acorns in Figure 1.) Why might this matter? Henderson and Ferreira suggest the following:

Downloaded by [Adrian Staub] at 08:18 03 September 2012

The boy will bounce/throw the ball.



The woman will pour/take the coffee.

Figure 1. Sample visual stimuli from (top panel) Altmann and Kamide (1999) and (bottom panel) present study. Corresponding restricting and control sentences are provided. Note that images were displayed in colour in the experiments. (Thanks to Gerry Altmann for providing the full set of experimental images.)

Downloaded by [Adrian Staub] at 08:18 03 September 2012

> [A]n array of, say, eight objects that includes a toaster does not have any kind of high-level description . . . and therefore the toaster could be anywhere. As a result, the visual system has *only* the linguistic input to guide eye movements. Indeed, with arrays, the participant is really engaged in something more like arbitrary visual search rather than true scene viewing. The implications of these points for research integrating vision and language are completely unknown at this stage. (p. 30, original italics)

In short, it is at least possible that linguistic guidance of eye movements is enhanced when the scene-internal factors that typically guide the eyes through a scene are artificially eliminated. The VWP may bias researchers towards the conclusion that there is a very tight linkage between language processing and eye movements, simply because the relevant competing pressures on the visual and cognitive systems (e.g., the need to recognize objects under varying viewing conditions, and to understand the situation in which they appear) have been removed.

Recently, however, Andersson, Ferreira, and Henderson (2011) conducted the first study investigating linguistically guided eye movements in realistic scenes. Andersson et al. found that listeners did reliably fixate mentioned objects while listening to short discourses and viewing photographs of complex real-world scenes. This effect was reduced when the speech rate was relatively fast, and when the objects were mentioned in succession, without intervening linguistic material. Still, it does appear that even when viewing realistic scenes, concurrent linguistic input plays a role in determining where people look.

The present study may be seen as building on the findings of Andersson et al. (2011), further investigating how spoken language influences eye movements when viewing realistic scenes. To motivate the present study, we first describe the classic VWP study by Altmann and Kamide (1999) that demonstrated anticipatory eye movements towards objects that were likely to be mentioned in a spoken stimulus; the present study builds directly open this previous study.

The participants in Altmann and Kamide (1999) listened to simple, transitive sentences like *The boy will eat the cake* or *The boy will move the cake*. At the same time, their eye movements were monitored while they viewed a visual stimulus consisting of a scene-like array of clip-art drawings; for the example just provided, this array included images of a boy, a cake, a toy train, a toy car, and a ball. The critical finding was that participants looked at the picture of the verb's object (the cake in the present example) sooner when the spoken sentence contained a selectionally restricting verb like *eat*, for which the cake is a particularly likely theme, than when the sentence contained a more neutral verb like *move*. Altmann and Kamide measured the latency from the onset of the critical verb in the spoken

Downloaded by [Adrian Staub] at 08:18 03 September 2012

stimulus until the beginning of the first saccade to the target object. They found that this latency was shorter when the verb was selectionally restricting than when it was not. Indeed, eye movement patterns in these conditions diverged even before the onset of the word *cake*; the probability of having initiated a saccade to the target object by the onset of the critical postverbal noun was greater in the *eat* condition than in the *move* condition. This suggests that listeners used the verb's selectional restrictions (e.g., the fact that *eat* requires an edible theme) anticipatorily, in advance of hearing any part of the critical noun itself, to guide eye movements towards the relevant object in the display. These basic patterns were obtained in two experiments with different task demands: In Experiment 1 participants were asked to judge whether the sentence corresponded to the image they were viewing, and in Experiment 2 they were told only to listen to the sentences while viewing the image. (Both experiments included filler trials in which a mentioned object was missing from the display.)

The first question addressed by the current study is simply whether similar effects are observed when the visual stimulus is a photograph of a coherent, realistic scene rather than an array of clip-art images. We regard the answer to this question as important from the perspective of the psycholinguistic literature: If it were the case that the anticipatory effects first obtained by Altmann and Kamide (1999), and since obtained in many subsequent studies (e.g., Altmann & Kamide, 2009; Kamide et al., 2003; Knoeferle & Crocker, 2006), do not emerge with more naturalistic visual stimuli, this would seriously undermine the theoretical significance of these findings.

However, the main question addressed by the present study is the following: If the standard anticipatory effect does emerge with realistic scenes as visual stimuli, how is this effect manifested in terms of the underlying eye movement control parameters that are of interest in the scene viewing literature? To motivate this question, consider two different ways the Altmann and Kamide (1999) effects might arise. One possible contribution to the latency difference between conditions is a difference in participants' *scan pattern* (e.g., Henderson, 2003) around the scene, starting approximately at the onset of the critical verb. That is, it is possible that hearing a restricting verb influences the order in which regions of the scene are fixated. Specifically, upon hearing *eat*, the eyes may reach the cake in relatively few fixations. For concreteness, we may consider the actual mean latencies to initiate a saccade to the target object in Altmann and Kamide's Experiment 2, which were 988 ms and 1246 ms from verb onset in the *eat* and *move* conditions, respectively. Given the durations of eye fixations in scene viewing (Nuthmann, Smith, Engbert, & Henderson, 2010), it is reasonable to infer that on average participants first fixated the target object approximately three to five fixations after verb onset. One way for the

Downloaded by [Adrian Staub] at 08:18 03 September 2012

observed 258 ms latency difference to come about would be for participants to saccade to the target object after approximately one fixation fewer in the *eat* condition than in the *move* condition. And because participants were more likely to saccade to the target object on an early saccade in the *eat* condition, they were more likely to saccade to the target object before noun onset in this condition (32% of trials) than in the *move* condition (18% of trials).

