



Referential and visual cues to structural choice in visually situated sentence production

Andriy Myachykov*, Dominic Thompson, Simon Garrod and Christoph Scheepers

Department of Psychology, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

Edited by:

Yury Y. Shtyrov, Medical Research Council, UK

Reviewed by:

Lucy Jane MacGregor, Medical Research Council, UK

Andrei Kibrik, Moscow State University, Russia

*Correspondence:

Andriy Myachykov, Department of Psychology, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB, UK.
e-mail: andriy.myachykov@glasgow.ac.uk

We investigated how conceptually informative (referent preview) and conceptually uninformative (pointer to referent's location) visual cues affect structural choice during production of English transitive sentences. Cueing the Agent or the Patient prior to presenting the target-event reliably predicted the likelihood of selecting this referent as the sentential Subject, triggering, correspondingly, the choice between active and passive voice. Importantly, there was no difference in the magnitude of the general Cueing effect between the informative and uninformative cueing conditions, suggesting that attentionally driven structural selection relies on a direct automatic mapping mechanism from attentional focus to the Subject's position in a sentence. This mechanism is, therefore, independent of accessing conceptual, and possibly lexical, information about the cued referent provided by referent preview.

Keywords: sentence production, visual attention, structural choice

INTRODUCTION

Many psycholinguistic theories of sentence production suggest that selecting words, grammatical roles, and structural configurations are not arbitrary processes as they necessarily reflect the organization of the conveyed conceptual message via the rules of a regular interface between language and cognition (e.g., Bock, 1982; Jackendoff, 2002; Vigliocco and Hartsuiker, 2002; Myachykov et al., 2007). The emphasis of this paper is on the interface between the speaker's visual attention on the event's referents, accessibility of the conceptual information associated with these referents, and the assignment of grammatical roles and consequent syntactic structures in a spoken sentence.

Speaking about events in a real time situated context is a seemingly effortless routine task, performed daily by every language user. Yet, producing even a single utterance about a simple event is a complex process involving rapid and well-orchestrated execution of both linguistic and non-linguistic operations in the speaker's mind (Jackendoff, 2002). These operations do not only include information retrieval, they are also inherently selective; that is, they involve selecting information for earlier or later processing. Consider a situation in which the speaker describes a simple event, for example, *a boy kicking a ball*. The first necessary step in generating a sentence about this event is creating a non-linguistic conceptual plan of the event, or its *message* (Levelt, 1989). This message will be eventually translated into an emerging sentence via selecting words and assigning to them specific grammatical roles and positions in a syntactic structure. The speaker's visual attention will guide the translation by progressively selecting information for processing. This selection will be based on a number of parameters that make a particular referent, word, or structure more relevant, available, or conspicuous than the other available alternatives. This selection process already starts at the earliest stages of message apprehension when the non-linguistic properties of the event (including the relative *salience* of the interacting referents)

are encoded. At this stage, a variety of factors act as *cues* increasing referential salience. Some cues may be part of the speaker's own perspective on the event or knowledge about the referents. These are *endogenous* cues. Other cues are *exogenous*; they are specific features of the referent itself, for example, its size, shape, motion, or color. Let us assume the *boy's* larger size acts as an exogenous cue, preferentially attracting the speaker's attentional focus to it over the smaller and less salient *ball*. As a result, the *boy* may be selected for earlier and deeper processing than the *ball* (e.g., Itti et al., 1998; Itti and Koch, 2000; Parkhurst et al., 2002)¹. In other words, the *boy* will be coded in the message as the referent that is more *accessible* for processing than the *ball* (Bock and Warren, 1985).

As the non-linguistic message is forwarded for linguistic formulation, the more accessible referent may receive preferential treatment by means of earlier *lemma retrieval* and also by being assigned a more important grammatical role during *structural assembly* (Levelt, 1989). Hence, at lemma retrieval (where concepts receive their lexical names accompanied by grammatical properties), the *boy's* name will be accessed earlier than that of the *ball*. At the stage of structural assembly, the *boy* may receive a more prominent role, e.g., the Subject. In English, this will almost inevitably lead to the selection of the active-voice frame *A boy kicked a ball* (where the *agent* assumes the Subject role), rather than the alternative passive-voice frame *A ball was kicked by a boy* (where the *patient* assumes the Subject role). This simple example portrays how attentional focus driven by purely perceptual properties of a referent may in principle predict the likelihood of Subject assignment and the resulting choice between available

¹We acknowledge that factors other than exogenous cues play a role in capturing visual focus during natural scene viewing (e.g., Corbetta and Shulman, 2002; Henderson, 2003). Here, we focus on the role of referential salience and, therefore, on exogenously captured visual attention in the process of sentence generation.

structural configurations (see Myachykov et al., 2011, for a recent review).

It has to be noted that visual salience is not the only factor that can influence Subject assignment. It is well known that *linguistic* cues, such as priming a word associated with a referent (Flores d'Arcais, 1975; Osgood and Bock, 1977; Bock and Irwin, 1980; Bates and Devescovi, 1989; Prat-Sala and Branigan, 2000), or priming aspects of structural configuration (Ferreira and Bock, 2006; Branigan, 2007; Pickering and Ferreira, 2008 for recent reviews), also exert strong influences on Subject assignment and the resulting structural choice. One can therefore hypothesize that the Subject role encodes both the non-linguistic (perceptual or conceptual) and the linguistic (lexical or structural) salience of a referent. Here, we focus on the role of non-linguistic salience as determined by visual and/or conceptual cues.

