Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 57 (2007) 502–518
Journal of Memory and Language
www.elsevier.com/locate/jml

The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing

Gerry T.M. Altmann [a,*], Yuki Kamide [b]

[a] *Department of Psychology, University of York, Heslington, York YO10 5DD, UK*
[b] *School of Psychology, University of Dundee, Dundee DD1 4HN, UK*

Received 29 June 2006; revision received 1 December 2006
Available online 29 May 2007

Abstract

Two experiments explored the representational basis for anticipatory eye movements. Participants heard '*the man will drink …*' or '*the man has drunk …*' (Experiment 1) or '*the man will drink all of …*' or '*the man has drunk all of …*' (Experiment 2). They viewed a concurrent scene depicting a full glass of beer and an empty wine glass (amongst other things). There were more saccades towards the empty wine glass in the past tensed conditions than in the future tense conditions; the converse pattern obtained for looks towards the full glass of beer. We argue that these anticipatory eye movements reflect sensitivity to objects' affordances, and develop an account of the linkage between language processing and visual attention that can account not only for looks towards named objects, but also for those cases (including anticipatory eye movements) where attention is directed towards objects that are *not* being named.

Keywords: Language-mediated eye movements; Affordances; Visual attention

Introduction

Sentence comprehension is not a passive process that projects an articulated world onto some inner mental screen. Instead, it is a process that results in active behaviours directed towards the contents of the concurrent world. These behaviours may be action-based, as when a participant is instructed to perform an action on an object, or they may be attention-based, as when a participant actively shifts his or her attention around their visual world in response to the language they hear. Indeed, it is likely that even when there is *no* action directed externally, a mental analogue of such actions (e.g., focusing on the mental representation of one discourse entity rather than another) nonetheless takes place. In these cases, attention is directed towards the contents of a mental world. One of the theoretical attractions of research on language-mediated eye movements is that it permits investigation of the interplay between the mental world and the external visual world.

Seminal studies by Cooper (1974) and by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) demonstrated how visual attention towards the external world can be mediated by the unfolding spoken language. As sentences unfold within visual contexts depicting the

* Corresponding author. Fax: +44 0 1904 433181.
E-mail address: g.altmann@psych.york.ac.uk (G.T.M. Altmann).

0749-596X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.12.004

objects mentioned in those sentences, so the eyes look towards those objects. Even as *words* unfold, the eyes begin to move towards the corresponding objects at, apparently, the earliest theoretical opportunity (cf. Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus, & Hogan, 2001). Indeed, even *before* a word unfolds, the eyes can move towards its corresponding object. Altmann and Kamide (1999) demonstrated that when a sentence such as '*the boy will eat the cake*' is heard in the context of a scene depicting a boy and a cake (and other things), the eyes begin to look towards the cake during the verb '*eat*'. Kamide, Altmann, and Haywood (2003) extended this demonstration of *anticipatory* eye movements to show that this effect was not driven solely by the relationship between the verb and whatever followed it. They presented participants with a scene depicting a man, a young girl, a motorbike, and a fairground carousel; after '*the man will ride*' the eyes looked towards the motorbike more than towards the carousel, but after '*the girl will ride*', the eyes looked towards the carousel more than towards the motorbike. Thus, the eyes looked towards whatever was most plausibly ridden by the sentential subject. Thus, anticipatory eye movements in this case did not simply reflect an association between the verb 'eat' and cakes; rather, they reflected the *combination* of the verb with its prior subject. Anticipatory eye movements have even been found when the sentence refers to the objects in a scene that is no longer present; Altmann (2004) repeated a version of the Altmann and Kamide (1999) study but removed the visual scene *before* the sentence began to unfold. Not only did the eyes look towards where the boy and the cake had been during '*boy*' and '*cake*' respectively, they also looked towards where the cake had been during '*eat*'. In the present paper, we explore the significance of these anticipatory eye movements, and ask what they tell us about the processes and representations that drive the eyes towards objects in a scene as a function of the unfolding language. We begin by considering the nature of the representations that mediate anticipatory eye movements, before progressing towards a consideration of why it is that the eyes move at all in response to sentences that unfold in the context of a concurrent visual world.

The data to-date suggest an interpretation of anticipatory eye movements in which they reflect the anticipation of what will be referred to next. Thus, after '*eat*' in '*the boy will eat...*' it will be something edible, and most likely something in the concurrent scene. Precisely *why* it will most likely be something in the concurrent scene is a topic we shall return to in the General discussion. For now, it is enough that the cake is a plausible object of the eating, given the scene, and that the eye movements towards the cake reflect the likelihood that the cake will be mentioned next. That the comprehender can anticipate upcoming information has been observed in a number of studies of reading. Schwanenflugel and LaCount (1988) proposed that highly constraining contexts increase the reader's expectation of what may follow. McDonald and Shillcock (2003) proposed that low-level contingencies between words may aid comprehension in the absence of 'higher-level' knowledge regarding the real world contingencies between, for example, the actions and entities that those words may refer to; thus, low-level statistical information about the co-occurrence of words like '*eat*' and words like '*cake*' may drive anticipatory processes (but see Frisson, Rayner, & Pickering, 2005, for an account based not on transitional probability *per se* but on a more general notion of *predictability*). Recently, a number of ERP studies have shed further light on the kinds of information that can be anticipated. For example, DeLong, Urbach, and Kutas (2005) presented participants with written sentences such as '*the day was breezy so the boy went outside to fly a kite/an airplane*'. The cloze probability for '*a kite*' was much greater than that for '*an airplane*'. ERP recordings showed that the article '*an*' elicited a *larger* N400 than the article '*a*' when, as in this case, it had a lower cloze probability. Overall, there was a significant negative correlation (over certain scalp areas) between cloze probability of the article (expected on the basis of the subsequent noun) and the size of the N400 amplitude (see Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005, for a similar result in the auditory domain). These data suggest not only that participants anticipated the noun that would follow (as one might expect from the cloze probabilities), but that they anticipated its phonological form; it was the noun's phonological form that determined whether the prenominal article would be '*a*' or '*an*'. These data reflect anticipatory processes in the absence of any concurrent visual scene, and they thus demonstrate that anticipatory processes are not restricted to language processing in concurrent visual contexts. But if a concurrent visual context does render certain post-verbal referring expressions more predictable than others (again, see below), these data suggest that anticipatory eye movements may well reflect the anticipation of their phonological form.

Given these effects of predictability, whether during reading or listening, what is the source of the information that renders something more or less predictable when language comprehension is situated in a concurrent visual context? One possibility, to be tested below, is that anticipatory eye movements reflect the fit between the unfolding language and what within the visual scene affords the objects/substances/events entailed by that language. Thus, participants may look at the cake because it is the one object in the scene that affords eating (or rather, being eaten). In other words, the language describes an act of eating, such acts require edible objects, and the fact that there is a cake present, and that

cakes are edible, makes the cake a good candidate for the required object. Thus, anticipatory eye movements are primarily driven, under this interpretation, by the anticipation of upcoming conceptual content. And this conceptual content is, of course, correlated with linguistic form. The distinction between linguistic and conceptual representation is, however, far from clear-cut. For the purposes of the present discussion (concerned with the anticipation of what may be referred to next) we can interpret linguistic representation as pertaining to form, and conceptual representation as pertaining to meaning. Of course, the boundaries are blurred, as evidenced by accounts of meaning based on the co-occurrence of form, whereby words that are similar in respect of the contexts in which they can occur will tend to have meanings that are more similar to one another than they are to the meanings of words that occur in other contexts (cf. LSA; Landauer & Dumais, 1997; and HAL; Burgess & Lund, 1997; see also Elman, 1990). However, our ability to anticipate what colorless green ideas do (i.e., sleep furiously) is due to generalization across form, and not across meaning, no matter how blurred the distinction. Thus, and notwithstanding these blurred boundaries (which become more or less blurry depending on one's theoretical pre-dispositions), we can distinguish between anticipation of upcoming linguistic form, and anticipation of upcoming conceptual content, where the latter is determined by the fit between the action denoted by the verb and the conceptual information associated with individual objects in the scene pertaining to the kinds of actions with which they may be associated. Anticipatory eye movements could, in principle, reflect upcoming linguistic form, upcoming conceptual content, or both.

