Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 57 (2007) 544–569
Journal of Memory and Language
www.elsevier.com/locate/jml

On the *give* and *take* between event apprehension and utterance formulation ☆

Lila R. Gleitman, David January, Rebecca Nappa, John C. Trueswell *

Department of Psychology, Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, Room 302C, Philadelphia, PA 19104-6228, USA

Received 2 July 2006; revision received 22 January 2007
Available online 16 July 2007

Abstract

Two experiments are reported which examine how manipulations of visual attention affect speakers' linguistic choices regarding word order, verb use and syntactic structure when describing simple pictured scenes. Experiment 1 presented participants with scenes designed to elicit the use of a perspective predicate (*The man chases the dog*/*The dog flees from the man*) or a conjoined noun phrase sentential Subject (*A cat and a dog*/*A dog and a cat*). Gaze was directed to a particular scene character by way of an attention-capture manipulation. Attention capture increased the likelihood that this character would be the sentential Subject and altered the choice of perspective verb or word order within conjoined NP Subjects accordingly. These effects occurred even though participants reported being unaware that their visual attention had been manipulated. Experiment 2 extended these results to word order choice within Active versus Passive structures (*The girl is kicking the boy*/*The boy is being kicked by the girl*) and symmetrical predicates (*The girl is meeting the boy*/*The boy is meeting the girl*). Experiment 2 also found that early endogenous shifts in attention influence word order choices. These findings indicate a reliable relationship between initial looking patterns and speaking patterns, reflecting considerable parallelism between the on-line apprehension of events and the on-line construction of descriptive utterances.

Keywords: Sentence production; Word order; Visual attention; Attention capture; Eye movements

People seem to think before they speak: Having understood and conceptualized some event or state of affairs, they construct and utter some phrase or sentence to describe it. On this picture, the relationship between apprehension and linguistic formulation is sequential, incremental, and causal. But is the progression from thought to speech always as tidy as this? Words sometimes seem to start tumbling forth before we fully apprehend a scene or organize our thoughts about it. In light of these contrasting intuitions, it is perhaps not surprising that debate concerning the timing and information characteristics of apprehension and linguistic formulation has a venerable psycholinguistic history. (See Bock, Irwin, & Davidson, 2004, for a recent review of the literature, which dates back most notably to Lashley, 1951; Paul, 1886/1970; and Wundt, 1900/1970; and also includes the recent experimental literature on sentence

☆ Authorship order was determined alphabetically by last name. We thank Katherine McEldoon for her helpful comments and assistance on drafts of this paper. This work was partially funded by a grant to L.R.G. and J.C.T. from the National Institutes of Health (1-R01-HD37507).
* Corresponding author. Fax: +1 215 898 7301.
E-mail address: trueswel@cattell.psych.upenn.edu (J.C. Trueswell).

0749-596X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2007.01.007

production.) Here we examine cases in which the apprehension of the visual world and the production of an utterance describing it suggest a surprisingly tight temporal coupling between perceptual and linguistic processes.

Factors controlling sentence production

Obviously there are several veridical ways to describe any single scene. For example consider Fig. 1.

Any of the following utterances (some of which are more natural than others) adequately describe this scene.

1	a. *A dog is chasing a man.*	b. *A man is running away from a dog.*
	c. *A dog is pursuing a man.*	d. *A man is fleeing a dog.*
	e. *A dog is being fled from by a man.*	f. *A man is being chased by a dog.*

How does the speaker choose among these options?

Beginnings

Several aspects of this problem can be characterized as "starting point questions" because the first-mentioned word or phrase constrains both the form and content of the remainder of the utterance (Bock et al., 2004). For instance, speakers typically begin their description of Fig. 1 with one of two noun phrases (henceforth, NP): *A man*. . .(as in 1b, d, f) or *A dog*. . .(as in 1a, c, e). This choice is often characterized as hinging, at least in part, on some notion of accessibility which itself branches into several subtypes.

One level of accessibility is perceptual and concerns just where the speaker's eyes land first—on the dog or the man. Plausibly, this property of initially inspecting the scene could have a corresponding influence on what is mentioned first. Effects of such visual landing sites have been studied indirectly in experiments in which attentional focus is drawn to a particular character. Notably, Tomlin (1997) repeatedly showed participants short cartoons of one fish eating another. Throughout, an arrow pointed to a particular fish and participants were to keep their eyes on that fish during the presentation. Under these conditions participants tended to mention the indicated fish first, choosing it as the Subject even when this meant using the ordinarily disfavored Passive structure (e.g. *The red fish is being eaten by the blue fish*). Thus, at least in some highly constrained situations, there appears to be an influence on sentence formulation of prior or simultaneous visual attention to some particular individual in the scene.

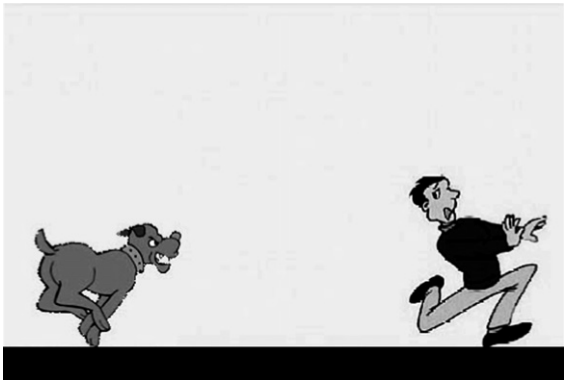

Fig. 1. A sample Perspective Predicate scene, depicting the verb pair *chase/flee*.

However, word order is also responsive to higher-level accessibility factors, and these may weaken or even obliterate any effect of first visual landing-site. For instance, some constructional types are preferred to others, e.g., all other things held equal, Active voice sentences (1a, b, c, d) are strongly favored over passives (1e or f) unless specific presuppositional supports are provided (e.g., Bock, 1986; Bock & Loebell, 1990; Slobin & Bever, 1982). Related accessibility distinctions hold on the conceptual side: For example, creatures higher in an animacy hierarchy tend to be in Subject position making 1b, d, and f preferred over 1a, c, and e (Dowty, 1991; see also Bock, 1986). These conceptual and linguistic preference factors themselves interact. Because frequent words are favored over infrequent ones and *chase* is a more frequent lexical item than either *pursue* or *flee*, this might promote the use of *chase* (1a or f) over the other descriptions (Griffin & Bock, 2000). Such a tendency may be enhanced by a semantic bias, across predicates, to favor descriptions in which the source is the logical Subject (1a, c, or f) and the goal is the Object over goal-to-source descriptions (1b, d, or e; Fisher, Hall, Rakowitz, & Gleitman, 1994; Lakusta & Landau, 2005).

Apprehension of gist

In summarizing factors guiding utterance formulation, we have so far implicitly envisaged an incremental process in which an utterance-initiating word or phrase—the "starting point"—is chosen, and further effects on the sentence are constrained by this first choice. But this picture is at best oversimplified and may even be a false characterization. That is, the so-called starting points may themselves be effects of a prior global apprehension of the scene in view, i.e., its conceptual-semantic gist. As Bock et al. (2004) have recently put this:

"What cements a starting point is not the relative salience of elements in the perceptual or conceptual underpinnings

of a message, but the identification of something that affords a continuation or completion; that is, a predication.'' (Bock et al., 2004, p. 270)

Indeed, the experimental evidence for visual-attentive factors guiding nominal (or other ''elemental'') starting points is quite weak. Consider again Tomlin (1997). In this experiment, an arrow superimposed on the picture told the participants which fish to look at, and they were instructed to maintain this fixation throughout the generation of their utterance. This rather blatant manipulation of attention leaves open the possibility that participants were aware of the intention of the study, thus producing the expected findings in contravention of their behavioral tendencies under more neutral conditions. Moreover, repeated description of the same event (all trials were fish-eating-fish events) essentially precludes generalization (see Bock et al., 2004, for discussion of this point). And the repetition of the fish-characters across trials might itself create confounds. For example, inspection of the Tomlin (1997) videos (available on the web at http://logos.uoregon.edu/tomlin/research.html) reveals that the cued fish on any given trial (e.g., the red fish) was always present on the immediately preceding trial, but the uncued fish (e.g., the blue fish) was never present on the previous trial. Thus, the cueing of a particular fish was perfectly confounded with which fish had been mentioned most recently by the participant. Given that recent mention of an entity promotes Subject status on its own, it is entirely plausible that this discourse factor, rather than attentional cueing, was determining the speakers' choice.

In fact, subsequent studies (Bock, Irwin, Davidson, & Levelt, 2003; Griffin & Bock, 2000) suggest that eye position may not be a cause of word order choice, but rather an artifact generated as a consequence of the more global semantic analysis of the scene. In the words of Bock et al. (2003)

''*...when speakers produce fluent utterances to describe events, the eye is sent not to the most salient element in a scene, but to an element already established as a suitable starting point.''* (Bock et al., 2003, p. 680).