Another possible contribution to the latency difference, however, would be a difference in fixation durations themselves. Specifically, it is possible that hearing a restricting verb like *eat* results in a rapid search through the scene for the verb's likely theme, in which individual eye fixations are relatively short in duration. In fact, there is evidence that fixation durations are shorter when a viewer is searching for a target object. Castelhano, Mack, and Henderson (2009) compared eye movements during scene viewing under two task conditions, scene memorization and visual search, and did not find significant differences in the duration of individual fixations. However, Nuthmann et al. (2010) performed a similar study with much greater power, and found a significant, and sizable, reduction in mean fixation duration during search (232 vs. 267 ms). This was the case for both early and late fixations on a trial. Nuthmann et al. suggested that "in search, where observers look for a predetermined target object, knowledge about the nature of the target and its relationship to the scene can be used to constrain the number of potential target locations competing for selection" (p. 391). They suggested that this reduced competition for saccade target selection reduces the duration of saccade programming, thereby reducing fixation durations. The relevance of this explanation to the present issue is clear: Hearing a restricting verb such as *eat* may constrain the saccade target locations competing for selection, and may thereby reduce fixation durations prior to finding the target object.

From the perspective of eye movement control models, the contrast just proposed, between an effect on the viewer's scan pattern and an effect on fixation durations, is a contrast between the *where* and *when* decisions (e.g., Rayner, 1998, 2009). A scan pattern effect is an effect on the *where* decision: It is an effect on where saccades are directed, or equivalently, where the eyes fixate. On the other hand, a fixation duration effect is an effect on *when* saccades are initiated. The present study asks whether effects of the type observed by Altmann and Kamide (1999) arise because a spoken verb's selectional restrictions affect where saccades are directed, or whether they also (or instead) influence when saccades are initiated.

It is important to note that the psycholinguistic literature has not offered precise accounts of how eye movement control and spoken language processing are related. In most experimental work using the VWP, the focus is on how the probability that the eyes are fixating an object changes over time, and the effect of experimental manipulations on this probability

Downloaded by [Adrian Staub] at 08:18 03 September 2012

(see Tanenhaus, 2007, for a review). And though there are several implemented computational models of data from the VWP (Allopenna et al., 1998; Kukona & Tabor, 2011; Mayberry, Crocker, & Knoeferle, 2009), these models have also treated fixation probability on an object over time as the dependent measure. For example, Allopenna et al. (1998) showed that changes in lexical activation in the TRACE model of spoken word recognition (McClelland & Elman, 1986) accurately predict the probability of the eyes fixating a given object at different time points. But the VWP literature has not addressed the question of whether language-mediated changes in fixation probability come about through effects on saccade targeting, fixation duration, or both; researchers using the VWP have not made predictions at this level of detail.

Thus, we present here an experiment based closely on Altmann and Kamide (1999, Exp. 2), but with two main modifications. The first was to use photographs of realistic scenes, rather than clip-art arrays, as visual stimuli.[1] The second was to introduce a new set of dependent variables. To assess the influence of verb type on participants' scan patterns, we compared, between conditions with a restricting verb like *eat* and a nonrestricting verb like *move*, how many fixations participants required to reach the target object, starting from verb onset. To determine whether hearing a restricting verb influences fixation durations, we compared the duration of fixations between verb onset and fixation on the target, in the two conditions.

Finally, a third modification to Altmann and Kamide (1999) involved the filler trials. The 16 experimental trials in the Altmann and Kamide experiments (eight with a restricting verb like *eat*, and eight with a control verb like *move*) were paired with an equal number of filler trials, in which the direct object in the spoken sentence was not included in the array of objects. In their Experiment 1, which was a verification task, the participant was to respond "no" to the fillers, as the sentence did not match the picture; in Experiment 2, the participant made no overt response to either the experimental sentences or the fillers. Still, as Altmann and Kamide noted, participants may have interpreted the task as an implicit verification task, which may have led to faster eye movements to the postverbal object (in the experimental trials in general, and especially in the restricting condition) than would have been observed in the absence of even implicit task demands. Moreover, two other aspects of the sentences are notable, in terms of the possible role of task demands and strategic behaviour. First, in all the experimental and filler items, the sentences used simple transitive structures.

[1] This change to the content of the visual stimuli means that, unlike in Altmann and Kamide (1999), we do not include experimenter-defined "distractor" objects in each image. But as in Altmann and Kamide's experiments, the critical comparison is always between conditions (e.g., looks to the same target object in the *eat* and *move* conditions), rather than between objects.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

Second, the direct object was always a relatively predictable object for the preceding verb. Thus, it is possible that eye movements to the direct object were speeded due to participants' implicit or explicit awareness of the extreme syntactic and semantic uniformity of the spoken sentences.

To explore these possibilities, we manipulated filler type between participants. All participants were presented with the same experimental trials, but different filler types. For the *standard* filler group, the spoken sentences in the fillers resembled the experimental sentences, but were paired with images in which the direct object was not pictured, as in Altmann and Kamide (1999). For the *unpredictable* filler group, the spoken direct object was pictured, but it was a somewhat unusual object for the verb (e.g., *The lady will punch the screen.*). An attenuation of anticipatory effects with unpredictable fillers would suggest that these effects depend in part on the within-experiment predictability of the verb's theme. For the *intransitive* filler group, the filler images were the same as those used for the unpredictable filler group, but the verb phrase in the corresponding spoken sentences did not include a direct object (e.g., *The boy will play all day.*). An attenuation of anticipatory effects with intransitive fillers would suggest that these effects depend in part on the repetition of a single syntactic structure within the experiment. Finally, the *no filler* group received no filler trials at all. To the extent that anticipatory eye movements to the verb's theme reflect strategic behaviour, the absence of fillers would amplify anticipatory effects.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

METHOD

Participants

Eighty-eight undergraduate students at the University of Massachusetts Amherst participated in the experiment in exchange for course credit. All were native speakers of English, had normal or corrected-to-normal vision, and reported no reading or other language-related disorders. All participants were naïve concerning the purpose of the experiment. The participants were randomly assigned to one of four filler type groups, each with 22 members.

Materials

The experimental materials consisted of 16 sets of critical stimuli consisting of a photograph and two accompanying sentences. Photographs were either obtained from the Internet or taken by the experimenters. They were selected to be similar to the images used in Altmann and Kamide (1999), in the sense that each included a person (or more than one person) surrounded by potential referents. The size, number, and visual salience of these entities were not controlled, as each photograph was chosen to reflect a naturalistic

scene. Moreover, the location and size of the target object varied substantially across images; sometimes the target was near the centre of the image, and was the object of gaze of a person in the scene, whereas in other images the target was more peripheral. Each image was resized to 1024×768 pixels. Figure 1 shows an example image, together with an example from Altmann and Kamide. The full set of images is available from the authors.