The tendency of salient referents to assume prominent grammatical roles in sentences was already noted in a number of early psycholinguistic experiments using a *referent preview* paradigm. One such study (Prentice, 1967) used a set of cartoon pictures portraying simple transitive interactions between two characters (e.g., *fireman kicking cat*). Some of the characters were human, others animals, and inanimate objects. The pictures were paired with slides of one of the event's characters: the agent or the patient. Therefore, one of the referents was cued before the full event was displayed. Participants first viewed the cue slide and then the whole event, of which they provided spoken descriptions. As a result, speakers were more likely to place the previewed referent first in their target-event descriptions, making it the sentential Subject, leading to a higher proportion of passive-voice descriptions (e.g., *A cat was kicked by a fireman*) in the patient-preview condition. Prentice explained this result by suggesting that referent preview acted as an attentional cue to the referent that participated in the subsequent event. Importantly, the cue slide was always presented *in the center* of the screen and not in the location where the corresponding referent would later appear. Hence, visual attention *per se* does not have to be invoked, as the structural choice effect most likely resulted from preferential access to the conceptual (and potentially, lexical) information about the cued referent, rather than from directing attention to the subsequent target's location. We will return to this issue below.

Experiments that followed Prentice (1967) used a similar setup. For example, Turner and Rommetveit (1968) presented children with active/passive sentences and later asked participants to recall these sentences. Both at the time of encoding and recall, sentences were presented to participants randomly paired with a picture of the agent, the patient, the whole event, or a blank. Among other things, Turner and Rommetveit found that the active-voice sentences were more likely to be recalled correctly if the visually primed referent was the agent, while the passive-voice sentences were better remembered if the primed referent was the patient. Although the latter study involved referent preview at both the encoding and the recall stages, the retrieval-picture effect and the storage-picture effect were attested separately. The authors found that the retrieval-picture effect was stronger, suggesting that the assignment of the referent's role in the sentence was affected more strongly by referent preview during production of the target

description than by the encoding of the target sentence for later recall.

These early studies seem to confirm the hypothetical scenario we outlined above: preferential attention to a referent can predict the choice of sentential structure via assignment of the Subject role to the most salient referent. However, the "attentional" manipulations in these studies employed a referent preview long enough (more than 600 ms) not only to bias attention toward the subsequently presented referent, but also for the participant to fully recognize the referent's identity, and potentially even activate its name. Also, the preview of a referent did not inform the participants about the corresponding referent's location in the subsequently presented target event. Therefore, although visual attention may have been implicated in the resulting structural choice effect, a plausible alternative explanation might be that referent preview primed the speaker's access to the conceptual (and possibly lexical) information associated with the primed referent, which in itself is enough to predict Subject selection without invoking any specific notion of attention.

Studies using a *visual cueing* paradigm directly address the question of how visual attention *per se* can predict Subject assignment and structural choice. In contrast to a referent preview paradigm, visual cueing studies use visual prompts that do not provide any information about the cued referent (Posner, 1980). Participants usually see a pointer, a dot, or a square, cueing the referent's location before the event presentation or simultaneously with it. Importantly, the cue itself does not provide any conceptual information about the cued referent; hence, any resulting structural choice effect must be attributed to visual attention and not to other factors, for example, prior higher memorial activation of conceptual and/or lexical information associated with the cued referent.

One of the earliest studies using a visual cueing paradigm was the Fish Film experiment by Tomlin (1995). In this study, English speakers described an animated film portraying one fish eating another. A visual cue (a pointer) directed participants' attention to the eventual Patient or Agent fish as the two fish approached each other; that is, before the eating event. When the cue was on the eventual agent, participants predominantly described the event with an active-voice sentence (e.g., *The blue fish ate the red fish*). When it was on the patient, they produced passive-voice descriptions most of the time (e.g., *The red fish was eaten by the blue fish*). Hence, the focally attended referent was consistently assigned the sentential Subject role, driving the choice between active and passive voice. Although Tomlin's results were very intriguing, both the cueing procedure and the repetitive nature of the Fish Film paradigm received criticism from some psycholinguists for being "too brutal" (Bock et al., 2004) or crude and suggestive about the experimenter's goal (Gleitman et al., 2007). From a methodological point of view, such criticisms are justified to some extent. First, although the experimental instructions did not tell participants anything about how to treat the cue in relation to the choice of event description, it considerably constrained their attentional focus to the cued referent making it not only perceptually, but also conceptually, more accessible. In this respect, presenting a pointer cue together with the stimulus (for a time long enough to recognize the cued referent) makes this cueing manipulation

very similar to the referential priming paradigm described above. Hence, any conclusion about independent contributions of visual attention to the selection of the sentential Subject remains only partially justified. Second, the Fish Film paradigm instructs participants to view and describe continuously *all* the interactions between the fish, including those preceding the target event. This inevitably increases the *givenness* (e.g., Bock, 1982; Givon, 1992) of the cued fish. Finally, the repetitive nature of the target event and the lack of interrupting filler materials make effects of syntactic persistence a possible concern (Bock, 1986). Nevertheless, this original finding and the Fish Film paradigm became in many ways ground-breaking; its variants were later used in studies of other syntactic structures (e.g., Forrest, 1996) and languages structurally different from English (Diderichsen, 2001; Rasolofo, 2006; Myachykov and Tomlin, 2008).

A more recent study (Gleitman et al., 2007) tried to avoid the methodological problems in Tomlin (1995) by separating the cue from the target event, using *implicit* rather than *explicit* cues, and monitoring attention through eye-tracking. Sentences with verbs of perspective (*give/receive*), conjoined noun phrases (*The boy and the girl/The girl and the boy*), voice alternating transitive sentences, and symmetrical predicates (*The boy meets the girl/The girl meets the boy*) were elicited with the help of still pictures presented on a computer screen. Participants' attention was directed to the location of one of the subsequently presented referents, *before* the target-event presentation, by flashing a black square on the screen for 75 ms. This short cue duration ensured that participants remained unaware of the manipulation itself, although their gaze (and the focus of attention) was attracted to the cued location *implicitly*. The success of the cueing manipulation was monitored by recording eye movements in real time. Once the picture was on the screen, participants extemporaneously described the presented event without any further manipulations of attention. The magnitude of the resulting visual cueing effect was smaller than the one reported by Tomlin; nevertheless, the cued referent was more likely to be assigned the sentential Subject position, triggering the choice between corresponding structural alternatives.