Bock et al. (2003) based this conclusion on experiments (Bock et al., 2003; Griffin & Bock, 2000) that employed a task similar to Tomlin's (1997) except that no visual cues or attention instructions were used. Instead, participants' eye movements were recorded as they carried out this task. When coupled with the content and the timing of the utterances, such eye movements can provide a strikingly fine-grained measure of the relationship between visual apprehension and linguistic formulation.

In Griffin and Bock (2000), participants viewed and described line drawings depicting simple agent-patient events such as the one in Fig. 2. In English, there is room for choice as to which character to mention first while preserving the general meaning of the sentence because the scene in Fig. 2 can be described with an Active *The girl is spraying the boy* or a Passive sentence *The boy is getting/being sprayed by the girl*.[1] Of course English speakers are disinclined to utter Passive-voice sentences; so to increase the likelihood of Passive production, three (out of a total of eight) stimulus pictures (which were then mirrored and role-traded to create four stimulus lists) involved one human and one non-human character. Because human characters tend to appear as sentential Subjects, this increased the number of Passive descriptions when the human participant was the Patient of the action.

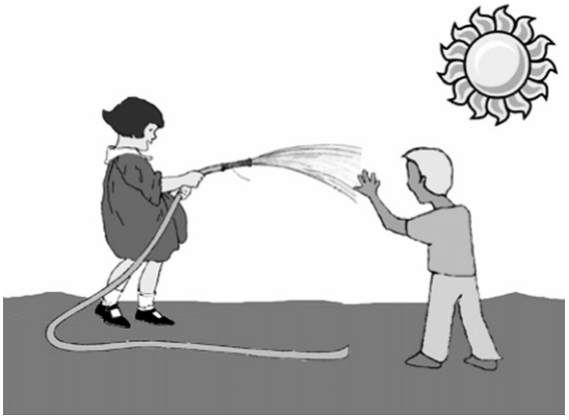

Fig. 2. A figure similar to those used by Griffin and Bock (2000), depicting a simple transitive action (*A girl spraying a boy*).

Griffin and Bock (2000) reasoned that if output from the early production stages involving *apprehension* of an event were expeditiously passed along to the later stages geared towards *formulation* of a linguistic characterization, then initial fixations to characters (and their sequential ordering) should be predictive of their

[1] Griffin and Bock (2000) manipulated experimentally which character played which role. Therefore as a between-Subject variable there was a ''role reversal'' variant for each of the eight original pictures (Each picture and its role-reversed variant are herein called a ''stimulus type.'') For the present example (Fig. 2), the role-reversed picture would have shown a boy spraying a girl/a girl being sprayed by a boy. It should also be mentioned that there was a single example of a perspective-verb pair type; namely a scene of chasing/fleeing (see Fig. 1 for an equivalent used in our own experiments). For such verbs there is a non-Passive Patient-first alternative, namely *flee* or *run away from*, thus potentially unconfounding constructional and first-mention factors. The flee-type responses were collapsed together with the Passives for analysis in these experiments, though their form is probably always Active voice (''was run away from by'' and ''was fled from by'' being awkward and therefore unlikely locutions).

description order. On the other hand, if the initial conceptualization depended solely upon the processes involved in apprehending the relations between characters in a scene, and linguistic considerations became a factor only later in the process, initial fixation on one character or the other would not predict which is mentioned first (and hence which is placed in grammatical Subject position, whether in an Active or Passive frame).

Results of Griffin and Bock's analyses supported the latter prediction: Speakers almost always uttered Active sentences, and the first 300 ms of the eye movement record showed no significant difference between looking times to the character that would ultimately be mentioned first versus the one that would be mentioned second; initial fixation on one character or the other did not predict Subjecthood in the upcoming utterance. A large difference in looking patterns did emerge beyond 300 ms after visual inspection began: Subject-referents were fixated more than Object-referents just prior to speech onset, and the opposite was true just after speech onset.

Griffin and Bock interpreted this pattern as consistent with a rapid initial apprehension period, during which the gist of the event is extracted. In their words,

"The evidence that apprehension preceded formulation, seen in both event comprehension times and the dependency of grammatical role assignments on the conceptual features of major event elements, argues that a wholistic process of conceptualization set the stage for the creation of a to-be-spoken sentence." (Griffin & Bock, 2000, p. 279).

Additional support for this conclusion was found in results from a separate group of participants who viewed the same pictures but were instead asked only to select the character being acted upon (the Patient). Here, eye-movements diverged between Patient and Agent approximately 300 ms into viewing. Given that Patient selection requires event apprehension, the data suggest that it is possible to achieve this gist-extraction process in the first 300 ms of viewing these stimuli.

These and subsequent supportive studies (Bock et al., 2003) suggested to the authors not only a separation of apprehension and formulation processes but a clear temporal dissociation as well. As Griffin and Bock (2000) put this,

"The results point to a language production process that begins with apprehension or the generation of a message and proceeds through incremental formulation of sentences" (Griffin & Bock, 2000, p. 279).

Open issues

Despite these useful findings, the current literature leaves a number of issues concerning utterance planning unresolved. Specifically, neither the more serial nor the more interactive accounts that have been proposed delve too deeply into questions involving the conceptualization stage itself. Many otherwise sequential models (e.g., Levelt, 1989; Levelt, Roelofs, & Meyer, 1999) allow for feedback between the conceptual stage of sentence planning and lemma representation, for example. Research exploring the question of the conceptual factors underlying word order choices has implicated variables such as concreteness, predicability, and particularly animacy as driving forces in Subject role assignment (see MacDonald, Bock, & Kelly, 1993, for a discussion) but has not investigated the time course with which any such conceptual factors contribute to the process of selecting thematic and syntactic roles when producing an utterance. For example, as one is apprehending a man participating in some event (an event not yet specified at the message level), will the production system generate a lemma candidate MAN to participate in the yet-to-be-determined proposition? Or is further apprehension of the relational information relevant to the man (e.g., Is he wearing a red hat? Or near a bicycle?) necessary before such linguistic planning can begin? Griffin and Bock (2000) endorse the latter account and support it with the aforementioned finding: Early fixations (in the first 300 ms of viewing a scene) in their studies simply did not predict the order in which fixated characters were mentioned in a descriptive sentence.

Griffin and Bock's results are, however, surprising not only in light of Tomlin (1997) but also from findings in the perception literature suggesting that initial gaze direction can exert a powerful influence on the outcome of the apprehension process itself (Ellis & Stark, 1978; Gale & Findlay, 1983; Pomplun, Ritter, & Velichkovsky, 1996). For instance, manipulation of a perceiver's first fixation influences his/her interpretation of ambiguous figures (Georgiades & Harris, 1997). In this study, participants viewed ambiguous images such as the classic mother-in-law/wife image, each of which had been preceded by a fixation crosshair that was designed to direct initial attention to certain aspects of the image. Attending first to visual features that are critical to the mother-in-law interpretation increased reports of a mother-in-law, and mutatis mutandis. These features of the scene are independent of any general salience factors having to do with mothers-in-law or wives, or, apparently, with visual properties of mothers-in-law and wives as portrayed in this image. Rather, the finding suggests, much as do Tomlin's findings, that what you first look at becomes, in virtue of that, the focus of your attention.

A related study concerns how attention influences the assignment of perceptual Figure and Ground. Vecera, Flevaris, and Filapek (2004) presented participants with simple images such as the one depicted in

Fig. 3. An image used by Vecera et al. (2004) to investigate contributions of attention to figure–ground assignment in visual perception.

Fig. 3. This image is ambiguous in that it can be interpreted either as a gray figure on a black background or as a black figure on a gray background. In the experiment, participants' attention was captured to one part of the image via a brief (50 ms) flash that accompanied stimulus onset. Such a cue is known to draw a participant's eye movements in a way that is rarely noticed by the participant (McCormick, 1997). Interestingly, Vecera et al. found that the cued region was more likely to be subsequently interpreted as the Figure.

In sum, the perception literature suggests that endogenous and exogenous contributions to initial attention can generate changes in interpretation of an image and even the assignment of Figure–Ground. In contrast, there was no trace of such an effect in Griffin and Bock (2000), seeming to suggest that the speaker's visual attention (as indexed by initial fixation and early looking-time preference) and his/her subsequent speech behavior (as indexed by first-mentioned character) are divided by a conceptual firewall that reorganizes the observed event for the sake of speech under quite different influences. This may simply be the fact of the matter, but the mismatch between these literatures provides at least some impetus for further investigation.

Indeed, it is important to reiterate that, thus far, published eye movement analyses of depicted events are currently limited to Griffin and Bock (2000), who studied just 8 pictorial items (and their role-traded variants). And these were so constructed that, with a single exception (the *chase/flee* example), they required participants to utter Passive-type sentences as the only environment in which to show effects of initial attention. But we know that English speakers, on independent grounds, tend to disfavor the Passive in speech (e.g., Slobin & Bever, 1982; Goldman-Eisler & Cohen, 1970). This imbalance in constructional preference rather than (or in addition to) any tendency to sequence utterance formulation may have accounted for the experimental results. The bias to utter a canonical Active-voice sentence may have overwhelmed any observable effects of initial attention.