At issue then, is *why* the eyes move towards the cake on hearing '*the boy will eat*', and whether an account based on the fit between the unfolding language and the affordances of the visual scene can explain anticipatory eye movements. The remainder of this paper proceeds as follows: First, we shall consider in more detail the relationship between the structure of the unfolding language and the interpretation of the concurrent visual scenes as used in our earlier studies. We shall describe a limitation of those studies which suggests a way to establish that anticipatory eye movements *do* reflect the fit between the unfolding language and what within the visual scene affords whatever objects are entailed by that language. We shall then describe two studies based on this method, before considering further not simply why the eyes move towards the cake on hearing '*the boy will eat*', but why they move *at all* in this paradigm. In developing this account, we shall focus not only on those cases where the eyes move to objects that are, or will be, referred to in the language, but on those cases also (which include the two studies described below) where the eyes move to objects that are *not* referred to in the language.

Linking visual scene interpretation to dynamical event representations

Sentences such as '*the boy will eat the cake*' are dynamically unfolding representations of *events*. The scenes that have typically been used in our corresponding visual world studies have simply been static representations of the objects that can take part in these events—they depicted *states*. Events, like sentences, unfold in time; they have a beginning and an end. Within a sentence, the combination of tense and aspect determine the beginning and end of the event which that sentence describes. But static scenes as used in our previous studies (e.g., Altmann & Kamide, 1999; Kamide et al., 2003) are, by themselves, indeterminate with respect to whether they denote the state of the world before an event unfolded, or the state of the world that resulted after the event had taken place (or an intermediate state during the unfolding of the event). Thus, when mapping sentences onto static scenes, the system must determine which part of the denoted *event structure* the scene depicts; it must determine whether the scene depicts the start state, the end state, or an intermediary state. Thus, the process of interpreting a sentence, in the context of mapping that sentence onto a scene, requires the interpretation of that scene with respect to the event structures (and their temporal properties) entailed by the sentence. Note that this is not a general requirement that applies 'across the board'—for example, studies that employ instructions to pick up an object and place it elsewhere, in the context of a set of objects laid out before the participant (whether in real space or the virtual space of a computer screen), do not require the participant to interpret the set of objects in front of them with respect to an event-related time line. Nor do studies in which the visual objects are presented on a virtual grid (cf. Huettig & Altmann, 2005) devoid of the other components that would normally make up a naturalistic scene (see Henderson & Ferreira, 2004, for discussion of scene interpretation in the context of different scene types). Nonetheless, studies in which participants are expected to map the event representation associated with an unfolding sentence onto that associated with a (semi-realistic) visual scene and the events that it affords do require this aspect of temporal interpretation (see also Knoeferle & Crocker, 2006). It is this aspect of the process that will allow us to explore the representational content that drives language-mediated anticipatory eye movements.

Altmann and Kamide (1999) presented participants with the sentence '*the boy will eat the cake*' as they viewed a scene depicting a boy and a whole cake. The cake was uneaten, and the future tensed '*will eat*' indicated that the eating had yet to happen. If the sentence had been '*the boy has eaten the cake*', the past tensed '*has eaten*' would have indicated that the eating had already happened (and was now over). A plate with a whole

cake would not be felicitous, in respect of depicting the end state, given the past tense verb morphology. A more felicitous scene might include, for example, an empty plate on which the cake (or a piece of cake) could be inferred to have been located prior to the eating. In principle, the comprehender could therefore determine whether the scene depicts the initial or the final state on the basis of tense morphology: '*will eat*' would indicate that the scene should be interpreted as the initial state, and '*has eaten*' would indicate that the scene should be interpreted as the end state. Behaviorally, '*will eat*' should result in the pattern of anticipatory eye movements described earlier (i.e., looks towards the cake), but '*has eaten*' is the more interesting case: if '*has eaten*' indicates that the scene should be interpreted as the end state, where should the eyes look? Not to the cake, as it is plainly uneaten. But if the empty plate *can* be inferred to have been the previous location of *something edible* (plates typically do function as such locations), perhaps the eyes will move towards this empty plate precisely because it affords the prior existence of something edible.

The experiments below test the hypothesis that in these past tense cases, the eyes will move towards a location that plausibly contained something that would satisfy the conceptual requirements of the main verb. We return in the discussion of these studies to their implications for the representational content that drives language-mediated anticipatory eye movements, and to the mechanism that would drive the eyes in this way.

Experiment 1

In this study, we manipulated the tense of the main verb in sentences such as (1) and (2) below:

(1) The man will drink the beer.
(2) The man has drunk the wine.

Accompanying each of these sentences was the same visual scene, which portrayed a man, a full glass of beer, an empty wine glass, some cheese, and some Christmas crackers (see Fig. 1).

The object corresponding to the empty wine glass in this example was chosen so that it violated the selectional restrictions associated with the verb (we use the term 'selectional restriction' to refer to the information associated with a verb that specifies the properties of the objects that can assume a thematic role associated with that verb—thus, a glass cannot assume the theme role associated with a verb such as 'drink').

Anticipatory eye movements following '*will drink*' should be directed towards the full glass of beer, following Altmann and Kamide (1999). Such a result is compatible with the prediction of upcoming linguistic material and with the fit between the action denoted by the verb and the affordances associated with full glasses of beer. But what of anticipatory eye movements (if any) following '*has drunk*'? Eye movements towards the empty glass would reflect sensitivity to the likelihood that the empty glass is a suitable location for whatever it

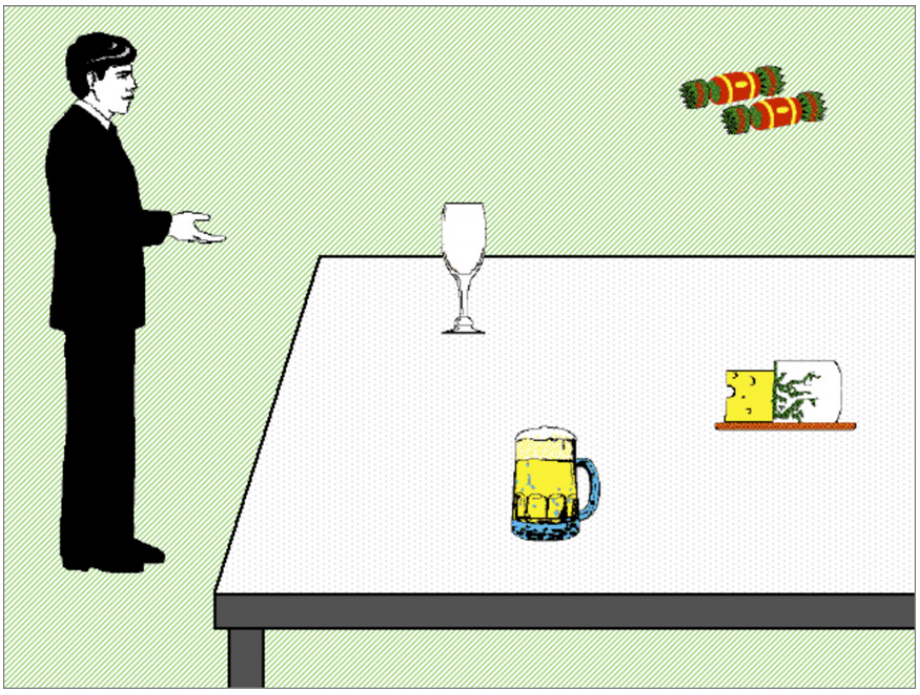

Fig. 1. Example scene from Experiments 1 and 2.

was that was drunk. We shall argue below that such anticipatory eye movements could not reflect the unfolding language *per se*.

Method

Subjects

Thirty-two subjects from the University of York student community took part in this study. They participated either for course credit or for £2.00. All were native speakers of English and either had uncorrected vision or wore soft contact lenses or spectacles.