Two sentences were recorded for each corresponding image by a female radio journalist who was unaware of the experimental hypotheses. She was instructed to speak at a normal rate with standard prosody. For each image, one of the accompanying sentences contained a verb whose selectional restrictions were judged to pick out a single object in the image as a potential theme, whereas the other sentence contained a verb that could take multiple objects as a potential theme.[2] We refer to the two conditions as the *restricting* and *control* conditions, respectively. The full set of sentences is presented in the Appendix.

Each sentence followed the format of the Altmann and Kamide (1999) items, consisting of a subject, a future tense transitive verb, and a direct object. All verbs were one or two syllables, and the verbs within each item were matched on length in syllables. (Altmann and Kamide did not match verb lengths in this way.) The waveform for each sentence was viewed in Praat (Boersma & Weenink, 2010), and each of several critical points in the audio file (verb onset, verb offset, noun onset, and noun offset) were identified. The restricting verbs were on average somewhat shorter in duration than the corresponding control verbs (417 ms vs. 456 ms) but this difference was not significant, $t(15) = 1.58$, $p = .14$. We note that the stimuli used by Altmann and Kamide had an almost identical verb duration difference, with mean durations of 383 ms and 423 ms for the restricting and control verbs, respectively. The mean duration of the subsequent determiner was also somewhat shorter for the restricting items (200 ms vs. 232 ms), $t(15) = 1.99$, $p = .07$. Unlike Altmann and Kamide, our stimuli did not include an identifiable prosodic break after the verb; Altmann and Kamide report separate durations for a postverbal break and the subsequent determiner, whereas our stimuli used a very natural prosodic contour in which no break was inserted between verb and determiner. The duration of

[2] We conducted a separate experiment that validated these intuitive judgements. In this experiment, 46 participants viewed each scene while hearing question variants of the sentences used in the present study, e.g., *What will the batter hit?* or *What will the batter see?* After each trial, participants were asked to choose which of two objects in the scene was the more likely answer to the question, e.g., the ball or the catcher. When the verb was restricting, participants chose the target object on 99.2% of trials. With a control verb, participants chose the target on 62.8% of trials.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

the following noun was similar in the two conditions (633 ms vs. 642 ms), $t(15) = 0.50$, $p = .63$. Note that the slight duration differences between the restricting and control verbs, and between the following determiners, worked against obtaining an anticipatory effect; because the restricting verbs were slightly shorter in duration than the control verbs, participants had less time to initiate a fixation on the target object before the offset of the verb.

The 16 experimental items were sorted into two lists, so that each participant heard one version of each pair of sentences, and eight in each condition. Each version of each item pair was heard by an equal number of participants in each of the four participant groups. Three different sets of 16 filler items were constructed, as described earlier. The experimental and filler items were intermixed and presented in an individually randomized order to each participant, except for the participants in the no filler condition, who heard only the 16 experimental sentences in an individually randomized order.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

Procedure

Participants were seated in front of an SR Research EyeLink 1000 eyetracker with remote desktop camera (SR Research, Toronto, Ontario, Canada), a 19-inch computer monitor, and powered speakers. The sampling rate of the tracker was 500 Hz, and the spatial resolution of the tracker, given the obtained level of calibration accuracy, was less than one degree of visual angle. Saccades were detected using the standard Eyelink remote parser settings for cognitive research: Velocity threshold 40°/s, acceleration threshold 80000°/s^2, motion threshold 0.2°. Participants were seated with their eyes approximately 70 cm from the monitor; head movements were unrestricted, although participants were encouraged to remain still throughout the experiment. Participants were instructed that they would be viewing pictures and listening to sentences and their task was simply to look and listen. Thirteen-point calibration of the eyetracker was performed before the experiment began and again after every eight trials. Calibration took approximately 30 s. Additional calibration between trials was performed as needed.

Before each trial participants fixated on a centrally located dot on the screen to allow the eyetracker to perform a drift correction if necessary, and to allow the experimenter to trigger the beginning of the next trial. Once the experimenter triggered the trial, it began with immediate onset of the visual stimulus. The first word of the auditory stimulus began shortly after the appearance of the image, with an average delay of 195 ms; the range was from 102 ms to 247 ms, due to slight variation in the amount of initial silence in the audio file and in the timing with which the experimental software was able to initiate playing the file. (We note that Altmann & Kamide, 1999,

reported simultaneous onset of visual and spoken stimuli. However, given that in both experiments the subject has the entire duration of the preverbal material, e.g., *The boy will*, to inspect the scene, this small difference in scene preview is unlikely to be functionally significant.) Trials automatically terminated after 6 s, giving participants additional time to survey the scene after the offset of the auditory stimulus. The experiment was programmed using the Experiment Builder (SR Research, Toronto, Ontario, Canada) software package. The duration of each experimental session was approximately 20 min, with somewhat shorter duration in the no filler condition.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

RESULTS

We begin by assessing whether the present experiment replicated Altmann and Kamide's (1999) basic findings: That the latency from verb onset to initial fixation on the target object is shorter in the restricting condition, and that participants are more likely to fixate this object before noun onset in the restricting condition. To anticipate the results, the experiment did replicate the basic findings, and the effects do not appear to be modulated by filler type. We then proceed to new analyses that examine whether these anticipatory effects are due to changes in participants' scan patterns, fixation durations, or both.

Trials were excluded from all analyses if any fixation during the trial had a duration greater than 2000 ms. Based on visual inspection (see Figures 4 and 5), these appeared to be outlier values (consisting of .002 of all fixations), and were most likely due to track loss or blinks that were not correctly registered by the software. This criterion eliminated 27 of 704 trials (3.8%) in the restricting condition, and 23 of 704 (3.3%) in the control condition. (The discrepancy between the proportion of fixations greater than 2000 ms and the proportion of trials containing such a fixation is due to the fact that because of the length of these fixations, no trial was likely to contain more than one, while many shorter fixations would occur on each trial.)