Stimulus types used in the present study

Following the methodology of Griffin and Bock (2000), in the present study we asked participants to describe novel depicted scenes, but these were designed to elicit various kinds of linguistically different but semantically equivalent utterances. (By "semantically equivalent," we mean two utterances that have roughly equivalent meanings, but may have different discourse or focusing properties.) Such types allow us to see what is driving linguistic choice when the conceptualization of the event is held constant (or close to constant). In addition to the Active/Passive alternation we examined three further productive word-order alternations. Each is exemplified in Fig. 4.

We chose these linguistic alternations because, although they are all semantically equivalent, each type differs in the extent to which the alternatives share the same linguistic-structural forms, the same discourse implications (e.g., Given vs. New), and the same information structure (e.g., Figure versus Ground).

(1) Active/Passive Pairs are often put forth as the classic structural alternation in English that preserves propositional meaning: If the cat drinks the milk, it follows that the milk is drunk by the cat. Not only are Active/Passive pairs usually semantically equivalent descriptions of events,[2] they are both descriptions of the very same event. It strains credulity to suppose, for example, that Jane could observe the cat drinking the milk while George simultaneously observes that (very) milk being drunk by that (very) cat, and yet the two of them are observing "different events." However, these alternative forms differ considerably in other regards that may be relevant in linguistic processing tasks, e.g., the Active form is more frequent than the Passive, less complex, acquired earlier, and more accessible.

(2) Perspective Predicates describe the same scene from the standpoint of one or the other character in the event. For Fig. 4B (repeated here from Fig. 1

[2] We say that Passivization only "usually" yields a semantically equivalent sentence because, among other exceptions, it notoriously interacts with quantification; thus *Every boy kissed at least one woman* does not entail that *At least one woman was kissed by every boy*. Stimuli in this experiment do not implicate such problems.

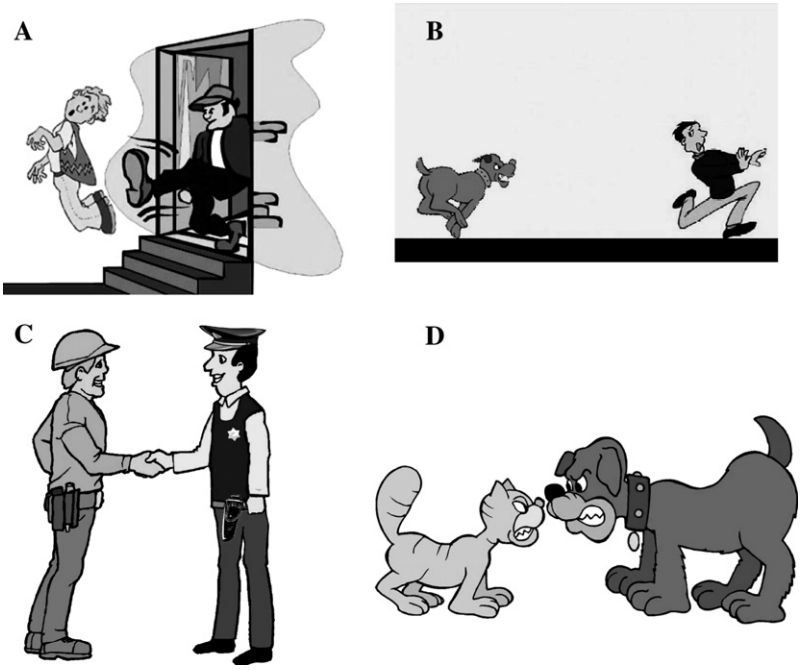

Fig. 4. Example stimulus scenes used in Experiments 1 and 2; Active/Passive (A), Perspective Predicate (B), Symmetrical Predicate (C), and Conjoined Noun Phrase (D).

for expositional clarity), one can differentially frame the *verbal description*, taking either the perspective of what the dog is doing (chasing) or what the man is doing (fleeing, or running away). In contrast, under most circumstances the chasing and fleeing *events* themselves cannot be decoupled. If the man opts to stop and confront the dog, he is no longer fleeing, and the dog can no longer be said to be chasing him (for a more detailed description of these framing structures, see Gleitman, 1990).

Many perspective predicates exist in English, including *buy/sell*, *chase/flee*, *win/lose* and *give/receive.* For each pair, both members support the use of the Active canonical form, making moot the structural and discourse constraints discussed earlier for the Passive. As such, perspective-taking scenes are ideal candidates for investigating how the speaker's attentional state influences and interacts with linguistic formulation. To the extent that attention is focused on the "chasingness" in Fig. 4B, the dog is promoted to Subject position; insofar as one is attending to "fleeingness," the man is necessarily the Subject.

It should be noted that for almost all of these predicate pairs, in the absence of extra contextual and presuppositional information, speakers have a clear preference for one event description over the other (Fisher et al., 1994). The items used here were therefore normed in advance to verify that both alternatives were readily available to the speaker (see below).

(3) Symmetrical Predicates are often conveyed using verbs such as *match*, *meet*, *argue*, and scores of others that under linguistically specifiable conditions obey the symmetrical entailment (for all x, y, $R(x, y)$ iff $R(y, x)$; see Tversky, 1977, on *similar/different* and Gleitman, Gleitman, Miller, & Ostrin, 1996, for an analysis of the class of symmetrical predicates). Symmetrical verbs are recognizable by their appearance in plural (but not singular) intransitive structures (e.g., *The men met; John and Bill met*, and with reciprocal inference structure, that is, with rough equivalence to *The men met each other; John and Bill met each other*). Notice that non-symmetricals differ from symmetricals both by not requiring the plural Subject (e.g., *John fled* sounds fine whereas *John met* is awkward or anomalous) and by never implying reciprocity (symmetrical entailment) without overt *each other* (e.g., *John and Bill fled* does not imply that they fled from each other; rather, that each fled from somebody or somewhere else).

Symmetrical Predicates (universally, across languages) permit framing effects analogous to those just noted for the Perspective Predicates, and it is the Symmetricals in these framing environments that we are studying here. That is, symmetrical predicate alternations make framing distinctions by reversing the structural position of nominal arguments. Thus Fig. 4C shows a policeman and a construction worker shaking hands. The relation "shaking hands" is necessarily true of the pair of men, i.e., this predicate obeys the symmet-

rical entailment.[3] Even so, the sentence *A policeman shakes hands with a construction worker* frames this relation in terms of the policeman, and *A construction worker shakes hands with a policeman* frames it in terms of the worker. Talmy (1978) aptly borrowed the terms Figure and Ground from perception research to describe the conceptual effect of these ways of expressing and interpreting (logically) symmetrical states and events; speakers place Ground information in the predicate, while the Figure is preferentially the Subject (for the experimental proof, see Gleitman et al., 1996).

Note that the difference in which character is treated as the Figure vs. Ground of the event is also distinguished as well in the two other word-order alternations already discussed (Active/Passive and Perspective Predicates). In each case the nominal that captures grammatical Subject position is what the sentence is "about," the Figure to the complement's Ground. However, as already noted, the Perspective Predicates differ also in the verb lexical item (*chase* versus *flee*) and the Active/Passive pairs differ in syntactic form and discourse requirements.

(4) Conjoined NPs: This type consists of sentence pairs differing only in how two nominal phrases are sequenced around the conjunction *and* in sentences which describe a joint (but not symmetrical) activity, e.g., *The cat and the dog*/*The dog and the cat are growling at each other*. Semantic and discourse factors (such as animacy and concreteness/imageability) have been found to have little or no influence on ordering of NPs in these conjunctive phrases (Bock & Warren, 1985; Kelly, 1986; MacDonald et al., 1993), leaving aside the special case of their use in Symmetricals as discussed in the preceding section. Rather, only form-related factors (length, prosodic and frequency differences between the conjuncts) appear to affect ordering, such that the more accessible lexical item tends to be mentioned first (Bock, 1987; Cooper & Ross, 1975; Fenk-Oczlon, 1989; Kelly, 1986; MacDonald et al., 1993). Semantic effects have been found but seem to be reducible to issues of lexeme accessibility (Kelly, Bock, & Keil, 1986; Osgood & Bock, 1977). Thus the order of NPs in Conjunctions with *and* offers a test case of highly flexible ordering in English, in which the order of mention plays little or no communicative role and has no stable syntactic or semantic consequences. To compare these with our other stimulus types (Sections 1, 2, 3 above): (1) Unlike for Active/Passive pairs, there are no syntactic or discourse differences between alternative word orders in NP Conjunction; unlike for Perspective Predicates, there is no difference in the lexical heads of Conjoined NPs (*and* is used regardless of the order), and (2) unlike for Active/Passive pairs, Perspective Predicates and Symmetrical Predicates, Conjoined NP alternates do not differ in Figure/Ground assignment.