Overall, the studies reviewed here, as well as a number of similar studies (see Myachykov et al., 2011 for a review) consistently showed that a visual cue to a specific referent in an event, uninformative with regard to the cued referent's conceptual and/or linguistic properties, reliably predicts the selection of that referent as the Subject (and associated structural choice). As a result, some theoretical proposals claim a direct link between visual attention on (or salience of) a referent on the one hand and assignment of the Subject role to that referent on the other (Tomlin, 1997; Myachykov et al., 2011). While this is a relatively simple and straightforward proposal, its validity is difficult to assess in the absence of studies that directly compare the effects of referential and visual cueing. One possibility is that referent preview provides more information about the cued referent than visual cueing. At least in principle, given enough preview time, speakers can extract both conceptual and lexical information about the referent. This is not so in the case of a purely visual cueing scenario. Indeed, if directing attention to the location of a referent (via a conceptually uninformative cue) provides only a part of the information provided by referent preview, then visual cueing might

have a weaker effect on subsequent Subject selection than referent preview.

The issue of cue informativity introduced above is related to the psycholinguistic concept of *conceptual accessibility* or the ease of retrieval of the conceptual information about the referent from working memory (Bock and Warren, 1985). Although the concept itself is very broadly defined as related to notions such as "codeability," "imageability," "retrievability," etc., the concept itself has been repeatedly invoked in psycholinguistic studies in order to explain why information associated with some referents (or, more broadly, concepts) is accessed or retrieved ahead of the information about other referents or concepts. A number of referent-related properties were shown to be responsible for an increase in conceptual accessibility, such as *givenness* (Bock, 1977; Arnold et al., 2000), *animacy* (Clark, 1966; Sridhar, 1988; Bock et al., 1992; McDonald et al., 1993; Prat-Sala and Branigan, 2000; Christianson and Ferreira, 2005; Altman and Kemper, 2006), *definiteness* (Grieve and Wales, 1973), and *prototypicality* (Kelly et al., 1986). What is important here is the fact that, similarly to lexical priming of a referent's name (e.g., Tannenbaum and Williams, 1968; Flores d'Arcais, 1975; Bock and Irwin, 1980; Bock, 1986; Bates and Devescovi, 1989) priming a referent's conceptual accessibility has also been shown to be a strong predictor of Subject selection and the resulting structural choice (e.g., Bock, 1977; Bock et al., 1992; Arnold et al., 2000; Prat-Sala and Branigan, 2000; Christianson and Ferreira, 2005). If conceptual accessibility is related to enhanced memory trace for the corresponding referent's mental representation, then additional memorial activation provided by referent preview should increase the conceptual accessibility of the referent beyond directing attentional focus to it. Hence, the bias to assign the Subject role to the cued referent and to alternate structure accordingly should be particularly strong in cases where a referent preview cue provides information about the cued referent's identity *as well as* points to its location. The effect of a purely visual cue to the location of a referent should, therefore, be weaker because such a cue provides no conceptual information about the referent. An alternative prediction stems from theories that emphasize a special role of attentional focus among non-linguistic factors affecting Subject assignment (Tomlin, 1997; Gleitman et al., 2007; Myachykov et al., 2011). If what matters is only the attentional focus on the cued referent, then there should be no difference in the strength of referential and visual cueing effects. The experiment reported below therefore directly compares the effects of perceptual and referent preview on structural choice. Specifically, we compare the effectiveness of cues that provide only location information with the effectiveness of cues that provide both location *and* referential information.

EXPERIMENT

DESIGN

Two factors were independently manipulated at two levels each: (1) Cue Location (Agent/Patient) and (2) Cue Type (Referent/Dot). Both manipulations were within-subjects and between-items. Cue Location was manipulated by means of presenting a visual cue in the location of one of the subsequently presented visual referents (agent or patient). The dependent variable was the probability of producing Passive-Voice sentences.

PARTICIPANTS

Twenty-four native English speakers (Glasgow University undergraduates; 12 female) with normal or corrected-to-normal vision took part. They either received course credits or £6 subject payment. The mean age of the participants was 20.3 years.

MATERIALS

The target pictures consisted of 64 black-and-white cartoon drawings showing simple transitive events (see example in **Figure 1**) and employed eight different event types (*chase, kick, pull, punch, push, scold, shoot, and touch*). Each event type appeared equally often in the Dot-Cue and the Referent-Cue conditions.

The materials were counterbalanced for left–right orientation (i.e., the agent was either on the left or on the right on an equal number of trials), size, animacy, color, and referent role suggestibility (i.e., both referents were equally plausible as being an agent or a patient). The human referents used in the target stimuli appeared in both the agent and the patient role in an equal number of trials. Since it was important that the visual referents were easily recognizable and distinguishable from one another, it was difficult to match them for familiarity. To compensate for this, we provided a practice session at the beginning of each experiment which familiarized participants with all the characters and events they would encounter (see Procedure). The materials were not controlled for corpus frequency; therefore participants previewed the single pictures of each referent during the practice session and became familiarized with the referents they encountered later in the experimental session.

We included 130 filler pictures showing various arrangements of geometrical shapes presented in different regions of the screen (e.g., a square diagonally above and right of a heart); participants had to describe those visual arrangements in the filler trials by producing a locative sentence describing the shapes and the relationship between them. Randomization was constrained so that there were always four fillers at the beginning of each session and each prime–target pair was preceded by at least two filler trials.

APPARATUS

The experiment was implemented in *SR-Research Experiment Builder*. An EyeLink II head-mounted eye-tracker monitored

participants' eye movements in order to ensure the efficiency of the cueing manipulation. Other than that, we will not report any eye-movement data since the focus of this paper is on how the experimental manipulations affect speakers' structural choices. The experimental materials were presented on a 17" CRT monitor of a DELL Optiplex GX 270 desktop computer running at a display refresh rate of 75 Hz. Also connected to the PC was a pair of stereo speakers. A SONY DAT recorder was used for speech recording. The audio clips were later uploaded onto a PC and analyzed with the help of Adobe Audition 2.0. The eye-tracking data were extracted and filtered using *SR-Research Data Viewer*.