Stimuli

Sixteen experimental pictures (see Fig. 1) were paired with two sentential conditions corresponding to (1) and (2) above. The visual scenes were created using commercially available ClipArt packages, and were constructed using a 16-color palette. They were presented on a 17″ viewing monitor at a resolution of 640×480 pixels. Each scene contained five objects: an Agent, two objects corresponding to the glass of beer and the wine glass depicted in Fig. 1, and two further distractors. The target object in the past tense condition was chosen so that it could not potentially satisfy the selectional restrictions of the verb (glasses cannot be drunk), and there was only one object in the scene that satisfied those selectional restrictions (the beer). Nonetheless, in a small number of items (6 out of the 16), the target object *could* be referred to post-verbally, as in 'the man has drunk a glass of wine' (although we note that even in this case, it is not the glass that is drunk, only the wine). We return to this issue below, in our discussion of the results of this study. For the majority of the experimental items (12 out of 16), the object corresponding to the empty wine glass in the above example was a prototypical location for the item that would be referred to post-verbally (that is, the object's *function* was to serve as such a location). For the remaining items, some other object was used to indicate such a location (e.g., an open doorway for delivering a parcel, a pile of feathers indicating the location where a cat killed a bird, and so on). See Appendix A. A further 32 sentences were added as fillers. Eight of these employed similar pictures to the experimental items but were designed as foils—for example, the scene might show a full suitcase, an empty briefcase, and a lamp, and the accompanying sentence would be either '*the man will pack the briefcase*' or '*the man has packed the suitcase*'. These were designed to avoid a situation in which the 'empty' object was always referred to following a past tense, and the 'full' object was always referred to following a future tense. A further 8 fillers employed similar pictures but were accompanied by a passive sentence (such as '*the lamp will be knocked over by the man*'). The final set of 16 fillers used a variety of different scenes each accompanied by either an active construction ('*the rabbit will eat the*cabbage') or a passive one ('*the rabbit will be eaten by the fox*').

The materials were arranged in a fixed-random order so that no successive items belonged to the same condition. Two lists of stimuli were created containing each of the 16 experimental pictures but just one version of each sentence pair.

The sentences were recorded by a male native speaker of British English (GTMA), and sampled at 44.1 kHz. The sentences were uttered with normal, but 'careful' prosody, with a mean utterance duration of 1900 ms. The sound files were presented to participants via a mono channel split to two loudspeakers positioned either side of the viewing monitor. The onsets and/or offsets of critical words in the stimulus sentences were marked using a sound editing package for later analysis.

Procedure

Participants were seated in front of a 17″ display with their eyes between 20″ and 25″ from the display. They wore an SMI EyeLink head-mounted eye-tracker, sampling at 250 Hz from the right eye (viewing was binocular). Participants were told that they would be shown some pictures that would be accompanied by a short sentence spoken over loudspeakers. In respect of their task, participants were simply told that '*we are interested in what happens when people look at these pictures while listening to sentences that describe something that might happen or might have happened in the picture*'. Although not given any explicit task, prior studies of anticipatory eye movements have demonstrated virtually identical effects whether or not a subsidiary task such as sentence verification was used (Altmann & Kamide, 1999, and see Altmann, 2004 for discussion). There were two practice trials before the main experimental block. Between each trial, participants were shown a single centrally-located dot to allow for any drift in the eye-track calibration to be corrected. This dot was then replaced by a fixation cross and participants would press a response button for the next trial. The onset of the visual stimulus preceded the onset of the spoken stimulus by 1000 ms. The trial was automatically terminated after 6 s. After every fourth trial, the eye-tracker was recalibrated using a 9-point fixation stimulus. Calibration took approximately 20 s. The entire experiment lasted approximately 25 min.

Results

Each object was colored by hand into a single color (with an error of 1 or 2 pixels beyond the object boundary, corresponding to approximately 0.1° of

visual angle), and the resulting bitmap representation of colors in the scene was subsequently equated (in software) with the coordinates output by the eye-tracker. Eye movements that landed beyond the boundaries of each object in the scene were not counted as fixations on that object. No claims are being made here in respect of the precision of the eye tracking system, nor indeed in respect of the precision of participants' eye movements. Increasing the extent of the region of interest beyond the actual pixels occupied by the object would change the absolute numbers of fixations recorded, and consequently, our empirical predictions are concerned more with changes in the bias to look towards one region or another than with absolute numbers of looks (which, given that we focus here on *anticipatory* looks, will generally be less frequent than looks towards an object engendered during the linguistic expression that refers directly to that object). For objects in or on a container, we included the container as a part of the object (e.g., the cake *and* the plate it stood on). For the purposes of analysis, we analyzed the proportion of trials on which participants were fixating either the empty wine glass or the full glass of beer (or the other objects in the scene, for the purposes of statistical analyses—see below) at two different positions within the spoken sentence; at the onset of '*will*' or '*has*', and at the onset of the determiner that preceded the sentence final word '*beer*' or '*wine*' (see Table 1). There should be no bias to be looking more towards one or other of the target objects before the onset of the verb complex, but according to the hypothesis outlined earlier, there *should* be a bias to look towards the objects, as a function of the tense of the verb, at the onset of the final referring expressions '*the beer*' and '*the wine*' (reflecting prior *anticipatory* looks towards either of these two objects). These positions were calculated on a trial-by-trial basis given the actual acoustic waveform presented to participants on a given trial. The minimum fixation duration on which basis a fixation was deemed to have occurred was 100 ms.

For the purposes of inferential analysis we used hierarchical log-linear models (see Scheepers, 2003, for discussion). Hierarchical log-linear models are an extension of chi-square, allowing for multiple categorical variables—that is, where the outcome on a particular trial might be entered into a contingency table which counts the number of times a particular outcome is observed (e.g., looking at the wine glass, the beer, the distractors) in each of the two conditions (future or past tense). ANOVA is better suited to continuous data (e.g., predicting reaction times), whereas log-linear models are better suited to predicting categorical outcomes. The logic of the log-linear analyses applied to the present data is as follows: Imagine that the data show more looks towards the empty wine glass in the 'has drunk' condition than in the 'will drink' condition, and more looks towards the full glass of beer in the 'will drink' condition than in the 'has drunk' condition. There will generally, in these data, be fewer looks to the empty wine glass than towards all the other objects added together (unlike in a reaching task where the eyes will tend to remain on the to-be-reached-for object during the reaching movement, participants in our task are free to look around the scene before, during, and after the critical referring expressions have unfolded; and given that the empty wine glass is just one of several objects in the scene that will be looked at, the sum of all looks towards the wine glass is likely to be less than the sum of all looks elsewhere in the scene). Similarly for looks to the full glass of beer. But what is at issue is whether the difference in looks to the empty wine glass (or to the full glass of beer) relative to looks to all the other objects *changes* as a function of Condition (i.e., past vs. future tense). Of course, finding such an *interaction* does not tell you whether the effect of Condition has really manifested itself on looks to the wine glass (or glass of beer) or whether it in fact has manifested on looks towards one of the *other* objects in the scene. Perhaps the effect of changing tense was, contrary to the original hypothesis, to cause more or fewer looks towards one of the other objects in the scene. The problem arises because if a participant is looking at one object, he/she cannot also be looking at another object (hence the analysis of the data in terms of a contingency table). To avoid this ambiguity (i.e., whether the effect is being carried by the wine/beer glasses or, for some unknown reason, by the distractor objects), a secondary prediction is that looks to the empty wine glass *and* to the full glass of beer (taken together), relative to looks to all the other objects, will

Table 1
Experiment 1. Proportion of trials on which the eyes were fixating the wine glass or the beer in the future and past tense conditions

Object:	Beer		Wine glass	
Condition:	will drink	has drunk	will drink	has drunk
At onset of '*will*/*has*'	.16	.13	.10	.11
At onset of '*the wine*/*beer*'	.32	.26	.16	.25

Proportions are shown at two positions within the sentence: at the onset of the auxiliary verb and at the onset of the sentence-final referring expression.

not change as a function of Condition. Finding the first interaction, but not the second, would show that looks to the empty wine glass increased *at the expense of* looks to the full glass of beer, and vice versa as a function of Condition. Note that this secondary prediction is not required by the log-linear analysis *per se*, but rather by the contingent nature of eye movements.

Participants and items were entered, separately, as factors in the computation of partial association likelihood ratio chi-squares ($LRCS_1$ and $LRCS_2$, respectively). This enables the generalizability of effects to be assessed across items and participants (see Scheepers, 2003, and see Huettig & Altmann, 2005, for discussion of how to interpret consistency across items or participants). We do not report statistical comparisons of the proportion of fixations on the empty wine glass relative to the proportion of fixations on the full glass of beer (and their equivalents across the different trials) because differences in their relative size would inevitably cause there to be unequal numbers of fixations even if the fixations were to be located at random across the scene. Thus we are primarily concerned with establishing that, for each of the two objects, looks towards those objects are mediated by the effect of the tense manipulation (this also avoids any confounds due to other aspects of the visual salience of one of these objects relative to the visual salience of the other).