Each eye fixation during a trial was coded with respect to whether it fell in a rectangular region that included the target object (i.e., the verb's direct object), extending 17 pixels, or approximately one degree of visual angle, from the top, bottom, left, and right edges of the object. Data analysis included trials on which the participant was fixating the target region at verb onset, or was in a saccade to this region; in this case, the next fixation on the target was counted as the first valid fixation on the target, as this is the first fixation that could, in principle, be guided by the selectional restrictions of the verb. This procedure is identical to that used by Altmann and Kamide (1999). We note that the probability that the participant was fixating the

target during verb onset, or in a saccade to the target, did not differ between conditions (.202 in the restricting condition vs. .225 in the control condition; $p = .35$ by mixed effects logistic regression, as described later). The probability of any fixation on the target prior to or during verb onset also did not differ between conditions (.430 in the restricting condition vs. .419 in the control condition; $p = .51$). Participants were very likely to have fixated the target object between verb onset and trial completion; they did so on 95.9% of trials in the restricting condition, and 97.1% in the nonrestricting condition ($p = .26$). The latency analysis excludes the few trials on which participants did not fixate the target by trial completion.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

Latency to fixate the target

Table 1 presents the mean latency from verb onset to initiate a target fixation, by experimental condition and by filler type.[3] Several patterns in these data are notable. First, the absolute latencies are similar to those reported by Altmann and Kamide (1999). It does not appear that substituting photographs for clip-art arrays strongly affects the time from verb onset until participants fixate the verb's direct object; if anything, eye movements to the target occur somewhat earlier with photographs. These latencies are also relatively uniform across the manipulation of filler types. Second, the verb type effect is in the predicted direction overall, and for each of the individual filler types. Numerically, the verb type effect is substantially smaller with the unpredictable fillers than with the standard fillers, but is actually larger with the intransitive fillers than with the standard fillers. The no filler condition, far from increasing the size of the effect, actually reduced it somewhat. Finally, though the size of the verb type effect (105 ms) is smaller than that reported by Altmann and Kamide, in their Experiment 2 (258 ms), this difference is due entirely to faster looks to the target object in

TABLE 1
Mean latency from verb onset, in milliseconds, to initiate a fixation on target object, by verb type and by filler type

	Overall	*Standard fillers*	*Unpredictable fillers*	*Intransitive fillers*	*No fillers*
Restricting	942	900	971	917	978
Control	1047	1044	1010	1073	1060
Difference	105	144	39	156	82

[3] Note that although Altmann and Kamide (1999) reported the latency until the start of the first saccade to the target, we report the latency until the start of the first fixation on the target. These values will differ only by the duration of a saccade, i.e., approximately 30–50 ms.

the control condition of the present experiment, rather than slower looks to the target object in the restricting condition.

To assess the statistical reliability of the verb type effect, and its potential interaction with filler type, we constructed linear mixed-effects models (Baayen, Davidson, & Bates, 2008) with latency as the dependent variable, participants and items as random effects, and verb type and filler type as fixed effects. Because the distribution of latencies was extremely right-skewed, the latency variable was log transformed. Verb type was centred prior to entering it into the model. Our modelling approach was as follows, for this and subsequent analyses. We first constructed a model with only random intercepts for participants and items. We then tested, by comparison of log likelihoods, whether inclusion of random participant or item slopes for the fixed effect variable(s) was justified, choosing the larger model when the difference between log likelihoods was significant, by a chi-square statistic, at the .05 level. Except where noted, models with only random intercepts could not be rejected in favour of more complex models, and thus the results from the simplest model are reported. Importantly, except where noted, critical effects showed the same patterns of significance across all models. (In a few cases, the most complex model, with random slopes not only for fixed effects but also for their interactions, did not converge, and so could not be evaluated.) All analyses were carried out using the lme4 package (Bates, Maechler, & Bolker, 2011) for the R statistical programming environment (R Development Core Team, 2011).

The initial analysis included only verb type as a fixed effect. The effect of verb type on latency to fixate the target object was significant, $b = .18$, $SE = 0.05$, $t = 3.80$ (throughout, we treat t-values greater than 2 as significant; Baayen, 2008). The filler type factor was then added to the model. Filler type was coded using treatment contrasts, with the standard fillers as the reference level. Thus, the model intercept corresponded to the latency to fixate the target object for participants in the standard filler condition, and the main effect of verb type was evaluated for these participants; this effect was significant, $b = .22$, $SE = 0.10$, $t = 2.30$. There were no significant main effects of the filler manipulation, comparing each of the other three filler types with the standard fillers (all ts $< .5$). Moreover, none of the three interactions between verb type and filler type approached significance (ts $< .7$). The model that included filler type and its interaction with verb type did not improve model fit over the simpler model with verb type alone, based on log likelihood comparison, $\chi^2(6) = 1.28$, $p = .97$.

Figure 2 illustrates the effect of verb type on the distribution of latencies. It appears that relative to control verbs, restricting verbs increase the frequency of latencies in the range from about 250 ms to about 1000 ms (each bin is 125 ms), and decrease the frequency of latencies between 1000