PROCEDURE

Participants were positioned approximately 60 cm from the display. They had a direct view of the monitor throughout the session. Viewing was binocular, but only the participant's right eye was tracked. Before the main experimental session, each participant was run through a practice session during which they saw the pictures of the referent characters that would later be presented in the target trials and sample pictures of both the target and the filler materials. The referents appeared one at a time in the center of the screen simultaneously with their names. Participants were instructed to read out the referent names and to remember them for the following tasks. Also, each participant had to describe eight sample event pictures (one for each event type) during the practice session. The pictures of the events were presented in the middle of the screen. No specific instruction as to how to describe these event pictures was given to participants, except that participants should always make reference to the event and both interacting characters.

The instruction for the experiment proper was to describe a picture extemporaneously and in a single sentence using the present tense. Participants were unaware of the nature of the experimental manipulations, any difference between target and filler trials, or the exact purpose of the study. They were told that the study was concerned with speaking about what they see on the computer screen. **Figures 2 and 3** illustrate typical target trial sequences.

Each target trial began with the presentation of the central fixation cross. Shortly after the participant fixated it, a dislocated fixation cross appeared on the screen. This ensured that participants would not be looking at the center of the screen at the time of cue presentation and that they would always have to make a saccade to the cued location or, if the cue was overlooked, to another location in the target picture once it appeared on the screen. The dislocated fixation cross was equally distant from the cued locations. The presentation of the cue was contingent on fixating the dislocated fixation mark. Participants fixated the dislocated fixation mark for a minimum of 200 ms, after which either a Dot-Cue or a Referent-Cue screen was displayed. The Dot Cue was a red circle (25 pixels in diameter), which appeared in the approximate center of one of the subsequently presented referents (agent or patient). The Referent Cue was operationalized via previewing one of the event referents (agent or patient) prior to the target display presentation. The previewed referent always appeared in neutral posture, preventing any thematic role (agent or patient) suggestibility. As with the Dot Cues, Referent Cues appeared in the locations corresponding to its location in the subsequently

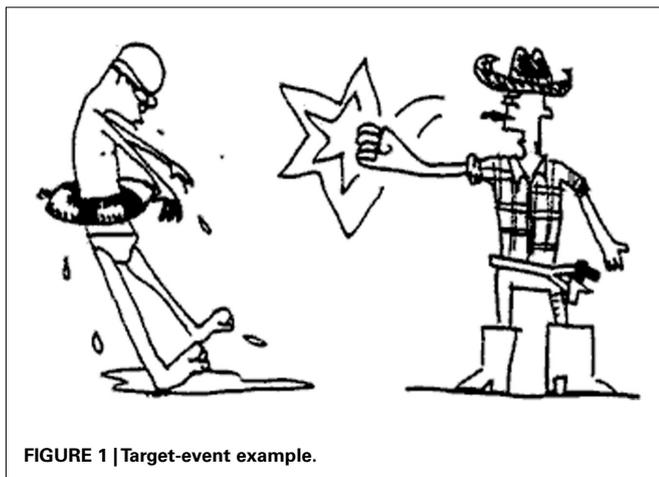
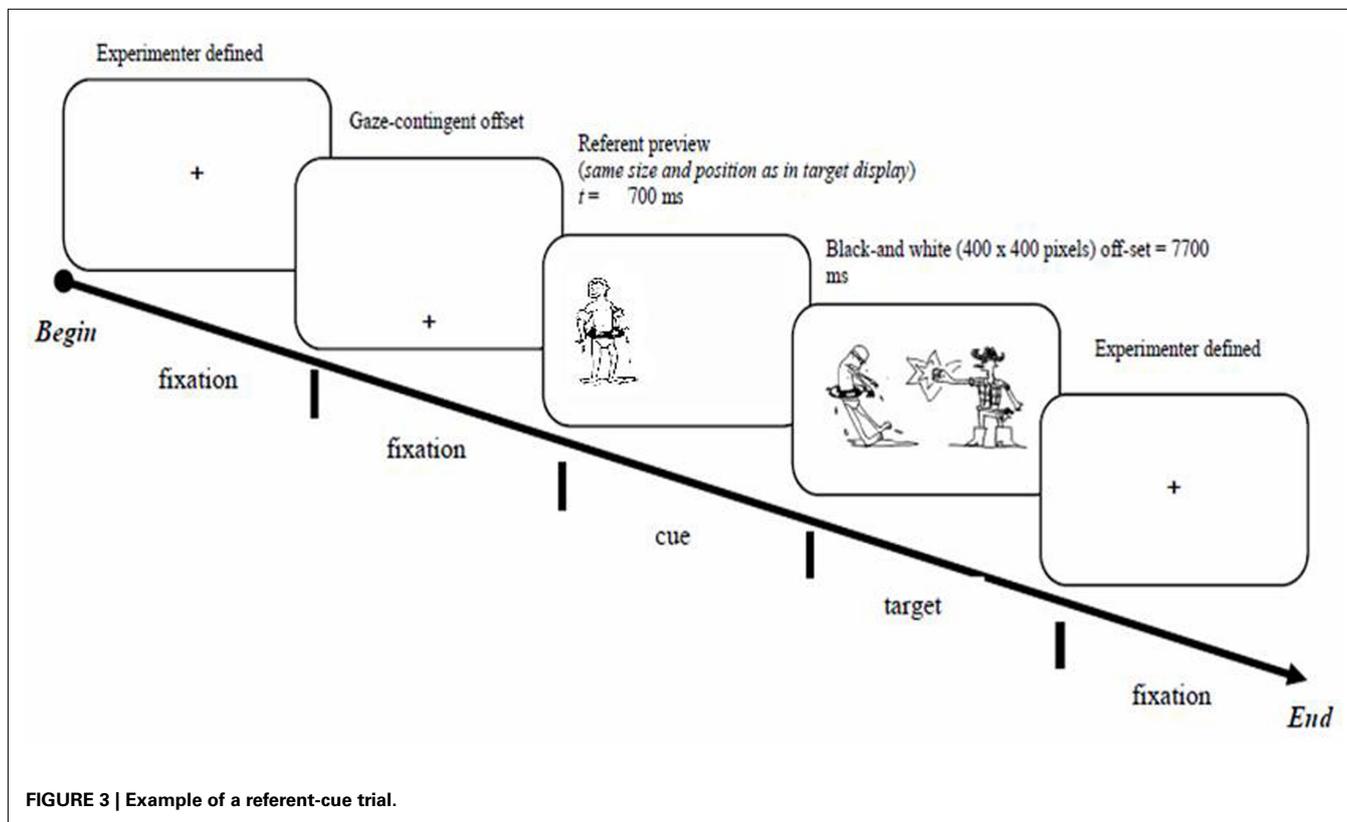
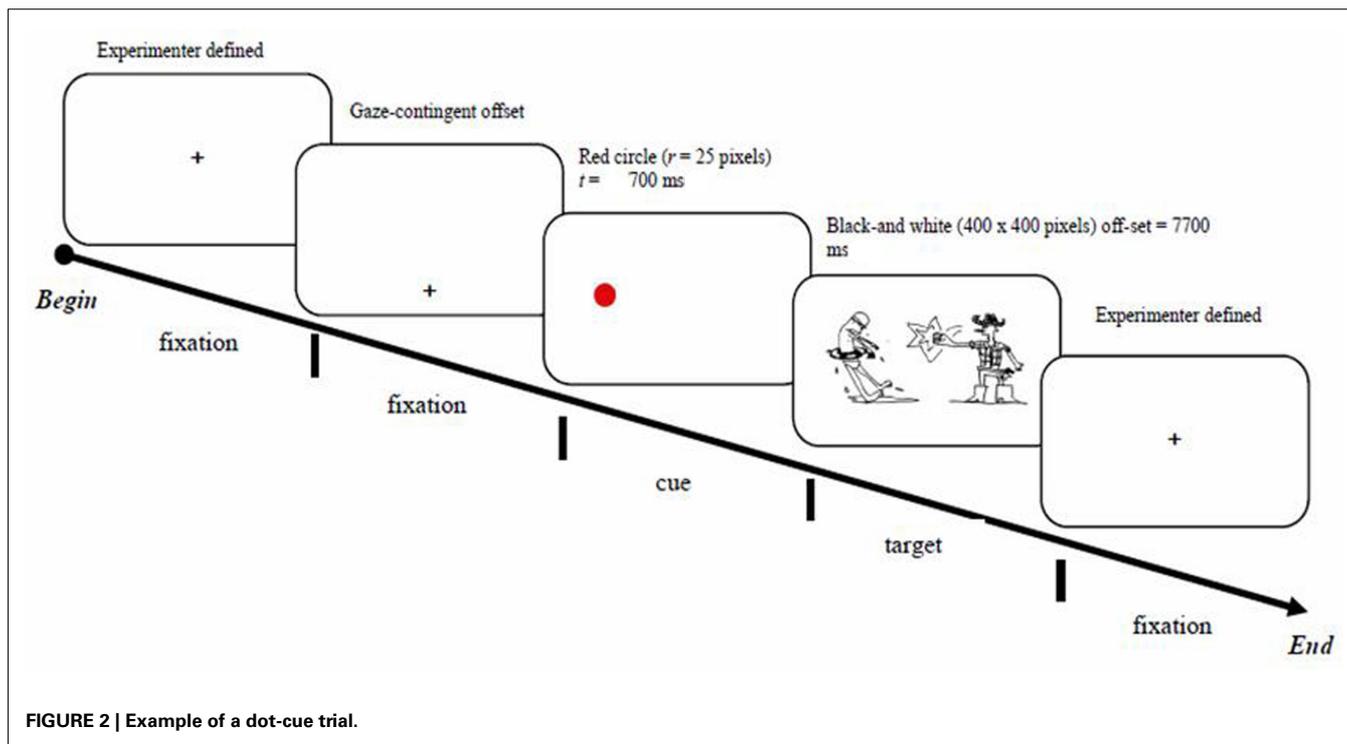