At the onset of the auxiliary, there were no significant differences in fixations to the target objects as a function of Condition (all $p > .2$). However, this pattern changed significantly at the onset of '*the wine/beer*'. For fixations on the empty wine glass, there was a significant effect of Condition ($LRCS_1 = 6.5$, $df = 1$, $p < .02$; $LRCS_2 = 6.8$, $df = 1$, $p < .01$). Although the effect generalized across participants and items, the *magnitude* of the effect (that is, the size of the difference in looks between the two conditions) was not consistent across participants ($LRCS = 53.5$, $df = 31$, $p < .01$); it was, however, consistent by items ($LRCS = 12.3$, $df = 15$, $p > .6$), indicating that different participants, whilst exhibiting the effect, exhibited it to a greater or lesser extent. In contrast, the apparent difference in fixations on the full glass of beer as a function of Condition was not significant ($LRCS_1 = 2.5$, $df = 1$, $p = .11$; $LRCS_2 = 2.7$, $df = 1$, $p = .10$). The increased looks to the empty wine glass in the 'has drunk' condition were at the expense of looks to the full glass of beer and not at the expense of looks to any of the other objects ($LRCS_1 = 0.43$, $df = 1$, $p > .5$; $LRCS_2 = 0.38$, $df = 1$, $p > .5$). As a measure of effect size, we computed odds ratios (see Field, 2005, for discussion). Odds ratios indicated that fixations were 1.73 times more likely on the empty wine glass in the 'has drunk' condition than in the 'will drink' condition, and that fixations were 1.34 times more likely on the full glass of beer in the 'will drink' condition than in the 'has drunk' condition (although the latter was not statistically noteworthy).[1]

Discussion

The data demonstrate that anticipatory looks were directed towards the empty wine glass (or its equivalent across the 15 other items) as a function of the tense of the verb. At the onset of the verb complex there was no bias to look more at the empty wine glass, or the full glass of beer, in one condition rather than in the other. However, by the onset of the sentence-final referring expression ('*the wine*' or '*the beer*'), there was a bias to look significantly more often at the empty wine glass in the 'has drunk' condition than in the 'will drink' condition (and given that in both conditions, this particular example could complete '*... a glass of...*', we cannot attribute the result to the anticipation of the linguistic form '*glass*'). The corresponding bias to look more often at the full glass of beer in the 'will drink' condition than in the 'has drunk' condition was not statistically significant, and most likely reflects the fact that it was not clear whether or not the man might already have drunk some (small amount) of the beer. This same visual indeterminacy generalized across several of the items. Thus, the sentence fragment '*the man has drunk*' was as compatible with the continuation '*some of the beer*' as it was with the continuation '*the wine*'. To test this hypothesis, and to establish that the effect of tense could be found on both the empty wine glass *and* the full glass of beer (as our hypothesis predicts), we repeated Experiment 1 with new stimuli designed to avoid this indeterminacy.

Experiment 2

To avoid the indeterminacy of whether the man might have drunk some of the beer, we modified the sentences from Experiment 1 to become:

[1] The odds of looking at the wine glass in the 'has drunk' condition were calculated as the ratio of the frequency of fixations on the wine glass to the frequency of fixations on all other objects (including the full glass of beer and the background). The odds ratio was the ratio of these odds to the odds of looking at the wine glass in the 'will drink' condition. Similarly for the odds ratio associated with fixations on the full glass of beer. The odds ratio is, in effect, a standardized measure of the strength of association between categorical predictor variables and categorical outcomes in a contingency table. Hence its suitability as an index of effect size (see Field, 2005, for a simple exposition). Unlike other indices (such as Pearson's r), it is not limited between 0 and 1, but like Pearson's r, its interpretation is relatively transparent.

(3) The man will drink all of the beer.
(4) The man has drunk all of the wine.

Accompanying each of these sentences was the same visual scene used in Experiment 1 (see Fig. 1). However, in six other cases, we had to modify the visual scenes to accommodate the change in the sentences (we elected to use '*all of*' in all our experimental items, not just those exhibiting the indeterminacy). Thus, and to take just one example, the sentence pair '*the policeman has discharged the woman*' (supposedly engendering looks to an empty police cell) and '*the policeman will discharge the man*' (engendering looks to a neighbouring police cell with a single man in it) became '*the policeman has discharged all of the women*' and '*the policeman will discharge all of the men*'. We thus had to ensure that the police cell now held more than one man. Broadly speaking, though, the accompanying visual stimuli were largely the same as those used in Experiment 1.

Method

Subjects

Forty-two subjects from the University of Manchester student community took part in this study. They participated either for course credit or for £2.50. All were native speakers of English and either had uncorrected vision or wore soft contact lenses or spectacles.

Stimuli and procedure

The stimuli and procedure were identical to those employed in Experiment 1, except for the changes detailed above (see Appendix B). The mean utterance duration was 2700 ms. Eye tracking was performed in this study with an SR Research EyeLink II eye-tracker sampling at 250 Hz.

Results

At the onset of the auxiliary, there were no significant differences in fixations as a function of Condition (all $p > .4$) [see Table 2]. At the onset of '*the wine/beer*', there was a significant effect of Condition on fixations on the empty wine glass, ($LRCS_1 = 6.0$, $df = 1$, $p < .02$; $LRCS_2 = 5.7$, $df = 1$, $p < .02$). Although the effect generalized across participants and items, the magnitude of the effect was not consistent across items ($LRCS = 27.2$, $df = 15$, $p < .03$); it was, however, consistent by participants ($LRCS = 43.5$, $df = 41$, $p > .3$), indicating that different items, whilst engendering the effect, engendered it to a greater or lesser extent. The difference in fixations on the full glass of beer as a function of Condition was also significant ($LRCS_1 = 7.4$, $df = 1$, $p < .007$; $LRCS_2 = 7.2$, $df = 1$, $p < .008$). The magnitude of the effect was not consistent across participants ($LRCS = 57.9$, $df = 41$, $p = .04$); it was, however, consistent by items ($LRCS = 11.7$, $df = 15$, $p > .7$). The increased looks to the empty wine glass in the 'has drunk' condition were at the expense of looks to the full glass of beer, and conversely for increased looks to the full glass of beer in the 'will drink' condition ($LRCS_1 = 0.13$, $df = 1$, $p > .7$; $LRCS_2 = 0.13$, $df = 1$, $p > .7$). Odds ratios indicated that fixations were 1.56 times more likely on the empty wine glass in the 'has drunk' condition than in the 'will drink' condition, and that fixations were 1.58 times more likely on the full glass of beer in the 'will drink' condition than in the 'has drunk' condition.

Discussion

The data demonstrate that anticipatory looks were directed towards the empty wine glass or the full glass of beer as a function of the tense of the verb. At the onset of the verb complex there was no bias to look more at the empty wine glass, or the full glass of beer, in one condition rather than in the other. At the onset of the sentence-final referring expression, participants were significantly more likely to fixate the empty wine glass after '*has drunk all of*' than after '*will drink all of*'. Conversely, they were significantly more likely to fixate the full glass of beer after '*will drink all of*' than after '*has drunk all of*'. Evidently, the inclusion of '*all of*' relative to Experiment 1 had the desired effect in respect of engendering significant effects of tense on looks to empty wine glass *and* on looks to the full glass of beer. Although we would argue that this is because the inclusion of '*all of*' rules out the possibility

Table 2
Experiment 2. Proportion of trials on which the eyes were fixating the wine glass or the beer in the future and past tense conditions

Object:	Beer		Wine glass	
Condition:	will drink	has drunk	will drink	has drunk
At onset of '*will/has*'	.09	.10	.10	.11
At onset of '*the wine/beer*'	.30	.21	.18	.25

Proportions are shown at two positions within the sentence: at the onset of the auxiliary verb and at the onset of the sentence-final referring expression.

that the man 'has drunk some of the beer', it is conceivable that the difference between the two experiments is also due to the fact that participants had more time between verb offset and the onset of the '*the wine/beer*' in which to direct their eyes to the appropriate object. Notwithstanding this possibility, Experiment 2 demonstrates the expected effect of tense on looks towards the empty wine glass or the full glass of beer.