[3] The linguistic-interpretive properties of symmetricals show considerable complexity, with event predicates (e.g., *kiss*) showing constraints that the formal stative predicates (e.g., *equal, match*) do not. For example, if the shirt matches the button, then it is true (framing effects aside) that the button matches the shirt and that the shirt and the button match, and match each other. On the other hand, if John and Mary kiss each other, it does not always follow that they kiss (if John kisses Mary's hand and simultaneously or even sequentially Mary kisses John's hand, then they kiss each other but do not kiss). Our stimuli always depicted symmetrically interpretable events (as it were, symmetrical kissing).

Attention manipulation and predictions

The present experiments explore how speakers' initial attentional state influences their description choice for the aforementioned stimulus types. The onset of each of these stimuli was preceded by a manipulation of the speaker's attention, using techniques reminiscent of Tomlin (1997) but far less overt or reportable. Across Experiments 1 and 2, we used the attention capture technique of a sudden onset, which is undetectable to the speaker but nevertheless influences initial saccades to characters (similar to Vecera et al., 2004). Eye movements in these experiments were also recorded (for the sake of brevity, however, detailed eye-movement analyses are only presented for Experiment 2, as findings were largely similar across both experiments).

The predictions for these experiments differ depending on what one believes to be the relationship between the apprehension of an event and the formulation of a description of that event. According to Bock and colleagues, effects of attention should be small and difficult to replicate across stimulus types (see Bock et al., 2003). Such a finding would be consistent with the view that linguistic factors are the main determinant of word order choice and would be in line with Griffin and Bock's (2000) observation that initial eye position did not predict word order.

If, however, one accepts the view suggested by the perception literature that the perceiver's initial attentional state influences the apprehension outcome itself and that in particular it influences the assignment of Figure and Ground, attentional manipulations should have an effect on only those stimulus types whose word-order pairs contrast Figure and Ground in their use: Active/Passives, Perspective Predicates, and Symmetricals. As discussed above, the first NP in each of these three types is in a structural position (Subject) that communicates what is perceived as the Figure, or aboutness, of the event, whereas the second NP is in a syntactic position that communicates Ground. The alternative orderings within Conjoined NPs do not contrast Figure/Ground (both NPs remain in Subject position); as such, manipu-

lations of a speaker's attentional state should not influence ordering choices in these stimuli.

Finally, it is possible that attentional manipulations influence more than just Figure/Ground assignment: If visual apprehension and linguistic formulation processes are tightly coupled, initial attention to a character should immediately increase the accessibility of the corresponding lemma (looking to a dog will activate the lemma DOG). In such an incremental interactive system, our manipulations of initial attention are expected to influence all stimulus types including Conjoined NP constructions. Such a finding would be at odds with the conclusions of Griffin and Bock (2000), who suggest that that apprehension and linguistic formulation are actually dissociable at this time scale.

Experiment 1

In this investigation of attentional effects on event interpretation and description, participants viewed still pictures that are naturally described using sentences containing Perspective Predicates or Conjoined NP Subjects. We captured participants' attention to one character or the other in these pictures by preceding each image by a sudden-onset, briefly flashed spatial cue (Jonides & Yantis, 1988). The participants' eye movements were recorded (using a remote eyetracker) along with the utterances they used to describe the pictures.[4]

Methods

Participants

Thirty-six monolingual students in an Introductory Psychology course at the University of Pennsylvania participated in the study in return for course credit.

Stimuli

Picture norming study. In order to select the images for the present experiment, a pencil and paper norming study was first conducted on a separate group of 21 monolingual English speakers. These participants wrote down a single sentence description for each of 52 images that had been designed to elicit either Perspective verb or Conjoined Subject sentences. Each image consisted of a simple color cartoon drawing depicting an event involving one or more characters and objects. The experimenters created these images by altering clip art images within a professional image editing software package. Target images for the primary experiment were selected from this larger set based on their flexibility in eliciting both alternations from norming participants (i.e., choice of verbs for the Perspective Predicate items and noun phrase order for the Conjoined NP items). Specifically, each alternative had to occur at least once among the sample of descriptions for that image. In addition, Target images had to have a very low rate of eliciting uninformative Subjects (e.g., *A man is chasing another man* or *Two people are running*). The result was that all selected Target pictures contained characters that participants routinely and spontaneously distinguish in their descriptions. That is, pictures typically contained two humans of different gender (e.g., a boy and a girl), two humans of different occupations (e.g., a policeman and a construction worker) or two different animals (e.g., a man and a dog, or a horse and a pig).

This allowed us to select 12 Perspective Predicate items and 12 Conjoined NP items. Rates of first-mentioned scene character for the pairs of items varied (see Appendix A) but for each Perspective Predicate item, participants showed some degree of bias toward one interpretation and/or verb choice; there was a Preferred Verb and a Dispreferred Verb, and hence a corresponding Preferred Subject and Dispreferred Subject. (Among the sentences that norming participants produced for the 12 Perspective Predicate Items, passives were rare, occurring only 6 times across all 252 sentences.) Preferred Subjects and verbs were produced by our norming participants 69% of the time, dispreferred Subjects and Verbs were used 27% of the time (the remaining 4% were uncodable, e.g., *A race between some animals*). Unlike the case for Perspective Predicates, baseline rates of first-mentioned scene characters in Conjoined NP stimuli did not show a bias for one character over the other, so scene characters were arbitrarily dubbed Character A and Character B for the coding and data analysis stages (for which, see Appendix A). An additional factor driving word order in Conjoined NPs, but not Perspective Predicates, was a bias to mention the leftmost depicted character first (Flores d'Arcais, 1975). This effect was seen in the current norming study too, with leftmost characters mentioned first 78% of the time for conjoined NP items (as compared to 53% of the time for perspective items).

Experimental stimuli. Sixty-four images (consisting of the 12 Perspective Predicate items, the 12 Conjoined NP items, and 40 Filler items) were used in the primary experiment. Fig. 4 presents an example Perspective Predicate image (Fig. 4B) and an example Conjoined NP image (Fig. 4D). The Filler images were taken from

[4] Two other pilot investigations of different attentional manipulations—a crosshair like the one used by Georgiades and Harris (1997) and a gaze-following manipulation—used only the Perspective Predicate items to determine how various manipulations of attention influence predication. Both investigations (with fewer stimulus items and participants) showed non-significant trends in the same direction as the findings reported in Experiments 1 and 2: Attentional manipulations drove scene interpretations and word order choices.

prior norming studies and pilot studies, and looked similar to the Targets in artistic style. Fillers were designed so as not to elicit high rates of either Perspective Predicates or Conjoined NPs.

Procedure and design

Participants sat approximately 18 in. from a 17-in. monitor, set to 1024×768 pixels, with a refresh rate of 75 Hz. The space onscreen in which scenes were presented was approximately 10.2×13.6 in., subtending approximately 32° of visual angle horizontally and 42° of visual angle vertically. Scenes varied to some degree in size, but they typically occupied most or all of this 10.2×13.6 in. of space.

On each trial, participants were first presented with a crosshair (which they had been instructed to fixate), which appeared for approximately 500 ms and was neutrally located between scene characters. This fixation point was then followed by a brief attention-capture manipulation. This consisted of a small black target area (subtending an area of approximately $0.5° \times 0.5°$ of visual angle) against a white background, onscreen for 60–80 ms, followed immediately by the stimulus (see Fig. 5 for a demonstration of stimulus presentation). Although no participant reported noticing the subliminal cue (see below), it was highly effective in capturing attention: Across Experiments 1 and 2, participants looked first to the cued location approximately 75% of the time.

Both the location of the attention-capture cue and the left-to-right orientation of the scene were counterbalanced across four stimulus lists. Manipulations were within-participants, with each participant randomly assigned to one of these four lists.

Scenes were presented (and randomized) by E-prime version 1.0 software, which progressed by way of a button-press from participants (i.e. participants were under no time pressure and paced themselves). An ISCAN tabletop remote eye-tracker system was used to collect and store eye-tracking data, which consisted of the participants' eye position sampled at 60 Hz (approximately 17 ms intervals). The scene image and the superimposed eye position, along with all auditory stimuli (i.e. participants' utterances), were recorded by a frame-accurate digital video recorder (a SONY DSR-30).