Downloaded by [Adrian Staub] at 08:18 03 September 2012

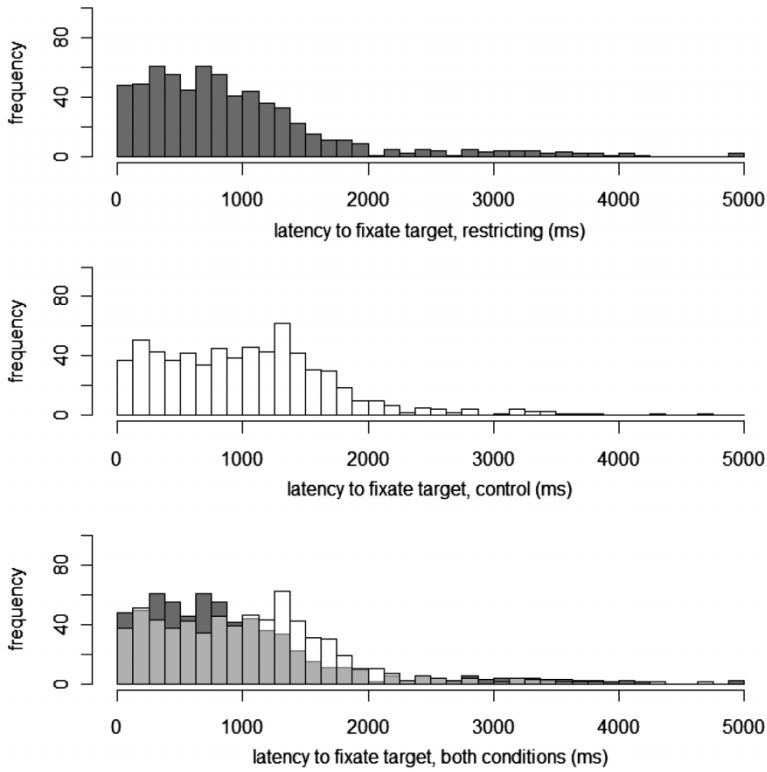


Figure 2. Histogram of latency from verb onset to fixate target object, in restricting condition (top panel), control condition (middle panel), and with both conditions superimposed (bottom panel).

and 2000 ms. Latencies of more than 2000 ms are relatively infrequent in both conditions.

A final latency analysis investigated how latency to fixate the target object may be modulated by having initiated a fixation on the target at any time before verb onset. Entering this variable into the model as a binary predictor did not reveal a significant effect of this variable or an interaction with verb type ($ts < 1$), and adding this predictor did not improve model fit over the model with only verb type as a fixed effect. Thus, it appears that the latency to fixate the target object is not modulated by having previously fixated the target object; looks to the target object are not slowed when the object was previously fixated, as would be predicted by an inhibition-of-return mechanism (Abrams & Dobkin, 1994), nor are they speeded, as would be predicted if a previous fixation is necessary to determine the identity of the objects in the scene.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

Probability of fixating the target by verb offset, noun onset, and noun offset

Table 2 presents the proportion of trials on which participants initiated a fixation on the target between the onset of the critical verb and the offset of this verb, between the onset of the verb and the onset of the subsequent noun, and between the onset of the verb and the offset of this noun. Mixed-effects logistic regression models (Jaeger, 2008) were constructed to assess verb type effects on the probability of fixating the target by each of these time points. Fixating the target between verb onset and verb offset was significantly less likely with a nonrestricting verb, $b = -.36$, $SE = 0.14$, $z = 2.66$, $p < .01$ (the parameter estimate is on the log odds scale). The same pattern held for the probability of fixating the target between verb onset and noun onset, $b = -.42$, $SE = 0.12$, $z = 3.41$, $p < .001$. For the probability of fixating the target between verb onset and noun offset, a model with random slopes for items was justified by model comparison, and in this model the verb type effect was only marginally significant, $b = -.40$, $SE = 0.22$, $z = 1.78$, $p = .07$, though it was fully significant in the other tested models.

Adding filler type in models of the probability of fixating the target by each of the critical time points yielded no significant filler type main effects (all ps $> .2$), and no significant filler type by verb type interaction effects (all ps $> .3$, with the exception of one marginal—$p = .07$—interaction in the analysis of probability of fixating by verb offset). Again, including filler type in the model was not justified by log likelihood comparison (all ps $> .25$).

Summing up, the present experiment replicated the finding that a verb's selectional restrictions affect the probability of fixating a postverbal direct object before the critical noun is heard, extending this finding to visual stimuli that are photographs of scenes. This effect actually appeared even before verb offset. Like the effect of verb type on the latency variable, the effect on the probability of an early target fixation was somewhat smaller here than in Altmann and Kamide's (1999) results. Still, this effect was highly significant. Finally, the filler type manipulation did not interact with the verb

TABLE 2
Proportion of trials on which participants initiated a fixation on the target between verb onset and each of the critical points in the linguistic stimulus, by verb type

	By verb offset	*By noun onset*	*By noun offset*
Restricting	.258	.397	.733
Control	.204	.320	.658
Difference	.054	.077	.075

Downloaded by [Adrian Staub] at 08:18 03 September 2012

type effect, either with respect to the latency of fixation on the target or with respect to the probability of fixating the target anticipatorily. In the remaining analyses, we combine across levels of the filler type variable.

Scan pattern analysis

We now turn to the question of whether the effects reported in the previous sections are due to modulation of participants' scan pattern. We begin with a point related to data analysis. Intuitively, the most straightforward method for assessing differences in the number of fixations required to reach the target would be simply to treat the number of fixations itself as a dependent variable; indeed, this type of analysis has been performed in the scene perception literature (e.g., Henderson et al., 1999). In the present experiment, participants averaged 2.77 fixations to reach the target in the restricting condition, and 3.23 fixations in the control condition, suggesting a substantial effect of verb type. However, inspection of the distribution of this variable reveals a decidedly nonnormal shape that cannot be made approximately normal by any transformation; the modal value is 1, and an extremely long tail extends to the right, with values into the teens. As a result, using number of fixations as a dependent variable seriously violates the parametric assumptions of our statistical models. (This problem is likely to be less severe when ANOVA over participant means is used to assess differences between conditions, as in Henderson et al., 1999, as this analysis takes advantage of the central limit theorem; even if the variable itself is not normally distributed, the participant means may be.)

Consequently, instead of examining the number of fixations directly, we used mixed-effects logistic regression to assess the effect of verb type on the probability that the target was reached on the first fixation after verb onset, by Fixation 2, by Fixation 3, etc. Figure 3 shows the cumulative proportions of trials on which the participants had fixated the target as a function of the number of fixations. Participants were more likely to reach the target object in exactly one fixation in the restricting condition than in the control condition (.306 vs. .258), $b = -.32$, $SE = 0.13$, $z = 2.39$, $p < .02$. The difference between conditions was greater still when considering scan patterns of two or fewer fixations (.483 vs. .410), $b = -.41$, $SE = 0.12$, $z = 3.32$, $p < .001$, and three or fewer fixations (.665 vs. .562), $b = -.56$, $SE = 0.12$, $z = 4.49$, $p < .001$. However, scan patterns of four, five, and six fixations were more common in the control condition, so in terms of cumulative proportion, the control condition began to catch up at this point. Still, in the restricting condition the target object was more likely to be reached in four or fewer fixations (.774 vs. .692), $b = -.52$, $SE = 0.13$, $z = 3.89$, $p < .001$, and five or fewer fixations (.847 vs. .791), $b = -.46$,