The increased bias to look towards the empty glass in the past tense condition, and to do so even before the post-verbal referring expression was encountered, has implications for the nature of the information that anticipatory eye movements reflect. As detailed earlier, previous studies (e.g., Altmann & Kamide, 1999; Kamide et al., 2003) were compatible with the idea that anticipatory eye movements reflect what, given the current input, is likely to be referred to next. But in the two studies reported above, the anticipatory eye movements towards the empty wine glass reflect more than just what will be referred to next; the wine glass (and its equivalent across the items) violated the selectional restrictions of the verb, and thus it could not have been looked at because it would be referred to next (but to the extent that in a small number of trials it *could* be referred to next, as in '... *drink/drunk the glass of*...', this could not explain the pattern of results across the 'will drink/has drunk' conditions). Instead, it was looked at because the system presumably anticipated after '*drunk*' that whatever would be referred to next would be something that (a) was drinkable (thus satisfying the selectional restrictions of the verb) and (b) had been, but was no longer, in the depicted world (thus satisfying the tense morphology of the verb). At this point, knowledge of real world contingencies would have identified that whatever was drunk would have been associated with some probability with the empty wine glass—the equivalent of a semantic priming effect (based on the real world contingency between events of drinking and subsequently empty glasses). We return to this effect below, in the General discussion. In essence, and somewhat paradoxically, the glass was looked at precisely because it did *not* contain the very thing that the verb '*drink*' selects for.

Anticipatory eye movements do more, then, than reflect what will be referred to next; they reflect the mapping of information regarding what may be referred to next onto the *affordances* of the objects in the scene. According to Gibson (1977), the affordances of an object are the actions which an object permits its viewer. Unlike Gibson, we and others (e.g., Glenberg, 1997; Kaschak & Glenberg, 2000) use the term 'affordance' to refer to knowledge based on our experience of how we interact with an object, and how that object interacts with other objects (this experience can, of course, include experience of form as well as function)—it is knowledge, in effect, of what kinds of events the object can participate in. For example, the affordances of a man-made artefact include knowledge of the uses to which that object can be put, how it is put to that use, and what the consequences of that use might be. In using the term 'affordance', we do not necessarily subscribe to Gibson's view that an object is perceived in terms of its affordances rather than in terms of its abstract physical properties, or that the affordances of an object are perceived on the basis of invariant cues in the visual stream. We simply use the term to refer to what constitutes the conceptual knowledge of an object. Thus, our knowledge of a glass–its affordances—include the fact that it can be drunk from. However, the knowledge deployed by the participants in our experiments was more subtle than just this; they did not look at the empty glass simply because it could be drunk from; if they had, they would have looked as often towards it in response to '*the man will drink*' as in response to '*the man has drunk*', but they did not. They looked at the wine glass because it was *empty* — the affordances of the *empty*wine glass attracted the eye movements, reflecting our experience of what empty glasses signify, in terms of how they become empty and what they may have contained.

In conclusion, Experiments 1 and 2 showed that participants anticipated at the verb that something would be referred to that had been drunk; the eyes were then drawn to those objects in the visual scene whose affordances were compatible with the conceptual specification associated with this anticipated reference. The fact that these affordances implicitly encoded a past state (when the glass contained something) constitutes an implicit interpretation of the scene as reflecting the state of the world after the event described in the sentence (the drinking event) had taken place. There are thus two aspects of these data that are noteworthy: first, that anticipatory eye movements reflect the mapping of linguistically derived information onto the affordances of the objects within the scene; and second, that the concurrent scene is dynamically interpreted with respect to the event structure conveyed by the unfolding sentence (for example, as reflecting the initial or end state). With respect to the latter observation, it appears that the scene is interpreted, at the earliest possible moment, in respect of its *temporal* relationship to the event denoted by the unfolding sentence; hearing '*the man will*...' or '*the man has*...' causes the scene to be located in time relative to the propositional content of the unfolding sentence (that is, as reflecting a state of the world at a time before, or a time after, the described event).

The fact that the scene is interpreted at all requires the existence of some representation of the visual scene that is abstracted from the image of that scene. Proponents of 'situated vision' (e.g., Ballard, Hayhoe, Pook, & Rao, 1997; O'Regan, 1992), suggest that the visual system uses the scene itself as a kind of external memory,

using oculomotor coordinates (defined relative to the configuration of cues within the scene) as pointers towards this external memory (see also Richardson & Spivey, 2000). However, whilst the scene can support the external memory of the objects that are, or were, present in that scene, some more abstract level of representation is required to encode the interpretation of those objects in respect of how they inter-relate, and how they participate in whatever event might be denoted by a concurrent sentence (on the assumption that those roles are not directly represented within the scene). Similarly, the affordances of an object are not 'out there', but are experientially-based encodings requiring a representational substrate that encodes information that goes beyond that conveyed by the scene itself. Thus, anticipatory eye movements do not reflect the mapping of language onto the scene itself, but rather, onto some dynamically interpretable representation of that scene and the affordances it contains.

This last observation goes further than previous demonstrations of the influence of affordances on language-mediated eye movements (e.g., Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers, Tanenhaus, & Magnuson, 2004; see also Kamide et al., 2003). For example, Chambers et al. (2004) showed that language-mediated eye movements during reference resolution are guided by the behavioral goals of the listener and the affordances of the objects laid out in front of the listener. Participants were either shown two eggs that were pourable (i.e., out of the shell, raw, in glass containers) or just one that was pourable (there were still two eggs, but the other was solid). The referring expression '*the egg*' in the sequence '*pour the egg*' was interpreted as indeterminate (i.e., not permitting a unique interpretation of which egg was intended) when both eggs were pourable, but as determinate (i.e., permitting a unique interpretation of which egg was intended) when only one was pourable. Similarly, when the listener was presented with two whistles, one of which had attached to it a loop of string, and was told '*put the whistle*…', the expression '*the whistle*' was interpreted as determinate when the listener held a hook with which to do the putting, but indeterminate otherwise. These cases show very elegantly that the behavioral goals of the listener, and the affordances of the objects within the visual world, can restrict the domain of reference. Both cases can be interpreted as reflecting a process by which the eyes are directed to whatever in the concurrent scene is named—the affordances of the objects restricted the set of objects in the scene that matched the target name to a subset of such objects (in fact, to just one) whose affordances matched the behavioral goals of the listener. In the cases we reported here in Experiments 1 and 2, the affordances of the objects in the scene did not simply circumscribe the domain of reference; they caused looks to be directed towards objects that were *neither* named *nor* referred to. Below, we consider the implications of such demonstrations for theories of why, and how, language mediates eye movements. We shall argue that prior formulations of the 'linking hypothesis' between language and eye movements (e.g., Tanenhaus, Magnuson, Dahan, & Chambers, 2000) require modification in light of such data.

General discussion

The fact that participants direct their visual attention towards objects in a concurrent scene whose names they hear is not particularly surprising. After all, one of the purposes of language is precisely to direct visual attention. More interesting (and perhaps more constraining, theoretically) are the conditions under which language-mediated eye movements occur towards objects that are *not* referred to in the language. Our own data, presented here, are just one example of how visual attention is not necessarily tied to direct (or anticipated) reference. To these data can be added demonstrations by Dahan and Tanenhaus (2005) and Huettig and Altmann (2004, 2005, in press) showing that the eyes will move rapidly towards objects whose visual forms are related to the objects named in the language, or whose conceptual properties overlap with those of the objects named in the language. Thus, on hearing '*snake*' the eyes will move to an electric cable, even when that cable is visually unambiguous and has been in visual field for several seconds. Similarly, on hearing '*piano*' the eyes will look towards a trumpet (cf. Yee & Sedivy, 2006). In the latter case, the 'semantic distance' between pianos and trumpets predicted the probability of directing an eye movement to one on hearing the other (where semantic distance could be defined either within a multi-dimensional space derived from association norms, or within a multi-dimensional space derived from co-occurrence statistics; Huettig & Altmann, 2005; Huettig, Quinlan, McDonald, & Altmann, 2006). In a similar vein, Myung, Blumstein, and Sedivy (2006) showed that '*piano*' engenders more looks towards a typewriter than towards a couch—an effect mediated by featural overlap between pianos and typewriters in respect of manipulation features. Taken together, these data offer an account in which *conceptual overlap* (cf. proximity in multi-dimensional space) mediates the direction of visual attention by language. They suggest that formulations of the linkage between language and eye movements need to take account of this mediating influence. Indeed, one of the earliest accounts of such linkage was limited (because of the data available at that time) to explaining the relationship between the lexical activation of the name of a picture and saccadic eye movements towards the corresponding picture (e.g., Allopenna et al., 1998; Tanenhaus et al., 2000). In the remainder of this discussion

we shall formulate an extension of this linking hypothesis that goes beyond the name of the picture (so as to explain language-mediation of eye movements to objects that are not being referred to). We shall develop the account to explain *why* the linguistically-mediated activation of conceptual information determines the probability that a participant will shift their visual attention.