Coding and analyses

Perspective Predicates. Transcriptions for each of the 432 Perspective Predicate trials (12 targets, 36 participants) were analyzed for choice of Preferred or Dispreferred Subject. Trials containing disfluencies (e.g., *um, uh*) were not excluded from analyses—as natural speech contains significant numbers of disfluencies, and these investigations were aimed at approximating the factors at play in normal sentence production. Utterances containing repairs that altered word order were excluded (e.g. *A dog, um, or, a man is running from a dog*). As we were most interested in the position of scene characters as constituents in the utterance, we included utterances that did not contain either of the most commonly used verbs (e.g. *chase/run away* for the scene in 4b). For example, if a participant said for 4b *A man is scared of a dog*, this would have been coded as a Preferred Subject utterance, just as if the participant had said, *The man is running away from the dog*. If participants produced both forms of the description we considered only the first clause in our analyses; thus, an utterance like, "A dog is chasing a man, who is running away from him" was coded as a Dispreferred Subject Utterance. We excluded any utterances that did not contain both NPs (e.g. *Two people are having a conversation* or *This is a boxing match*). We also excluded any utterances that did not contain a Subject, Verb and Object (e.g. *The dog and the man are running*), as very different information comes into play when ordering NPs within a NP conjunction and within a sentence containing such a conjunction (e.g. thematic role assignment takes place in the latter case). Lastly, a handful of trials were excluded due to experimenter or participant error (e.g. audio recordings were not intact, or the participant mistakenly button-pressed and skipped a trial). Just over 14% of trials (61 trials) were removed from further analysis for one of the above reasons.

Conjoined NPs. Transcriptions for each of the 432 target Conjoined NP trials (12 targets, 36 participants) were analyzed for the order of NPs within the conjoined NP Subject. Because this was the experimental focus, coding ignored other differences in sentence structure and verb choice (e.g., coded equivalently as "Participant

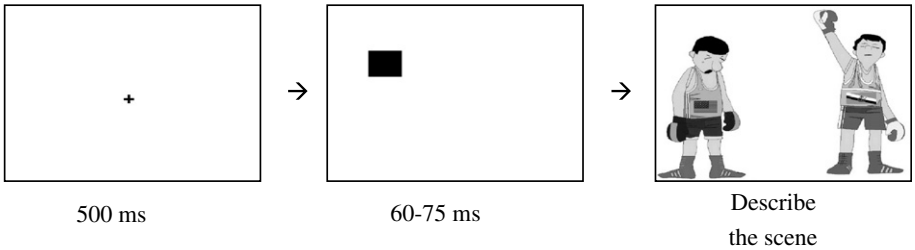

Fig. 5. Display sequence for Experiments 1 and 2. Participants saw the first panel (on the left) for 500 ms, the second panel (in the middle) for 60–75 ms, and then viewed the scene (on the right) and described the event taking place therein.

B mentioned first'' were the responses *''A dog and a cat are growling at each other,'' ''A dog and a cat are nose to nose, about to go at it,'' and ''There's a dog and a cat.''*). Trials containing disfluencies or false starts were included here too, unless the change altered word order (e.g. if a participant said, ''A dog, um, or, a cat and a dog are about to fight''), in which case the item was excluded. Lastly, a handful of responses were excluded due to experimenter or participant error. Just over 10% of conjoined NP trials (44 trials) were removed for one of the above reasons.

Results

Post-experiment questionnaire

A post-experiment questionnaire was administered to every participant. This began by asking what participants ''thought the experiment was about,'' and increased in specificity to a final question as to whether they had noticed any kind of flash or disruption in the presentation of the scenes in the experiment. No participant reported being aware of the attention capture cue, and all were quite surprised to discover that attention had been manipulated.

This lack of awareness may seem surprising in light of prior research on RSVP (Rapid Sequential Visual Processing), in which participants demonstrate the ability to recognize a visual stimulus as part of the set of visual stimuli presented during an earlier very rapid (each item presented for 80 ms) stimulus-presentation session (Rosenblood & Pulton, 1975), and recent findings on URVC (Ultra-Rapid Visual Categorization), in which participants are able to determine whether a natural scene contains an item within a particular category (e.g. an animal, or a vehicle) with as little as 20 ms of scene presentation time (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). One must consider the task demands of these different investigations, however. Our task consists of a sort of backwards masking, which results in perceptual interference and a general disruption in participants' ability to detect an otherwise quite visible stimulus (the attention cue—a black square against a white background, subtending approximately $0.5 \times 0.5^{\circ}$ of visual angle). In almost all investigations of backwards masking, participants are instructed that a ''target'' will be present, and their task it to attempt to detect it (see Breitmeyer & Ogmen, 2000, for a review of the backwards masking literature). In our task, however, the target corresponds to our subtle attention-manipulating cue, which is never addressed pre-experimentally. That is, participants in our experiment had no expectation that a ''target'' would be present at all. Additionally, each filler scene was preceded by a ''flicker,'' during which the scene that they were about to describe appeared onscreen, disappeared briefly (for approximately 60 ms), then reappeared. Thus, our target trials did not stand out in their presentation as appearing to have any special visual disruption prior to scene onset.

Moreover, it is important to note that the attention capture cue had always been preceded by a crosshair, displayed for only 500 ms and located equidistant from each of the two possible cue positions. Participants had been instructed to fixate this crosshair, and as such eye position and attention were located away from the attention capture cue at the time of its presentation. It is well established that the detection of a ''target'' object (such as our briefly displayed square) diminishes when the locus of attentional resources is directed to a location other than where a target will appear (Bashinski & Bacharach, 1980), and recent research has emphasized the importance of spatial attention in the nature and magnitude of masking effects (Enns & DiLollo, 2000, 1997); when the location of the target is not abundantly clear prior to stimulus presentation, masking effects are massively enhanced, even at long delays between the onset of the target and the subsequent mask. Thus, our participants' uniform failure to detect the attention cue is what many models of visual attention and masking would predict (e.g., Enns & DiLollo, 1997).

Attention capture

Eye movement analyses of all target trials revealed that across both participants and items, the location of the attentional cue had a reliable effect on looking patterns. In particular, participants correctly fixated the centrally located crosshair on a large percentage of trials (82.3%). Out of these trials, participants then went on to fixate the cued character first 72% of the time. One sample, two-tailed *t*-tests were performed on both participant and item mean proportions of first looks to the cued character (Table 1). These analyses showed that first looks to cued characters were reliably greater than would be expected by chance (0.72, 95% CI $= \pm 0.03$, where chance is less than 0.50, as Subjects will not always look first to one scene character or the other, and their first fixation may be to an unrelated region of the display).

Utterances

Perspective Predicate items. Fig. 6A presents the mean proportion of trials on which the Preferred Subject was used in Perspective verb items. (Confidence intervals and averages were computed using subject means.) As can be seen in the figure, more Preferred Subjects were uttered

Table 1
Statistical tests for Experiment 1: One-sample, two-tailed *t*-tests on mean proportion of trials that began with a look to the cued character as compared to chance (0.5)

	df_1	t_1	p_1	df_2	t_2	p_2
First look to cued character	35	42.64	<.01	23	19.42	<.01

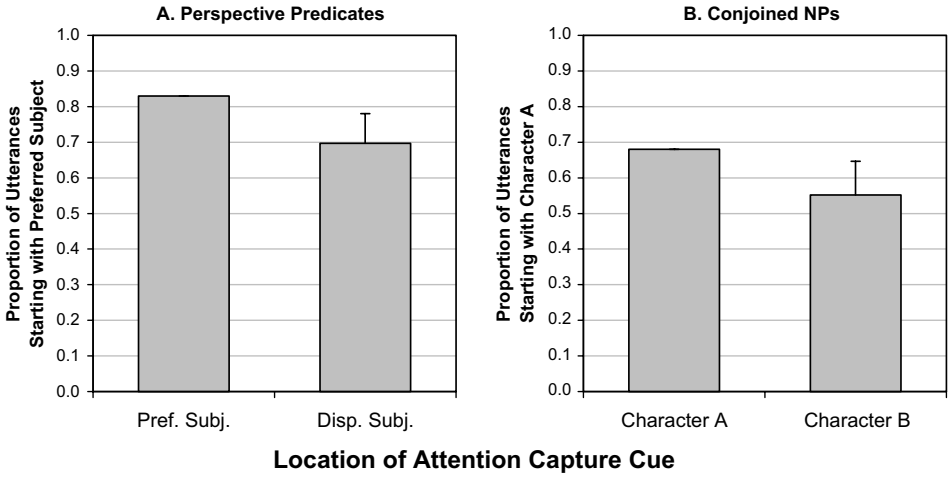

Fig. 6. Results of Experiment 1: effects of attention capture on word order for Perspective Predicates (A) and Conjoined NPs (B). Error bars indicate 95% CIs for the pairwise comparison and by convention are placed on the lower of the two data series.

when the Preferred Subject had been cued than when the Dispreferred Subject had been cued. Participant and item means were separately computed for the proportion of trials on which the Preferred Subject was used. The means were entered into separate Analyses of Variance (ANOVAs) having three factors (List (four lists), Attention Capture location (two positions), and the Left–Right orientation of the characters (two orientations)). Results of these analyses are shown in Table 2. A main effect of cue location was found, such that the mean proportion of trials on which participants began their utterance with the Preferred Subject was 0.83 when the attention-capture cue was in the Preferred Subject location, and was 0.70 when the attention-capture cue was in the Dispreferred Subject location (95% CI = ±0.08) (see Fig. 6A). No effect of left–right orientation was found on Perspective Predicate items: Scene characters on the left-hand side of the scene were no more likely to be the sentential Subject than those on the right. Additionally, no significant interactions were found between location of the attention cue and left–right orientation.[5]