FIGURE 1 | Target-event example.



presented target display. Hence, Dot Cues only provided location information whereas Referent Cues provided both location and referential information. Cue duration was 700 ms regardless of

the Cue Type. There was no specific instruction as to how the cues should be treated. After the cue presentation, the target picture appeared on the screen. Participants were instructed to describe

the target picture in a single sentence, and to press the space bar to move on to the next trial. In case the participant did not respond, the picture disappeared from the screen after 7700 ms. Filler trials employed a comparable presentation sequence: the trial would begin with a central fixation mark, after which a dislocated fixation mark appeared, followed by a visual cue (identical to the procedure in the target trials), and finally, the presentation of the target display.

RESULTS

CUEING EFFICIENCY

In order to analyze initial fixations on visually cued versus non-cued referents, the pictures were pre-coded to include separate areas of interest: one for each referent (agent and patient) and one for the background. The referent areas included the referent itself plus a surrounding area of about two degrees of visual angle. Both Dot and Referent cueing manipulations were highly effective in attracting initial visual attention to the cued location. In approximately 96% of the experimental trials, presenting the cue led to the execution of a saccade to the cued location. When the (Dot or Referent) cue was replaced with the target picture (700 ms after cue-onset), participants continued to look at the cued referent, accounting for approximately 90% of initial fixations in the target trials.

TARGET STRUCTURE

Target responses were coded by a naïve coder as Active Voice, Passive Voice, or Other. To be coded as Active Voice, the description had to employ a transitive verb referring to the depicted event, a subject NP referring to the agent, and a direct object NP referring to the patient (e.g., *The cowboy is punching the boxer*). To be coded as Passive Voice, the description had to employ a passivized transitive verb referring to the depicted event, a subject NP referring to the patient, and a by-phrase referring to the agent (e.g., *The boxer is [being] punched by the cowboy*). Note that truncated passives (not including a by-phrase) were hardly ever produced since they were explicitly discouraged in the practise session. All remaining responses (including missing responses) were coded as Other. The latter accounted for less than 1.5% of the data and will not be considered further.