In formulating a hypothesis that links language processing to eye movements in the studies discussed thus far, we shall start by considering that in most cases, the visual objects have been present *before* the occurrence in the language of the 'trigger' to move the eyes. Thus, the conceptual representations activated through the interpretation of the scene pre-date the arrival of the language. When a sequence of words is encountered, the representations that *they* engender can then make contact with the pre-existing representations already 'in place' due to the interpretation of the concurrent (or previous) visual scene. To explain how this process of making contact can direct the eyes back towards the objects in the scene, we must make a number of assumptions. We shall illustrate these using the '*piano*'-trumpet finding of Huettig and Altmann (2005)—our previous data on anticipatory eye movements, as well as the data from the two studies above, add further constraints on this process to which we return below. First, we assume that the conceptual representations engendered by words *and* objects are not indivisible wholes, but are featural in composition (cf. McRae, de Sa, & Seidenberg, 1997; Rogers & McClelland, 2004). Thus, seeing a trumpet activates a featural representation that includes information about form (presumably, both actual as evidenced by the scene, and abstract as associated with the prototype), function, color, mode of physical interaction, sound, associations and contextual dependencies (e.g., marching bands, Satchmo, the school concert, Sketches of Spain), and so on. We follow McRae, Ferretti, and Amyote (1997) and Rogers and McClelland (2004) in assuming an experiential basis to these features (that is, we assume that 'features' are emergent abstractions across multiple experiences of each object). Hearing the word '*piano*' similarly activates a composite featural representation. A second assumption we make concerns the consequences of overlap between the featural specification associated with '*piano*', and the featural representation associated with the previously (or concurrently) seen trumpet. We assume here that featural overlap results in an activation 'boost'—that is, the activation of conceptual features by '*piano*' that are common to the conceptual representation already activated by the trumpet causes that representation (of the trumpet) to itself increase in activation. We view this increase in activation as akin to a form of re-activation of the episodic trace associated with the perceptual experience of the (previously or concurrently) viewed object. 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). In our '*piano*'-trumpet example, the episodic trace associated with the trumpet will increase in activation more than the traces associated with the other objects in the scene (unless they too overlap significantly with the conceptual representation activated through hearing '*piano*'). Moreover, we assume that such changes in activation constitute changes in the attentional state of the cognitive system; in effect, we are proposing that such changes constitute a shift in attention (although whether they *constitute* a shift in attention, or *cause* a shift in attention, is not central to the account). A final assumption is that this shift in attention will, other attentional factors being equal, increase the likelihood of a saccadic eye movement towards the spatial location associated with that episodic trace. Why the eyes move when the episodic trace associated with an object increases in activation is beyond the remit of this paper. One simplistic explanation is based on Hebbian learning: Orienting towards an object increases the activation of the mental representation of that object and its encoded location. Successive pairings of this kind during development may result in the opposite pattern, with increases in the activation of a mental representation of an object and its location resulting in the increased likelihood of an orientation response towards that location. The idea that object representations mediate shifts in visual attention is shared with other object-based accounts of visual attention (e.g., Christ & Abrams, 2006; Law & Abrams, 2002), and shares many features with the Theory of Event Coding proposed by Hommel, Müsseler, Aschersleben, and Prinz (2001) (and see the Richardson & Spivey, 2001 response to this target article).

With these assumptions in place, the '*piano*'-trumpet data of Huettig & Altmann (2005) are explained in terms of featural overlap between the pre-existing representation of the trumpet (and its location) and the language-induced conceptual (featural) representation of pianos. The greater the overlap (i.e., the smaller the featural distance), the greater the likelihood of a saccade back towards the trumpet. Similarly, the effects described by Dahan & Tanenhaus (2005) and Huettig & Altmann (2004, in press), whereby '*snake*' engenders looks to an electric cable, can be explained in terms of the featural (form) overlap between snakes and cables, while the effects described by Myung et al. (2006) in respect of looks to a typewriter on hearing '*piano*' can be explained in terms of the featural (motoric) overlap between pianos and typewriters. The account also explains why, even in the absence of a concurrent scene, the eyes move towards where named objects *had been* (Altmann, 2004). In such cases, as in the concur-

rent case, it is the *episodic trace* of an object that drives the eye movements towards a particular spatial location, not the object (or rather its retinal projection) itself.

The account does not in fact rely on *pre-existing* episodic traces. For example, if the language *preceded* the visual scene, the word '*piano*' would first activate a conceptual representation of pianos. On subsequently encountering the trumpet-object in the visual scene, a representation of this perceptual experience would become activated. The overlap with the language-induced conceptual representation of pianos would ensure that the activation of the trumpet-induced conceptual representation would be higher than it would otherwise be if '*piano*' had not been heard previously (and conversely, the language-induced representation of pianos would receive a boost because of this overlap also). Thus, the trumpet would still be attentionally more salient even though in this case, the language preceded the visual scene. A study by Moores, Laiti, and Chelazzi (2003) is relevant to this case: in a visual search paradigm, they presented participants with a visual word ('key') for 1000 ms which was replaced by a fixation cross for 800 ms. Then, an array of four objects, including (in the condition of interest) a lock and three unrelated distractors, appeared for 1000 ms. The participants' task was to report whether the previous (visually) named object was present in the array. 23% of initial saccades were directed towards the lock, compared with 17% towards each of the unrelated distractors (pre-testing had ensured that each object could be resolved at the presentation eccentricity used in this study). Thus, the object (lock) that was related to the target word ('key') appears to have been more salient than the unrelated distractors even in this case where the linguistic target preceded the visual objects.

The situation becomes more complex, however, when the linguistic 'triggers' that cause the eye movements are not nouns, as in the case of looks towards named objects or towards semantic or visual competitors (the '*piano*'-trumpet and '*snake*'-cable cases). Verbs also drive eye movements (and more specifically, anticipatory eye movements), and they can do so whether alone ('*the boy will eat the cake*' engendering looks towards the one edible object at '*eat*'; Altmann & Kamide, 1999), whether with their subject ('*the girl will ride…*' engendering looks towards whatever is most plausibly ridden by the girl; Kamide et al., 2003), or whether with their auxiliary tense marker ('*the man will drink/has drunk…*' engendering looks towards whatever in the scene affords wine in the future or, alternatively, in the past; see above). Consideration of these cases (taken all together) leads naturally to the idea that the conceptual fit between words in the language and objects in the visual scene can occur at the level of affordances (i.e., properties of the possible *interactions* those objects could have engaged in, do engage in, or will engage in). Thus, it is the fit between the cake as a potential object of eating, and the action of eating denoted by '*eat*', that drives the eyes to the cake at '*eat*'. Similarly, it is the fit between the carousel as a child-centred object of riding, and the action of riding denoted by '*ride*' in the context of the sentential subject '*the girl*', that drives the eyes to the carousel after '*ride*'. And finally, it is the fit between an action of drinking that took place in the past, and the affordances of an empty glass (affording something drinkable in the past) that drives the eyes towards the empty wine glass after '*the man has drunk*'.