Conjoined NP items. Fig. 6B presents the mean proportion of trials on which Character A was used in conjoined NP items as a function of whether Character A or Character B was cued. As can be seen in the figure, attentional cueing had a similar effect on word order for these items. Participant and item means were separately computed for the proportion of trials on which Character A was mentioned first. The means were entered into separate ANOVAs having three factors (List (four lists), Attention Capture location (two positions), and the Left–Right orientation of the characters (two orientations)). Results of these analyses can be seen in Table 3. A main effect of cue location was found, such that the mean proportion of trials on which participants began their utterance with Participant A was 0.68 when the attention-capture cue was in the Participant A location, and was 0.55 when the attention-capture cue was in the Participant B location (95% CI = ±0.09). Unlike Perspective Predicate items, a significant main effect of left–right orientation was also found for the Conjoined NP items, such that the mean proportion of trials on which participants began their utterance with Participant A was 0.74 when Participant A was on the left and was 0.48 when Participant B was on the left (95% CI = ±0.09). No significant interaction between the location of the attention cue and left–right orientation was found, however; despite the tendency to mention leftmost scene characters first, the attention cue had equivalent effects on first-mention regardless of left–right position within the scene.

Eye-contingent utterance analysis. One question that arises is whether the attention capture cue had an effect on word order choice independent of whether it was effective at capturing the first shift in visual gaze. It is possible for instance that the cue affected the accessibility of this character independent of initial eye movements (e.g. at some later stage of processing). We explored this possibility by informally comparing utter-

[5] Throughout this paper, whenever inferential statistics are reported over proportions, similar analyses were conducted using log-odds ratios or log transformations of probabilities. For instance, in this case, the ln((Preferred Subject Probability)/(Dispreferred Subject Probability)) was calculated and the corresponding inferential statistics were performed on the resulting participant and item means. Unless otherwise noted, significant effects using proportions were also significant using log transformations.

Table 2
Statistical tests for Experiment 1: ANOVAs of the proportions of utterances beginning with the Preferred Subject (for Perspective Pairs Predicates) or Participant A (for Conjoined NP Subjects), with Character Type (cued vs. uncued) and Left–Right orientation of characters (left vs. right) as factors

	df_1	F_1	p_1	df_2	F_2	p_2	min F	$df_{\min F}$	$p_{\min F}$
Perspective Predicates									
Cue location	1, 32	12.69	<.01	1, 8	12.88	<.01	6.392	1, 26	.02
Left–Right orientation	1, 32	2.74	>.05	1, 8	0.77	>.05	0.601	1, 13	.45
Interaction	1, 32	0.14	>.05	1, 8	0.003	>.05	0.003	1, 8	.96
Conjoined NPs									
Cue location	1, 32	16.52	<.01	1, 8	16.23	<.01	8.186	1, 25	<.01
Left–Right orientation	1, 32	30.12	<.01	1, 8	20.96	<.01	12.359	1, 21	<.01
Interaction	1, 32	0.64	>.05	1, 8	0.50	>.05	0.280	1, 22	.60

Table 3
Statistical tests for Experiment 2: One-sample, two-tailed *t*-tests on mean proportion of trials that began with a look to the cued character as compared to chance (0.5)

	df_1	t_1	p_1	df_2	t_2	p_2
Attention capture	35	49.26	<.01	35	55.46	<.01

ance choices for trials on which the attention capture cue effectively shifted gaze (72% of trials) vs. trials on which it did not (28% of trials). As Fig. 7 shows, trials on which the cue was effective at shifting gaze were also the trials that carried the effect of this cue on order of mention. ANOVAs were not possible on this smaller subset of trials (72% of the trials) due to missing data. Nevertheless, this pattern suggests that it was the initial capture of attention toward one character over the other (rather than some later process) that was influencing order of mention.

Discussion

The main result of this experiment is that the visually captured character was more likely to be mentioned first in both Perspective Predicate items and Conjoined NP items. Given that alternative orderings of Conjoined NPs (*the cat/dog and the dog/cat*) are not believed to reflect Figure–Ground assignment, we conclude that, at least for these items, the attention capture increased activation of the lemma that corresponded to the character being attended. Occam would encourage us to conclude that the same is true for the Perspective Predicate items, but it is entirely possible that in these cases attention capture is affecting both Figure–Ground status and lemma accessibility simultaneously. Indeed, as previously discussed, others have found that attention-capture influences Figure–Ground assignment of arbitrary black-and-white shapes (Vecera et al., 2004). We would not want to suggest that lemma accessibility is also at work

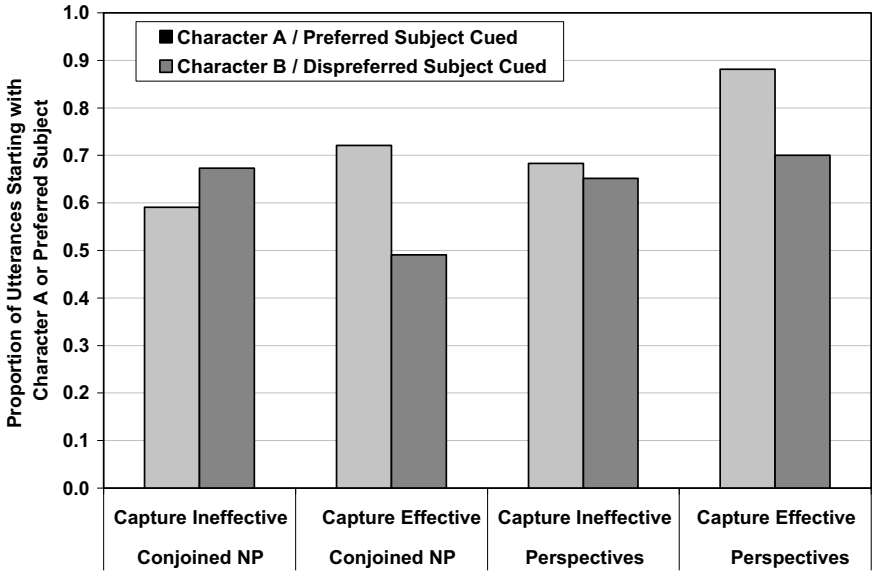

Fig. 7. Effects of attention capture word order choice in Experiment 1 as a function of whether the attention capture cue was Effective or Ineffective.

in the Vecera et al. study since the arbitrary shapes clearly had no names associated with them (and responses were collected using a non-verbal button-press).

Experiment 2

Experiment 2 utilized the same attention manipulation as Experiment 1, with three primary goals. First, we wanted to replicate the attention capture effects reported in Experiment 2. To this end, Conjoined NP items were again used, and these items were preceded by the same sudden-onset attention capture manipulation. Second, we wanted to explore the generality and robustness of this visual-attentive effect on word order. This was done by using the sudden-onset manipulation for images that elicit other linguistic forms that are flexible in terms of English word order: sentences containing Symmetrical verbs or in both Active and Passive structures.

Finally, we wanted to better understand how our findings relate to prior findings reported in Griffin and Bock (2000). Recall that Griffin and Bock (2000) found that initial self-generated shifts in attention did not predict word order choice in a small set of item types (one Perspective Predicate item type and two Active/Passive item types). If our attention capture manipulation does not influence word order choice in Active/Passive items, we would therefore conclude that Griffin and Bock (2000) failed to find effects of initial attention on word order because the majority of their items were of the Active/Passive sort.

However, we also decided to examine in a subset of our stimuli a situation much more like that of Griffin and Bock (2000). In particular, our Perspective Predicate items were used in the present experiment without any attention capture manipulation. They were simply preceded by a neutral fixation crosshair. Eye movements for these trials (i.e., self-generated attention shifts) will be analyzed to see whether they too predict word order or whether our attention effects on word order are limited to exogenous (attention capture) visual cues.

Methods

Participants

Thirty-six monolingual English-speaking students in an Introductory Psychology course at the University of Pennsylvania participated and received course credit in return.

Stimuli

The stimuli consisted of 88 images: 40 fillers and 48 critical items. The critical items were the 12 Perspective Predicate and 12 Conjoined NP stimuli from Experiment 1, and two new sets of stimuli: The first set consisted of 12 images designed to depict two animate entities engaging in a joint activity, such that they could be described using a Symmetrical Predicate (e.g., *hugging*, *kissing*, *arguing*). The other set consisted of 12 critical images depicting two animate entities, usually animals, engaged in a joint activity that could be described using either the Active or Passive form of a transitive verb (e.g., *kicking*, *scolding*, *splashing*).