Statistical analyses were performed in SPSS/PASW 19 using *Generalized Estimating Equations* (GEE, e.g., Hardin and Hilbe, 2003). Unlike ANOVA, GEE allows for specifying distribution and link functions that are appropriate for analyzing categorical frequencies. Here, we used a *binomial* distribution and *logit*

link function (cf. Jaeger, 2008) to model proportions of passive-voice responses as a function of Cue Location (agent or patient) and Cue Type (referent or dot). The two predictors were entered as *within-subjects* (respectively *between-items*) variables assuming a compound symmetry covariance structure for repeated measurements. **Table 1** and **Figure 4** present the results of our analysis.

The reliable intercept confirms that passive-voice responses were less likely overall than the active-voice responses. This finding is in line with existing corpus-analysis data (e.g., Svartvik, 1966; Roland et al., 2007) as well as previous findings using visual cueing and referent preview paradigms (see Myachykov et al., 2011 for review). Our analysis registered the presence of a reliable main effect of Cue Location: when the *patient* was cued, passive-voice responses were $23 \pm 10\%$ more likely by subjects and $23 \pm 6\%$ more likely by items, than when the *agent* was cued. This finding provides further support to the previously reported tendency of the attentionally focused referents to correspond to the Subject position in an English transitive sentence. More importantly, in our data there was no suggestion of a Cue-Type effect (dot versus referent) nor of an interaction between Cue Location and Cue Type.

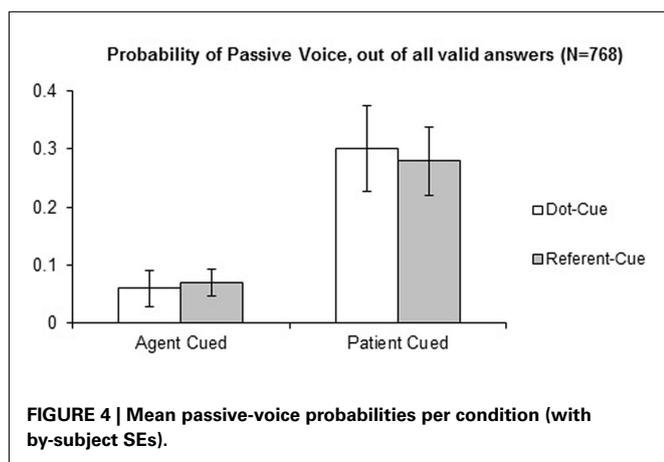
DISCUSSION

In this experiment, we analyzed how directing the speaker's attention to one of the event's referents via prior presentation of a referentially uninformative visual location cue or a referent preview (in the same location) predicts the assignment of the Subject position and the resulting structural choice during English transitive sentence production. Following on from the existing theories, we investigated whether speakers use a combination of perceptual and conceptual information provided by the cue, or only the perceptual information about the location of the cued referent when choosing the Subject and the resulting grammatical structure of a spoken sentence.

In general, cueing the location of the eventual patient resulted in a higher probability of selecting the patient as the sentential Subject and producing passive-voice responses. This finding is in line with previous reports suggesting that attentional focus plays a special role in determining Subject assignment (and corresponding structural choice) in visually situated sentence production (e.g., Tomlin, 1995; Gleitman et al., 2007; Myachykov et al., 2011). The novel finding is that the two different cueing manipulations were equally successful in predicting the choice of Subject, regardless of whether the deployed cue was referentially uninformative or

Table 1 | Results from logit binomial GEE analyses modeling proportions of passive-voice responses as a function of Cue Location (L) and Cue Type (T).

Effect	By subjects		By items	
	$gs\chi^2(1)$	<i>P</i>	$gs\chi^2(1)$	<i>P</i>
Intercept	37.28	0.001	29.64	0.001
Cue location (L)	13.69	0.001	19.24	0.001
Cue type (T)	0.00	0.949	0.04	0.852
L × T interaction	0.14	0.711	1.13	0.287



was a full referent preview long enough for the speaker to extract both conceptual and lexical information about the cued referent. Contrary to the prediction that conceptual accessibility plays an independent role in determining Subject assignment, the cueing effect on structural choice in the referent-cue condition was no different from that in the dot-cue condition. Our data, therefore, provide further support to the special role of attentional focus in the assignment of the constituents' roles and the resulting structural choice during visually situated sentence production.

Indeed, accessibility-based theories predicted that, in addition to cueing the location of an eventual referent, referent preview would establish a stronger memorial trace for the corresponding referent, which should have led to a further modulation of the overall cueing effect on structural choice. The fact that there was no such modulation might be interpreted as suggesting that participants did not access conceptual and/or lexical information about the previewed referent. This interpretation, however, does not seem very plausible given the amount of time participants were able to preview the referent in the referent-cue condition. It can also be argued that referent preview *alone* is not sufficient to increase the referent's conceptual accessibility, and that other properties of the referent, such as animacy or humanness, need to be manipulated in order to achieve such an increase. This is an interesting empirical question in itself, but its premise takes us back to a very loosely defined notion of conceptual accessibility in the first place. Taking the original definition that "*Conceptual accessibility is the ease with which the mental representation of some potential referent can be activated in or retrieved from memory*" (Bock and Warren, 1985, p. 50), previewing a referent should have achieved exactly that – a better memorial trace for the cued referent's mental representation. If such memorial facilitation played an independent role in Subject selection, then in the design implemented in the current study it should do so *in addition* to biasing attention to one of the subsequently presented referents. The fact that referent preview did not boost the effect of location cueing suggests that attentional focus is the primary driving factor in alternating Subject assignment, thus biasing structural choice.