These last effects, based on the fit (or rather, overlap) between the affordances of the objects in the scene and the conceptual representations activated by the language, are not restricted to verbs. Kamide et al. (2003) reported a study in Japanese showing that anticipatory eye movements can be made on the basis of the ordering of (what are in effect) case-marked nouns in Japanese (a verb-final language). In that case, the case-marking determined the roles of the depicted participants—thus, after the sequence (glossed here in English) '*waitress-nominative customer-dative*' the role of the customer as a recipient and the waitress as the agent suggests an event of transfer that entails an object being transferred from one to the other; hence the observed eye movements towards the object in the scene (a hamburger) that plausibly afforded such transference. These effects, and the earlier case of looks toward what is most plausibly ridden by the man or the girl after '*the man/girl will ride*' suggest that the conceptual representations activated by the language which make contact with the affordances of the objects in the scene are event-based representations which reflect not only single lexical items, but their *combination* also.

Thus far, our discussion has focused on language-mediated eye movements towards objects that have not been named (or have not *yet* been named) in the language. In principle, an affordance-based account of conceptual overlap (and the consequent activation boost to existing episodic traces) is not restricted to only these cases. Even in the conceptually simpler cases where the eyes move to named objects as the name unfolds (cf. Allopenna et al., 1998), the same affordance-based account holds; participants may look at the beaker on hearing '*beaker*' not because of the knowledge that the depicted beaker is called a 'beaker' (i.e., the phonological code associated with the conceptual representation activated by that depicted object), but because of the overlap between conceptual knowledge of the affordances of the beaker and the conceptual representation activated by the unfolding word. Thus, the phonological form of the name of the object need not in fact mediate the mapping between the conceptual representations activated by the language and the conceptual representations activated by the scene

(cf. Dahan & Tanenhaus, 2005); it certainly mediates the activation, by the language, of those conceptual representations (cf. its role in lexical access), but once they are activated (or in lexical access terms, 'accessed'), a fit against the affordances of the objects in the scene can be determined without recourse to any information about how those objects might normally be named. Thus, the effects showing a very fine synchronisation between acoustic-phonetic input and eye movements towards named objects (e.g., Allopenna et al., 1998; Dahan & Tanenhaus, 2004; Dahan et al., 2001; McMurray, Tanenhaus, & Aslin, 2002; Salverda, Dahan, & McQueen, 2003) necessarily reflect graded effects of phonetic variation on lexical access (that is, on the activation of conceptual knowledge about an object's form, function, and other properties); but beyond that, they do not *necessarily* reflect phonetic, or phonological, mediation of the mapping between language and vision. We are *not* claiming, however, that phonological information associated with knowledge about visual objects is not used (such information must be available because otherwise we could not name objects); phonological information may indeed constitute a part of the informational overlap between the representations activated by the language and the representations activated by the visual objects. All we claim is that phonological information is not a *necessary* component of that overlap. However, the notion of overlap is crucial, and relies on an approach to knowledge in which information (whether activated through language or through vision) is encoded as a distributed representation within a multi-dimensional space (cf. Cree & McRae, 2003; McRae, de Sa, et al., 1997; Rogers & McClelland, 2004; Tyler & Moss, 2001; Vigliocco, Vinson, Lewis, & Garrett, 2004; and see also Rumelhart & McClelland, 1986). We believe the conceptual information that is necessary to drive eye movements in many of the cases described above is abstracted knowledge about the *actions* or *events* that an object can take part in, and the role(s) it plays when taking part in those events. Such affordances fit as well within a multi-dimensional space account of conceptual knowledge as do any other kinds of knowledge.

In summary, we have presented an account of the linkage between language processing and saccadic eye movements that extends the framework originally proposed by Tanenhaus et al. (2000); this 'linking hypothesis' no longer relies on an object's name as the trigger to move the eyes, and no longer relies on the lexical activation of individual items. Instead, it is based on the activation of representations by the language (whether due to individual lexical items or to their combination) that overlap with those activated by the concurrent, or previous, scene (and as such is compatible with the view that the representational 'products' of language comprehension are laid down in the same representational substrate as the representational 'products' of scene interpretation). This overlap causes an increase in activation of the episodic trace associated with the experience of the object, with a consequent change in the attentional state of the system. The virtue of this account is that it can apply to those cases where the eyes move towards something that is *not* currently being referred to.

Finally, the reliance of our account on pre-existing episodic representations of the objects in the scene can explain why, after '*eat*' or '*ride*' or '*has drunk/will drink*' the listener assumes that the object of the eating/riding/drinking is most likely something in the concurrent scene. These pre-existing representations are, as the language unfolds, the currently most active. And as the language continues to unfold, they become even more active (because of that conceptual overlap). And if some representation becomes more active than any other because it best fits the conceptual specification of a (thematic) role associated with the main verb, we can assume that this higher activation state constitutes the implicit assignment of the object associated with that representation as the most plausible candidate for that role. Hence the 'preference' to interpret objects in a concurrent scene as the most likely participants in the event described by the unfolding language.

Conclusions

We have suggested that visual attention shifts to whichever objects in the scene engender conceptual representations that overlap with the conceptual structures activated by the language. This is a dynamic process, dependent in part on the fact that the conceptual structures activated by the language change as that language unfolds in time. This approach may also extend to include language *production*; eye movements during language production (e.g., Griffin, 2004; Griffin & Bock, 2000) may reflect the relationship between conceptual structures activated by the scene, and the conceptual structures activated during the processes of formulating, and preparing, an utterance.

Our account of the linkage between language processing and eye movements suggests 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). In the case of the two studies we described earlier, this conceptual overlap is at the level of experiential knowledge of the actions and events in which depicted objects can take part (which includes information about what other objects they may interact with). The behavioral consequences of this

overlap are threefold: The overlapping episodic object representations change their activation state; this change constitutes (or at least causes) a shift in attention; and this shift in attention is accompanied by saccadic eye movements towards the location of the corresponding visual object (as recorded in the temporary episodic trace of the object).

This account is not limited to explaining only how language causes visual attention to shift around the *visual* world. In our studies, the scenes preceded the language, but in other studies of language processing, discourse contexts often precede a fragment of language. And those discourse contexts may function in precisely the same way as our visual scenes functioned—namely, by activating conceptual structures pertaining to objects and their interactions. And as language subsequent to this context unfolds, it may activate conceptual structures which overlap with those previously activated by the context. This overlap would cause these earlier representations to 'reactivate', with a consequent change in the attentional state of the system. In effect, attention would not shift about the visual world, but about the *mental* world. Indeed, our account of the linkage between linguistic process and visual attention is predicated on the notion that *mental* representations guide attention (as distinct from the retinal image itself). The only difference is whether the shift in attentional state results in a saccadic eye movement. If the episodic trace associated with the mental representation includes a component of the perceptual experience corresponding to the location of the object, a saccadic eye movement will likely ensue. But if there is no such component, it most likely will not. Other than that, there is little functional difference between the visual world in our studies and the mental world created through the use of discourse context in other studies. In this respect, there is considerable overlap between our demonstrations here, and prior demonstrations of the role of situation models during text comprehension (cf. Zwaan & Radvansky, 1998). Several such studies have shown how spatial relations within the situation model modulate the accessibility of referents. For example, Glenberg, Meyer, & Lindem (1987) manipulated in their texts whether a protagonist moved with, or away from, an object described in the text. Subsequently, the text continued to focus on the protagonist, after which a probe-recognition task revealed greater accessibility for the target object when it had remained, in the situation described, close to the protagonist. Similarly, Zwaan (1996) and Zwaan et al. (2000) showed how accessibility can also be determined by temporal relations (cf. Rinck & Bower, 2000). To the extent that probe recognition requires a shift in attention away from the protagonist and back towards the target, these studies show that it is not the physical text that determines the ease with which objects referred to in that text can be retrieved, but rather it is the mental correlate of the situation described by the text that determines this ease of retrieval. Given this dependence on the mental world, it is perhaps not so surprising, therefore, that many contextually-determined phenomena, including anticipation, circumscription of the referential domain, and sensitivity to affordances, to name just a few, are common to studies of language whether situated in visual or linguistic contexts.