The newly added critical items were again selected based on a prior norming study with 21 monolingual English speakers. This norming was designed to identify baseline rates of word order selection for these particular items and ensure that each item had the necessary flexibility (i.e. both word orders were produced at least once). Participants viewed a series of 64 pictures and were asked to describe the event that was taking place in the scene using a simple sentence. Of these 64 pictures, the 24 critical items depicted Symmetrical and Active/Passive scenes.

Baseline rates of first-mentioned scene characters for the Symmetrical items in this norming experiment showed the same pattern of results as Conjoined NP items in Experiment 1: No particular scene entity was preferred for first-mention, but left–right orientation did predict first mention. Thus, scene characters were again arbitrarily dubbed Character A and Character B for the sake of coding and data analysis, and they are referred to as such from this point onward. Baseline rates of first-mentioned scene characters for the Active/Passive items showed strong preferences to maintain the active structure, thus making the Agent of the action the Subject of the sentence and causing it to become the Figure in the scene's Figure–Ground relationship. For baseline rates of mentioning each scene character first, for each of these sentence types, see Appendix B.

Procedure and design

For Conjoined NP, Symmetrical, and Active/Passive items, the location of the attention-capture cue and the left-to-right orientation of the scene were counter-balanced across four stimulus lists. Perspective Predicate items were only counter-balanced for left-to-right orientation; no attention-capture cue was used for these items, rather a brief full-screen flash appeared prior to image onset exactly as in filler trials. Manipulations were within-participants, with each participant assigned randomly to one of the four lists. Otherwise, the experimental set-up and procedures were identical to those in Experiment 1.

Coding and analyses

Criteria for coding and analyses of the Perspective Predicate and Conjoined NP items were identical to the criteria used in Experiment 1, resulting in the exclusion of just over 9% of Conjoined NP trials (41 trials) and just over 18% of Perspective Predicate trials (78 trials). We describe below the criteria for the new stimulus types:

Symmetricals. Transcriptions for each Symmetrical trial were analyzed for the order of NPs within the utterance. Trials containing disfluencies or false starts were accepted unless these altered word order, as were changes in verb (e.g., *introduces himself to* rather than *shake hands with*). Utterances not containing both NPs were excluded. A handful of trials were discarded due to experimenter or participant error (e.g. audio recordings were not intact, or the participant mistakenly button-pressed and skipped a trial). Just under 16% of Symmetrical verb trials (69 trials) were excluded from further analyses for one of the above reasons.

Active/Passives. These were analyzed for Subject choice. Trials containing disfluencies were accepted unless the change altered word order, as were uses of various verbs (e.g. *throw out* instead of *kick* for the scene in 3a). If a participant produced two different utterances in response to some stimulus, we coded only the first of these. All responses not containing two NPs were also excluded. Lastly, a handful of trials were discarded due to experimenter or participant error. Just under 14% of Active/Passive trials (60 trials) were removed for one of these reasons.

Results and discussion

Post-experiment questionnaire

The post-experiment questionnaire used in Experiment 2 was identical to that used in Experiment 1. No participant reported being aware of the attention-capture cue, and all were quite surprised to discover that attention had been manipulated.

Attention capture

As expected, the attention capture technique was quite effective at attracting early eye movements. A comparison of proportion of looks to the cued versus uncued scene character during the first three seconds after scene onset can be seen in Fig. 8. The manipulation was effective as shown by the fact that an early and significant divergence of looks was found. To test the effectiveness of the attention capture mechanism, one sample *t*-tests like those reported for Experiment 1 were performed. The results of these analyses (see Table 3) demonstrated that across both participants and items, the location of the attentional cue had a reliable effect on both looking patterns: Participants were significantly more likely to direct their gaze to the cued scene character first than would be expected by chance (mean proportion of trials on which participants fixated the cued character first = 0.80, 95% CI = ±0.03).

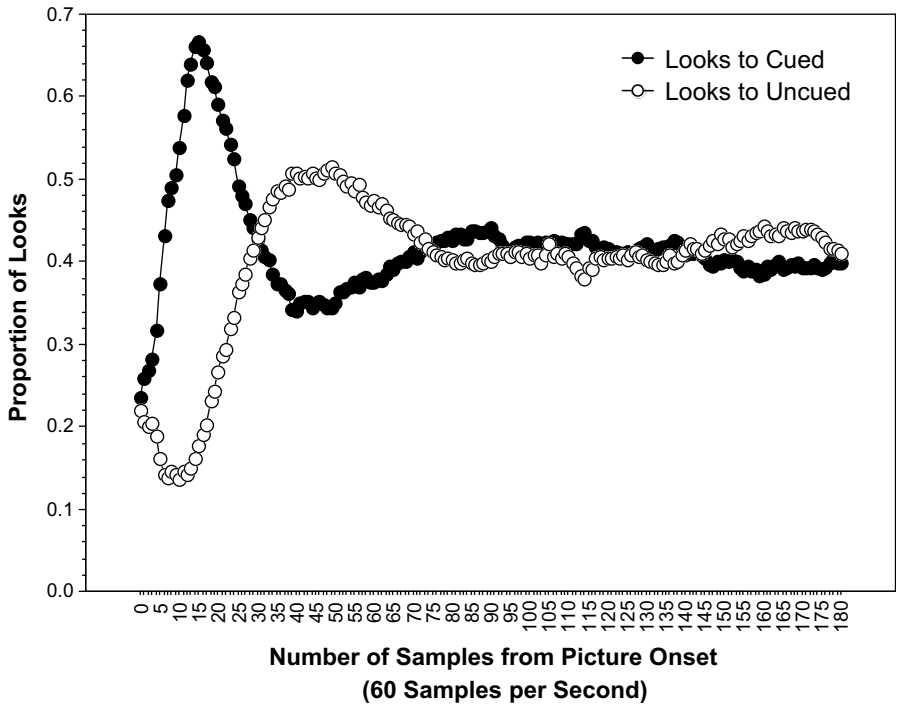

Fig. 8. Effects of attention capture on eye movement patterns from Experiment 2: changes in viewing across successive 17 ms intervals from picture onset through three seconds thereafter, collapsed across all Stimulus Types that utilized attention capture priming (CNP, SP, and A/P items).

Utterances

Conjoined NPs. Fig. 9A plots the proportion of trials for which Character A was mentioned first within the Conjoined NP stimuli across all four conditions. Participant and item means were separately computed for the proportion of trials on which Character A was mentioned first. The means were entered into separate ANOVAs having three factors (List (four lists), Attention Capture location (two positions), and the Left–Right orientation of the characters (two orientations)). Results of these analyses, replicating findings from Experiment 1, can be seen in Table 4. A main effect of cue location was found, such that the mean proportion of trials on which participants began their utterance with Participant A was 0.68 when the attention-capture cue was located where Participant A would appear, and was 0.53 when the attention-capture cue was located where Participant B would appear (95% CI = ±0.09). Left–right orientation was again significant for the Conjoined NP items, such that mean proportion of trials on which participants began their utterance with Participant A was 0.73 when Participant A was on the left, and was 0.49 when Participant B was on the left (95% CI = ±0.11). Again, no significant interactions between left–right orientation and location of the attention cue were observed.

Symmetricals. Fig. 9B plots the proportion of trials for which Character A was mentioned first for the Symmetrical Predicates stimuli across all four conditions. Participant and item means were separately computed for the proportion of trials on which Character A was mentioned first. The means were entered into separate ANOVAs having three factors (List (four lists), Attention Capture location (two positions), and the Left–Right orientation of the characters (two orientations)). Results of these analyses can be seen in Table 4. A main effect of cue location was found, such that mean proportion of trials on which participants began their utterance with Participant A was 0.61 when the attention-capture cue was located where