Downloaded by [Adrian Staub] at 08:18 03 September 2012

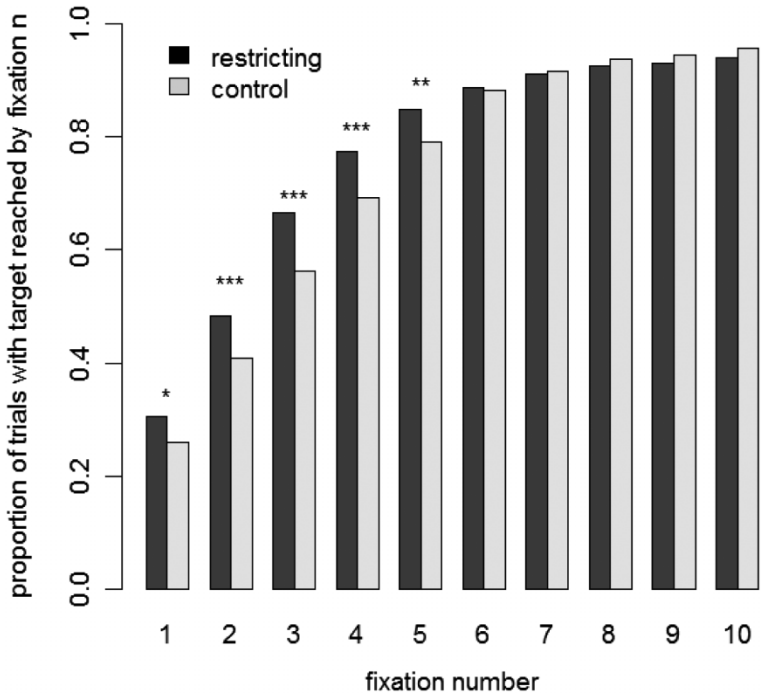


Figure 3. Cumulative proportion of trials on which the target object was reached on each successive fixation after verb onset. $*p < .05$, $**p < .01$, $***p < .001$.

$SE = 0.15$, $z = 3.01$, $p < .01$. However, by six fixations the control condition caught up completely (.888 vs. .882), $b = -.06$, $SE = 0.18$, $z = .34$, $p = .73$.

Clearly, short scan patterns were more likely in the restricting condition than in the control condition. Note the relationship between the latency data as shown in Figure 2 and the length-of-scan-pattern data as shown in Figure 3. The increased tendency for the target to be fixated in less than one second in the restricting condition corresponds to the increased likelihood of a short scan pattern prior to reaching the target.

Fixation duration analysis

We first analysed the duration of the fixation that was underway at verb onset. If the participant was in a saccade during verb onset, the subsequent fixation was treated as the relevant fixation. The mean duration of this fixation was 455 ms in the restricting condition and 468 ms in the control condition. At the outset, one important question is why these fixations are so long, relative both to previous reports of fixation durations in scene viewing and to other fixation duration measures reported later. In fact, these long

Downloaded by [Adrian Staub] at 08:18 03 September 2012

values are due to a simple, but nonintuitive, property of the analysis itself. As in any other measure, there is substantial variability in fixation durations, with these durations ranging from near 0 to almost 2000 ms (given our criterion for eliminating outliers), and with the distribution showing substantial right skew. Though values near the upper end of this range are relatively rare, the fixations that they represent will take up a disproportionate amount of total viewing time. Thus, even with the assumption that long fixation durations occur at random throughout a trial, for any experimenter-defined point p in the spoken stimulus the mean duration of the fixations that were ongoing when p occurred will be much longer than the overall mean fixation duration.[4]

For statistical analysis the duration data were log transformed. The effect of verb type did not approach significance, $b = .03$, $SE = 0.03$, $t = 0.79$. Indeed, the numerical effect of 13 ms is actually somewhat misleading, as the median duration was actually longer in the restricting condition (364 ms) than in the control condition (354 ms). Figure 4 presents a histogram of this variable for each condition. There are no evident differences between the two distributions in terms of either their location or their shape.

We then asked whether there was an effect of verb type on the fixation duration immediately prior to fixation on the target. (This analysis excludes the few trials on which the target was never fixated after verb onset.) In fact, the mean was longer in the restricting condition (335 ms) than in the control condition (325 ms), though the difference did not approach significance, $b = -.03$, $SE = .03$, $t = 0.95$. Figure 5 compares the two distributions of fixation durations; again, there is remarkable similarity.

Finally, we computed a measure of mean fixation duration for each trial by dividing the duration between verb onset and initial fixation on the target by the number of fixations during that interval. The count in the denominator included the fixation underway at verb onset, because this fixation accounted for some of the latency between verb onset and fixation on the target. Note also that saccade durations are included in the value used in the numerator. The means for the restricting and nonrestricting conditions were 315 ms and 318 ms, respectively; the difference did not approach significance, $b = .04$, $SE = .03$, $t = 1.35$.

[4] Here is a simple analogy. Assume that the probability that a runner's shoes will become untied is constant over time, i.e., shoe-untying is equally likely in each minute spent running. Some runs are very short (15 minutes) and some are very long (3 hours). An analysis of the length of the runs on which shoe-untying occurs will find that it occurs disproportionately often during long runs, and consequently, that the mean duration of the runs during which shoe-untying occurs will be much longer than the mean duration of all runs. This is simply because the long runs take up more of the total running time.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