We propose an alternative interpretation, according to which a stronger memorial representation associated with referent preview

plays no additional role in Subject assignment beyond directing attention to the cued referent – the general cueing effect observed in both experimental conditions. In other words, once the speaker commits to using an attentional cue as the predictor of the Subject position, an additional memorial facilitation of the referent-related information does not improve this bias any further. Comparison of our data with the earlier study by Prentice (1967)², in which *centrally* established referent preview successfully predicted the assignment of the referent as the Subject, helps to further elaborate our theoretical interpretation. If Prentice's interpretation of her own data was correct in that central referent preview acted as an endogenous cue, orienting participants' attention to the location of the subsequently presented referent, then our study in fact replicates this effect, this time with a lateral referent preview, and using a cue that was mixed: it was both endogenous (that is, prompting participants to identify the previewed referent once the target event was displayed) and exogenous (by virtue of being a laterally presented visual cue accurately predicting the previewed referent's location in the subsequent event; Posner, 1980). The lack of a Cue-Type effect in the current study suggests that referent preview generally acts as a memorial cue to search for the subsequently presented referent. A number of recent reports documented the ability of information held in working memory to affect the distribution of visual attention in perceptual processing tasks (e.g., Bundesen, 1990; Desimone and Duncan, 1995; Downing, 2000; Kumar et al., 2009). In other words, what people currently have in mind can affect what they attend to later. Importantly, these *memorial cues* do not have to be spatial, as linguistic information currently held in working memory has also recently been shown to determine the spatial deployment of visual attention (e.g., Soto and Humphreys, 2007; Hodgson et al., 2009; Mannan et al., 2010; Anderson et al., 2011; Salverda and Altmann, 2011). Our data suggest that once the attentional cue is established in the speaker's working memory, irrespective of whether it was established with the help of a pointer or a referent preview, this attentional cue biases the speaker to select the referent that later appears in the cued location as the sentential Subject. One prediction from this view is that one should observe comparable cueing effects on structural choice for a situation in which, in one condition, referent preview would be established centrally (hence, uninformative about the referent's location), and in the other, laterally (hence, informative about the referent's location). Another way to address this question is via the use of *conflicting* cues, i.e., when a patient referent appears in the agent location (or *vice versa*) at the time of cueing and before the target picture display. This scenario helps address the question of the speaker's selection bias arising from resolving a conflict between information from an endogenous cue (bias to locate the referent whose identity was revealed by the preview) and an exogenous cue (the location of the previewed

²It has to be noted that there are important differences between the design used in the current study and the one utilized by Prentice. These include lack of fillers in Prentice (1967), unclear description of the cue slides, and heterogeneous set of referents used in that study: some of the referents were human, others were animals and inanimate objects, and others were indefinite referents, such as *leaves* or *fire*. These features may have affected Subject assignment in their own right.

referent that, in our example, conflicts with its role in the target event).

So, what is the specific role of enhanced memorial activation associated with referent preview in the process of visually situated sentence generation? Our data do not answer this question directly other than suggesting that, as far as Subject selection and structural choice are concerned, there clearly were no cue-enhancing effects of referent preview. However, additional analyses of sentence onset latencies (the time from the onset of the picture to the onset of the participant's response) suggest that participants were on average faster (by 132 ms) to initiate their responses in the referent preview condition than in the dot-cue condition. Although this Cue-Type effect did not approach significance ($ps > 0.1$), the direction of this difference suggests that participants were more prepared to "fill in" the Subject slot by way of pre-activated conceptual and/or lexical access. This interpretation leads to intriguing theoretical implications. It would suggest, for example, that the choice of Subject (which our study showed to depend primarily on attentional focus) is a mechanism separate from the assembly of the corresponding committed structure. Apparently, in both the dot-cue and the referent-cue

conditions, participants were biased to assign the Subject role to the referent that later appeared in the cued location. However, in the referent-cued condition, they also knew the identity and the name of the referent, with which they wanted to fill the Subject slot. The difference in sentence onset latency, albeit statistically unreliable, suggests that this knowledge could matter; not at the stage of structural choice, but at the stage of lemma access and linear arrangement of the constituents in the chosen structure. This would be an interesting direction for further research.

In conclusion, we have shown that structural choice (assignment of the Subject role to either the agent or the patient of a transitive event) is primarily driven by attentional factors such as a visual cue to the location of a referent. Additional information about the referent's identity in the cue did not significantly modulate structural choice further, but there might be an influence of referential cueing on conceptual and/or lexical access.

ACKNOWLEDGMENTS

This research was supported by the Economic and Social Research Council grant PTA-026-27-1579 awarded to Andriy Myachykov.