Finally, we note that anticipatory eye movements are not unique to language-mediated visual attention. For example, Land, Mennie, and Rusted (1999) and Pelz and Canosa (2001) observed anticipatory eye movements during motor behaviour, with looks 'ahead' towards objects that would shortly require a reaching movement (see also Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). In these cases, eye movements, and the representations that engendered those eye movements, were modulated by participants' intentions given the task at hand. But even when there is no *explicit* intention to move the eyes in service of some other behaviour, anticipatory eye movements can occur; McMurray and Aslin (2004) and Gredebäck and von Hofsten (2004) found that six- to seven-month-old infants' can learn to anticipate, as evidenced by their eye movements, the reappearance of objects that had moved behind an occluder. Why such anticipatory movements occur (in the absence of any motoric behaviour such as reaching or navigation that could benefit from such anticipation) is a matter for conjecture, but accounts based on reinforcement learning may shed some light on this (see Hayhoe & Ballard, 2005, for review). In their review of eye movements and 'natural behavior', Hayhoe and Ballard (2005) concluded that 'eye movement patterns appear to be shaped by learnt internal models of the dynamic properties of the world' (p. 190). We believe the data we have presented here are a further example of precisely such behaviors. Anticipatory eye movements do not reflect the unfolding language; they reflect an unfolding (mental) world.

Acknowledgments

The research was supported by an award from The Medical Research Council (G0000224) to G.A. and an award from the British Academy (SG-37355) to Y.K. We thank Silvia Gennari for helpful discussion regarding event structures, and Jelena Mirkovic for helpful discussion on the distinction between linguistic and conceptual representation, and for reminding us that colorless green ideas afford form-based generalizations. The paper is considerably improved as a result of their comments on an earlier version. We

also thank Rolf Zwaan and two anonymous reviewers for their constructive comments that have also benefited the paper.

Appendix A

The sentential stimuli used in Experiment 1. The two versions of each item are presented on the same line, with the future tensed version first. Also shown is a list of the objects in the accompanying scene. The underlined object is the target object for the past tense version.

1. The man will drink the beer; The man has drunk the wine. [*man*, *glass with beer*, *empty wine glass*, *table*, *cheese*, *party crackers*]
2. The postman will deliver the letters; The postman has delivered the parcel. [*postman with letters*, *open door with woman at entrance*, *fence*, *tree*, *house*]
3. The boy will destroy the blocks; The boy has destroyed the sandcastle. [*boy*, *toy blocks in tower*, *pile of sand*, *tree*, *slide*]
4. The policeman will discharge the man; The policeman has discharged the woman. [*policeman*, *man in closed jail cell*, *open jail cell*, *chair*, *food dish*]
5. The woman will eat the cake; The woman has eaten the scones. [*woman*, *cake on plate*, *empty plate*, *coffee cup*, *napkin*]
6. The boys will enjoy the pizza; The boys have enjoyed the Chinese take-away. [*two boys*, *pizza in box*, *empty take-away food boxes*, *napkin*, *table*, *window*]
7. The scientist will evaporate the chemical; The scientist has evaporated the water. [*scientist*, *clamp with full flask*, *stand with empty flask*, *workbench*, *book*, *blackboard*]
8. The woman will finish the pie; The woman has finished the soup. [*woman*, *pie with fork*, *empty plate on napkin with spoon*, *vase with flowers*, *table*, *fridge*]
9. The woman will give away the toys; The woman has given away the clothes. [*woman*, *pile of toys*, *empty clothes cupboard*, *window*, *coffee mug*]
10. The dog will hide the bone; The dog has hidden the shoe. [*dog, bone*, *disturbed earth*, *cat*, *flowers*]
11. The cat will kill the mouse; The cat has killed the bird. [*cat*, *mouse*, *pile of feathers*, *water dish*, *fence*]
12. The woman will pour the coke; The woman has poured the wine. [*woman*, *coke bottle*, *empty wine bottle and cork*, *apple*, *table*, *wall clock*]
13. The fireman will put out the fire; The fireman has put out a fire. [*fireman*, *burning car*, *smouldering logs*, *stop sign*, *barrier*]
14. The boy will release the bird; The boy has released the dog. [*boy*, *bird in cage*, *empty dog kennel with lead*, *basket ball*, *plant*]
15. The grocer will run out of apples; The grocer has run out of oranges. [*grocer*, *2 bowls with apples*, *empty food crate*, *market stall*, *weighing scales*]
16. The waitress will serve the coffee; The waitress has served the dinner. [*waitress*, *tray with coffee*, *trolley with empty serving dish*, *worktop*, *sink*, *dustbin*]

Appendix B

The sentential stimuli used in Experiment 2. The two versions of each item are presented on the same line, with the future tensed version first. The numbering corresponds to the stimuli used in Experiment 1.

1. The man will drink all of the beer; The man has drunk all of the wine.
2. The postman will deliver all of the letters; The postman has delivered all of the parcels.
3. The boy will destroy all of the blocks; The boy has destroyed all of the sandcastle.
4. The policeman will discharge all of the men; The policeman has discharged all of the women.
5. The woman will eat all of the cake; The woman has eaten all of the scones.
6. The boys will enjoy all of the pizza; The boys have enjoyed all of the Chinese take-away.
7. The scientist will evaporate all of the chemical; The scientist has evaporated all of the water.
8. The woman will finish all of the pie; The woman has finished all of the soup.
9. The woman will give away all of the toys; The woman has given away all of the clothes.
10. The dog will hide all of the bones; The dog has hidden all of the shoes.
11. The cat will kill all of the mice; The cat has killed all of the birds.
12. The woman will pour all of the coke; The woman has poured all of the wine.
13. The fireman will douse all of the flames; The fireman has doused all of the logs.
14. The boy will release all of the birds; The boy has released all of the dogs.
15. The grocer will sell all of the apples; The grocer has sold all of the oranges.
16. The waitress will serve all of the coffees; The waitress has served all of the dinners.

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.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723–767.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*, 177–210.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47*, 30–49.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*, 687–696.

Christ, S. E., & Abrams, R. A. (2006). Just like new: newly segregated old objects capture attention. *Perception & Psychophysics, 68*, 301–309.

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

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.

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., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Language and Cognitive Processes, 16*, 507–534.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 498–513.

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

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.

Frisson, S., Rayner, K., & Pickering, M. (2005). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 862–877.

Gibson, J. J. (1977). The theory of affordances. In J. Bransford (Ed.), *Perceiving acting and knowing*. Killsdale, NJ: Erlbaum.

Glenberg, A. M. (1997). What memory is for? *Behavioral and Brain Sciences, 20*, 1–19.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26*, 69–83.

Gredebäck, G., & von Hofsten, C. (2004). Infants' evolving representation of moving objects between 6 and 12 months of age. *Infancy, 6*, 165–184.

Griffin, Z. M. (2004). The eyes are right when the mouth is wrong. *Psychological Science, 15*, 814–821.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*, 188–194.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. . *Journal of Vision, 3*, 49–63.

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). Hove: Psychology Press.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): a framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849–878.

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: Eye-tracking 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*, 23–32.

Huettig, F., & Altmann, G. T. M. (in press). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*.

Huettig, F., Quinlan, P., McDonald, S., & 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.

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*, 133–159.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: the role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*, 509–529.

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

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.

Land, M., Mennie, N., & Rusted, J. (1999). Eye movements and the roles of vision in activities of daily living: making a cup of tea. *Perception, 28*, 1311–1328.

Law, M. B., & Abrams, R. A. (2002). Object-based selection within and beyond the focus of spatial attention. *Perception & Psychophysics, 64*, 1017–1027.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research, 43*, 1735–1751.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy, 6*, 203–229.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*, B33–B42.

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.

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12*, 137–176.

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

Myung, J.-Y., Blumstein, S. E., & Sedivy, S. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition, 98*, 223–243.

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.

O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology, 46*, 461–488.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research, 41*, 3587–3596.

Richardson, D. C., & Spivey, M. J. (2001). The TEC as a theory of embodied cognition. *Behavioral and Brain Sciences, 24*, 900–901.

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

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition, 28*, 1310–1320.

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

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. 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.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: persistence of structural configuration in sentence production. *Cognition, 89*, 179–205.

Schwanenflugel, P. J., & LaCount, K. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 344–354.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (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*(5217), 1632–1634.

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences, 5*, 244–252.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*, 422–488.

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.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1196–1207.

Zwaan, R. A., Madden, C. J., & Whitten, S. N. (2000). The presence of an event in the narrated situation affects its activation. *Memory & Cognition, 28*, 1022–1028.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.