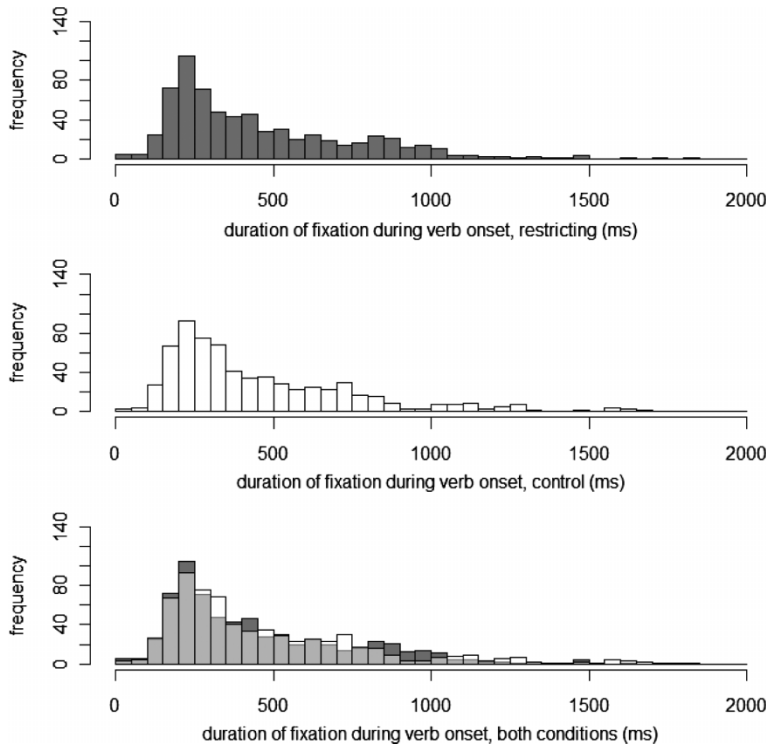


Figure 4. Histogram of the duration of the fixation underway during verb onset, in restricting condition (top panel), control condition (middle panel), and with both conditions superimposed (bottom panel).

DISCUSSION

As in Altmann and Kamide (1999), verb type affected how quickly participants fixated the postverbal direct object. When the verb selected a single object in the scene as a potential theme, this object was fixated earlier than when the verb allowed multiple objects as its theme. As in the previous experiments, these effects were truly anticipatory. There was a difference in the probability of fixating this object even before the offset of the verb itself. These effects emerged with several different types of fillers, and they emerged as subjects viewed photographs of realistic scenes, rather than clip-art arrays. Apparently, in the absence of any explicit task, a verb's selectional restrictions do result in anticipatory eye movements when viewing photographs of scenes. Considering the present result together with that of Andersson et al. (2011), it appears that in scene viewing, concurrent spoken language has a rather pervasive influence on where people look: They look at objects that are mentioned, and they look at objects that have not yet been mentioned, but are likely to be, given the constraints established by the verb.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

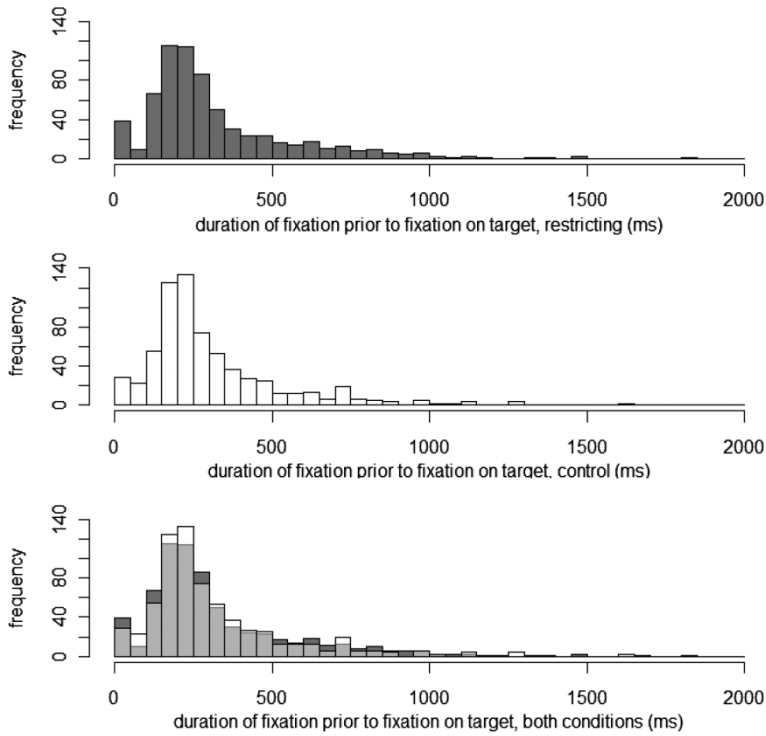


Figure 5. Histogram of the duration of the fixation prior to fixation on the target, in restricting condition (top panel), control condition (middle panel), and with both conditions superimposed (bottom panel).

Analysis of new dependent measures assessed whether the effect of verb type was an effect on scan pattern, in which the target was reached in fewer fixations in the restricting condition, or whether there was also (or instead) an effect on fixation durations. There were clear effects on scan pattern, and there was no evidence whatsoever of an effect on fixation durations. Short scan patterns were more likely when the verb was selectionally restricting; indeed, participants were more likely to reach the target on the first fixation after verb onset in the restricting condition. However, there was no evidence that participants made especially short fixations between hearing a restricting verb and reaching the target object. Neither the fixation that was underway at verb onset, nor the fixation that immediately preceded a saccade to the target, was shorter when the verb was selectionally restricting.

One obvious question is about the generality of these findings. Does spoken language quite generally influence scan patterns in scene viewing, but not the duration of fixations, or is this only the case for the verb-based anticipatory eye movements investigated here and by Altmann and Kamide

Downloaded by [Adrian Staub] at 08:18 03 September 2012

(1999)? Clearly, much more work is required to determine the answer to this question.

For the remainder of this discussion, we focus on two theoretical issues. First, we discuss implications of these results for an account of the linkage between spoken language comprehension and eye movement control. Second, we discuss these results in light of existing research on eye movements in scene viewing.

The present data suggest that although spoken language comprehension does have a very rapid influence on eye movements, its influence may be limited to one dimension of eye movement control. If spoken language influences where the eyes move, but not when, the point at which the influence of spoken language appears in the eye movement record, on a given trial, will be no earlier than the independently determined end of the fixation that is underway at the point at which a linguistic manipulation affects the language processing system. If, for example, at the critical point in the linguistic stimulus a participant is 50 ms into a fixation that would normally last 400 ms, then the effect of the linguistic manipulation will not appear in the eye movement record for 350 ms. At this point, the linguistic manipulation may result in a saccade to one location rather than another.