REFERENCES

- Altman, L. J. P., and Kemper, S. (2006). Effects of age, animacy and activation order on sentence production. *Lang. Cognitive Proc.* 21, 322–354.
- Anderson, S. E., Chiu, E., Huette, S., and Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychol. (Amst.)* 137, 181–189.
- Arnold, J., Wasow, T., Losongco, A., and Ginstrom, R. (2000). Heaviness vs. newness: the effects of complexity and information structure on constituent ordering. *Language* 76, 28–55.
- Bates, E., and Devescovi, A. (1989). "Competition and sentence production," in *The Crosslinguistic Study of Sentence Processing*, eds B. MacWhinney and E. Bates (New York: Cambridge University Press), 225–256.
- Bock, J. K. (1977). The effect of pragmatic presupposition on syntactic structure in question answering. *J. Verbal Learn. Verbal Behav.* 16, 723–734.
- Bock, J. K. (1982). Towards a cognitive psychology of syntax: information processing contributions to sentence formulation. *Psychol. Rev.* 89, 1–47.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychol.* 18, 355–387.
- Bock, J. K., and Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *J. Verbal Learn. Verbal Behav.* 19, 467–484.
- Bock, J. K., Irwin, D. E., and Davidson, D. J. (2004). "Putting first things first," in *The Integration of Language, Vision, and Action: Eye Movements and the Visual World*, eds J. Henderson and F. Ferreira (New York: Psychology Press), 249–278.
- Bock, J. K., Loebell, H., and Morey, R. (1992). From conceptual roles to structural relations: bridging the syntactic cleft. *Psychol. Rev.* 99, 150–171.
- Bock, J. K., and Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21, 47–67.
- Branigan, H. P. (2007). Syntactic priming. *Lang. Linguist. Compass* 1, 1–16.
- Bundesden, C. (1990). A theory of visual attention. *Psychol. Rev.* 97, 523–547.
- Christianson, K., and Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98, 105–135.
- Clark, H. H. (1966). The prediction of recall patterns in simple active sentences. *J. Verb. Learn. Verb. Behav.* 5, 99–106.
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.
- Diderichsen, P. (2001). "Selective attention in the development of the passive construction: a study of language acquisition in Danish children," in *Ikonicitet og struktur. Network for Funktionel Lingvistik*, eds E. Engberg-Pedersen and P. Harder (Department of English, University of Copenhagen).
- Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychol. Sci.* 11, 467–473.
- Ferreira, V. S., and Bock, K. (2006). The functions of structural priming. *Lang. Cogn. Process.* 21, 1011–1029.
- Flores d'Arcais, G. B. (1975). "Some perceptual determinants of sentence construction," in *Studies in perception. Festschrift for Fabio Metelli*, ed. G. Flores d'Arcais (Milan: Martello-Guinti), 344–373.
- Forrest, L. B. (1996). "Discourse goals and attentional processes in sentence production: the dynamic construal of events," in *Conceptual Structure, Discourse and Language*, ed. A. E. Goldberg (Stanford, CA: CSLI Publications), 149–162.
- Givon, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics* 30, 5–55.
- Gleitman, L., January, D., Nappa, R., and Trueswell, J. (2007). On the give-and-take between event apprehension and utterance formulation. *J. Mem. Lang.* 57, 544–569.
- Grieve, R., and Wales, R. (1973). Passives and topicalization. *Br. J. Psychol.* 64, 173–182.
- Hardin, J., and Hilbe, J. (2003). *Generalized Estimating Equations*. London: Chapman and Hall/CRC.
- Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends Cogn. Sci. (Regul. Ed.)* 7, 498–504.
- Hodgson, T. L., Parris, B. A., Gregory, N. J., and Jarvis, T. (2009). The saccadic Stroop effect: evidence for involuntary programming of eye movements by linguistic cues. *Vision Res.* 49, 569–574.
- Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506.
- Itti, L., and Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York, NY: Oxford University Press.
- Jaeger, F. T. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446.
- Kelly, M., Bock, J. K., and Keil, F. (1986). Prototypicality in a linguistic context: effects on sentence structure. *J. Mem. Lang.* 25, 59–74.
- Kumar, S., Soto, D., and Humphreys, G. W. (2009). Electrophysiological evidence for attentional guidance by the contents of working memory. *Eur. J. Neurosci.* 30, 307–317.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT.
- Mannan, S. K., Kennard, C., Potter, D., Pan, Y., and Soto, D. (2010). Early oculomotor capture by new onsets driven by the contents of working memory. *Vision Res.* 50, 1590–1597.

- McDonald, J. L., Bock, J. K., and Kelly, M. H. (1993). Word and world order: semantic, phonological, and metrical determinants of serial position. *Cogn. Psychol.* 25, 188–230.
- Myachykov, A., Posner, M. I., and Tomlin, R. S. (2007). A parallel interface for language and cognition in sentence production: theory, method, and experimental evidence. *Linguist. Rev.* 24, 455–472.
- Myachykov, A., Thompson, D., Scheepers, C., and Garrod, S. (2011). Visual attention and structural choice in sentence production across languages. *Lang. Linguist. Compass* 5, 95–107.
- Myachykov, A., and Tomlin, R. S. (2008). Perceptual priming and structural choice in Russian sentence production. *J. Cogn. Sci.* 9, 31–48.
- Osgood, C. E., and Bock, J. K. (1977). “Salience and sentencing: some production principles,” in *Sentence Production: Developments in Research and Theory*, ed. S. Rosenberg (Hillsdale, NJ: Erlbaum), 89–140.
- Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* 42, 107–123.
- Pickering, M. J., and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychol. Bull.* 134, 427–459.
- Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25.
- Prat-Sala, M., and Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: a cross-linguistic study in English and Spanish. *J. Mem. Lang.* 42, 168–182.
- Prentice, J. L. (1967). Effects of cuing actor vs. cuing object on word order in sentence production. *Psychon. Sci.* 8, 163–164.
- Rasolofo, A. (2006). *Malagasy Transitive Clause Types and Their functions*. Ph.D. manuscript, University of Oregon, Eugene.
- Roland, D., Dick, F., and Elman, J. L. (2007). Frequency of basic English grammatical structures: a corpus analysis. *J. Mem. Lang.* 57, 348–379.
- Salverda, A. P., and Altmann, G. T. M. (2011). Attentional capture of objects referred to by spoken language. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1122–1133.
- Soto, D., and Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 730–737.
- Sridhar, S. N. (1988). *Cognition and Sentence Production: A Cross-linguistic Study*. New York: Springer-Verlag.
- Svartvik, J. (1966). *On Voice in the English Verb*. The Hague: Mouton.
- Tannenbaum, P. H., and Williams, F. (1968). Generation of active and passive sentences as a function of subject or object focus. *J. Verbal Learn. Verbal Behav.* 7, 246–250.
- Tomlin, R. S. (1995). “Focal attention, voice, and word order,” in *Word Order in Discourse*, eds P. Downing and M. Noonan (Amsterdam: John Benjamins), 517–552.
- Tomlin, R. S. (1997). “Mapping conceptual representations into linguistic representations: the role of attention in grammar,” in *Language and Conceptualization*, eds J. Nuyts and E. Pederson (Cambridge: Cambridge University Press), 162–189.
- Turner, E. A., and Rommetveit, R. (1968). Focus of attention in recall of active and passive sentences. *J. Verbal Learn. Verbal Behav.* 7, 543–548.
- Vigliocco, G., and Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychol. Bull.* 128, 442–472.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 September 2011; accepted: 22 December 2011; published online: 18 January 2012.

Citation: Myachykov A, Thompson D, Garrod S and Scheepers C (2012) Referential and visual cues to structural choice in visually situated sentence production. *Front. Psychology* 2:396. doi: 10.3389/fpsyg.2011.00396

This article was submitted to *Frontiers in Cognition*, a specialty of *Frontiers in Psychology*.

Copyright © 2012 Myachykov, Thompson, Garrod and Scheepers. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.