This conclusion may appear paradoxical, as the VWP literature contains many apparent demonstrations of a manipulation having an essentially immediate effect on eye movements, modulo some estimate of saccade preparation time. (This estimate is often given as 200 ms, though Altmann, 2011, has recently suggested that the correct value may be much smaller. Indeed, the 200 ms estimate is logically suspect, as many fixations are shorter than 200 ms in duration; see, e.g., Figure 5.) However, these very rapid effects are actually not inconsistent with the current claims. Across trials, the time remaining in the fixation that happens to be underway at the critical point in the linguistic stimulus will vary widely. On some trials, there will be a great deal of time remaining in this fixation, and so the earliest point at which the linguistic manipulation can have an effect in the eye movement record will be relatively late. On other trials, there will be very little time left in this fixation; indeed, there may be so little time that a saccade programme has already been initiated, and can no longer be cancelled. But on still other trials, the critical point in the linguistic stimulus will arrive at just the right point with respect to exerting a rapid effect on eye movements; i.e., with just enough time left in the current fixation that the target of the next saccade can still be modified. (See Altmann, 2011, for related points.) Even if this is the case on only a minority of trials, it is still the case that when many trials are aggregated, the effect of a linguistic manipulation on the probability of fixating a given object may appear very early on. Indeed, Figure 2 suggests that in the present experiment verb type began to exert an influence in the eye movement record at around 250 ms after verb onset. Evidently, this is

Downloaded by [Adrian Staub] at 08:18 03 September 2012

due to the fact that verb type influenced the destination of the first saccade after verb onset, because verb type clearly had no influence on the duration of the fixation that was underway at verb onset.

We turn now to the relation between the present results and eye movements in scene viewing. As noted in the introduction, Henderson et al. (2009) have shown that, in a visual search task, it is an object's "cognitive relevance", as opposed to its visual salience, that primarily guides scan patterns. In the Henderson et al. experiments, a search target was very likely to be fixated before other objects in a scene, even though the target object in these experiments was less visually salient than competitor objects. While Henderson et al. do not define the notion of cognitive relevance precisely, they suggest that "cognitive relevance is based on knowledge of the task, semantic knowledge about the particular scene, and current scene interpretation and understanding" (p. 854). From this perspective, in the present experiments the target object presumably achieved earlier cognitive relevance, and was therefore fixated earlier, when the verb was selectionally restricting, because hearing this verb, in the context of the particular visual scene with which it was paired, rapidly activated representations associated with the target object.

Nuthmann et al. (2010) have also demonstrated that fixation durations are shorter during visual search than during a scene memorization task. However, in the present study fixation durations were not influenced by the "search" instruction provided by a restricting verb. We think there are two possible explanations for this discrepancy. The first is that the effect of hearing a restricting verb is not actually analogous to the effect of an explicit search instruction; perhaps fixation durations in scene viewing are reduced when subjects are strategically attempting to find an explicitly identified target object, but are not reduced by the automatic shifting of attention that results from hearing a selectionally restricting verb. The second possibility is that the difference between the two studies is actually in the control condition: It is possible that explicit memorization instructions (as in Nuthmann et al., 2010) actually make fixations longer than they would be with passive viewing instructions. We cannot, at present, decide between these alternatives.[5]

CONCLUSION

The goals of the present study were twofold. The first was to determine whether the effect of a spoken verb's selectional restrictions on eye movements, first observed by Altmann and Kamide (1999), occurs when the visual

[5] In support of this second alternative, unpublished recent work in our laboratory has found a dramatic increase in individual eye fixation duration on words (with mean durations greater than 400 ms) when subjects are required to remember these words for a later recognition memory test.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

stimulus is a photograph of a realistic scene. This question was answered in the affirmative. The second was to determine whether this is an effect on the decision of where to move the eyes, the decision of when to move them, or both. In fact, the effect of a verb's selectional restrictions was due entirely to participants making fewer eye fixations before arriving at this object in the restricting condition; there was no evidence of an effect on the duration of individual eye fixations. These results suggest a constraint on accounts of the linkage between spoken language processing and eye movement control: In scene viewing, concurrent linguistic input may influence only the decision of where to move the eyes, not the decision of when to move them.

Downloaded by [Adrian Staub] at 08:18 03 September 2012

REFERENCES

Abrams, R. A., & Dobkin, R. S. (1994). Inhibition of return: Effects of attentional cuing on eye movement latencies. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 467–477.

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. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, *137*, 190–200.

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

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *73*, 247–264.

Andersson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, *137*, 208–216.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. H., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Barr, D. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*, 18–40.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package (version 0.999375-40) [Computer software]. Available from http://CRAN.R-project.org/package=lme4

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.2.37) [Computer software]. Available from http://www.praat.org/

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*, 1–15.

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.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action* (pp. 1–58). New York, NY: Psychology Press.

Henderson, J. M., Pierce, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin and Review*, *16*, 850–856.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

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.

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

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, *35*, 1009–1051.

Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, *33*, 449–496.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*, 382–405.

R Development Core Team. (2011). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.

Tanenhaus, M. K. (2007). Spoken language comprehension: Insights from eye movements. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 309–326). Oxford, UK: Oxford University Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*, 766–786.

Manuscript received December 2011
Manuscript accepted July 2012
First published online August 2012

Downloaded by [Adrian Staub] at 08:18 03 September 2012

APPENDIX

Critical sentences pairs used in experiment; restricting verb presented first.

The batter will hit/see the ball.
The girl will fill/touch the pail.
The woman will wear/lift the backpack.
The woman will pet/watch the dog.
The cook will drink/try the wine.
The lady will clean/check the carpet.
The man will peel/like the oranges.
The woman will erase/admire the chalkboard.
The lady will pour/take the coffee.
The girl will pack/push the suitcase.
The boy will eat/move the cupcakes.
The man will douse/fix the fire.
The girl will destroy/enjoy the sandcastle.
The man will answer/locate the phone.
The woman will feed/hold the cat.
The woman will sip/shift the beer.

Downloaded by [Adrian Staub] at 08:18 03 September 2012