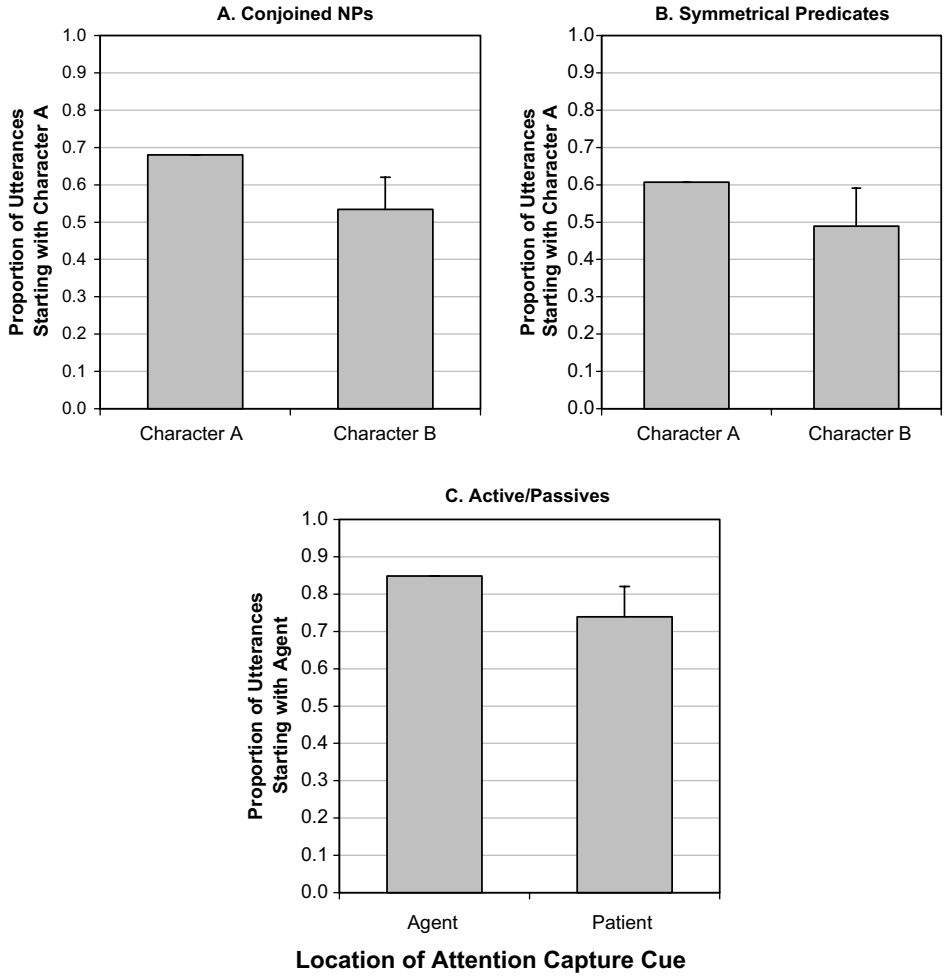

Fig. 9. Results of attention capture on word order choice in Experiment 2. Error bars indicate 95% CIs for the pairwise comparison and by convention are placed on the lower of the two data series.

Table 4
Statistics Tests for Experiment 2: ANOVAs on the proportions of utterances beginning with Participant A (for Conjoined NP Subjects and Symmetrical Predicates) or with the agent (for Active/Passive Predicates)

	df_1	F_1	p_1	df_2	F_2	p_2	min F	$df_{\min F}$	$p_{\min F}$
Conjoined NPs									
Cue location	1, 32	13.63	<.01	1, 8	6.04	.04*	4.185	1, 16	.06
Left–Right orientation	1, 32	18.41	<.01	1, 8	28.13	<.01	11.127	1, 32	<.01
Interaction	1, 32	0.10	>.05	1, 8	0.75	>.05	0.088	1, 38	.77
Symmetrical Predicates									
Cue location	1, 32	5.00	.03	1, 8	3.00	.10*	1.875	1, 19	.19
Left–Right orientation	1, 32	4.03	.05	1, 8	11.52	<.01	2.985	1, 39	.09
Interaction	1, 32	0.41	>.05	1, 8	0.92	>.05	0.284	1, 37	.60
Cue location within only SVO constructions	1, 14	8.17	.01	1, 9	0.31	>.05	0.299	1, 10	.60
Cue location within only CNP constructions	1, 22	19.27	<.01	1, 11	7.34	.02	5.315	1, 20	.03
Active/Passives									
Cue location	1, 32	4.75	.04*	1, 8	9.51	.02	3.167	1, 36	.08
Left–Right orientation	1, 32	1.06	>.05	1, 8	5.09	.08	0.877	1, 40	.35
Interaction	1, 32	0.04	>.05	1, 8	0.08	>.05	0.026	1, 36	.87

Factors were Character Type (cued vs. uncued) and Left–Right orientation of characters (left vs. right).
* Indicates the effect was only marginally significant when analysis was done on log-odds ratio.

Participant A would appear and was 0.49 when the attention-capture cue was located where Participant B would appear (95% CI = ±0.10). Left–right orientation was marginally significant for the Symmetrical Predicate items, such that the mean proportion of trials on which participants began their utterance with Participant A was 0.61 when Participant A was on the left and was 0.47 when Participant B was on the left (95% CI = ±0.09). Again, no significant interactions between left–right orientation and location of the attention cue were observed.

It is important to consider the Symmetrical Predicate findings in more detail, however. This is because two different sentence alternations are available for these items: (1) SVO framing structures (*A policeman is shaking hands with a construction worker*); or (2) unframed conjoined noun phrase structures (*A policeman and a construction worker are shaking hands*) that are identical on the surface to ordinary conjoined structures.[6] Because each of these forms was produced at least sometimes, we divided the responses into those two structures. We then considered only those participants who used each construction frequently enough to get a good estimate of the effect of attention capture on word-order choice. For example, many participants used the SVO structure only once or twice, making estimates for these individuals too sparse (consider the extreme: 1 SVO utterance for a participant, so the effectiveness of the cue is a 1 or a 0 for that participant). We drew the line at five instances (out of the 12 Symmetrical Predicates items) of a given structure; that is, if a participant used a structure fewer than five times, this participant was not included in the analysis of attention capture effects on this particular structure. The results of this ANOVA can also be seen in Table 4.

Eighteen participants used an SVO construction five or more times in the Symmetrical Predicates. For these 18 participants, when Character A was cued, it was then mentioned first (i.e. was in Subject position) 0.75 of the time, as opposed to only 0.44 of the time when Character B was cued (95% CI = ±0.11). Even with just these 18 individuals included in the analysis, this was significantly different from chance performance.[7] A similar analysis was done on the 26 participants who used Conjoined NPs five or more times, and again when Character A was cued, it was then mentioned first more frequently (0.63 of the time) than when Character B was cued (046 of the time) (95% CI = ±0.10). This too was significantly different from chance. Thus, regardless

[6] Recall from our initial description of these sentential types that in English ordinary and symmetrical coordination share the same structural format. Intransitive symmetrical predications can surface using a NP conjoined with *and*, e.g., *The horse and the rabbit meet* but so can nonsymmetricals, e.g., *The horse and the rabbit eat*. It is the inference structure of such sentence types that differs systematically. For the former type, a reciprocal relation is implied (i.e., that they met each other) but for the latter it is not (they both eat but are not implied to have eaten each other). So for the present analysis we are examining symmetrical predications that, on the surface, look like ordinary predications that happen to have a conjoined nominal grammatical Subject. In other languages, often the two types are differentiated morphologically, e.g., *Le cheval et le lapin se rencontrent* versus *Le cheval et le lapin mangent.*

[7] This effect was not significant by items, because of large differences in rates of using the SVO structure from one item to the next. For example, 30 participants used an SVO structure when describing the *hug* item, while only 4 participants used an SVO structure when describing the *marry* item.

of the structure, attention capture exerted effects on word order in the Symmetrical items.

Active/Passives. Fig. 9C plots the proportion of trials for which the Agent (i.e., the Preferred Subject) was mentioned first for the Active/Passive stimuli across all four conditions. Participant and item means were separately computed for the proportion of trials on which the Preferred Subject was used in Subject position. The means were entered into separate ANOVAs having three factors (List (four lists), Attention Capture location (two positions), and the Left–Right orientation of the characters (two orientations)). Results of these analyses can be seen in Table 4.

A small but reliable main effect of cue location was again found, such that the mean proportion of trials on which participants began their utterance with the Agent was 0.85 when the attention-capture cue was located where the Agent would appear and was 0.74 when the attention-capture cue was located where the Patient would appear (95% CI = ±0.08). Left–right orientation was not significant for the Active/Passive items (although marginal by items). And again, no significant interactions between left–right orientation and location of the attention cue were observed.

These findings again indicate a potent role for attention in linguistic choice and sentence production. The sheer force of this effect can be seen when one examines the latencies to begin an utterance within the Active/Passive items (see Fig. 10). Although no effects of first-mentioned character on utterance latency had been seen yet on other item types (regardless of attention capture or left–right orientation), we found that, within the Active/Passive-Pairs items, beginning an utterance with the Patient (i.e. Dispreferred Subject) led to somewhat delayed utterance onset (2324 ms) as compared with Preferred Subject utterances (2076 ms), regardless of which scene character had been cued (CI = ±241 ms).[8] It appears that activating and producing a passive structure (e.g. *The boy is kicked by the man*) requires significant cognitive time and effort, yet attention capture causes speakers to sacrifice this time and effort for the sake of satisfying effects of attention.

Eye movements

Eye-voice span. Before discussing effects of attention capture, we made an initial assessment of eye movements relative to utterance onset. If our participants are behaving like those in other eye-tracking production experiments, we would expect a tight coupling between eye position and referent mention (eye-voice span, Griffin & Bock, 2000). To examine this question, we re-plotted the eyegaze data based on the referent choice that speakers made for each trial. In particular, following Griffin and Bock (2000), Fig. 11 plots (over time) the proportion of trials for which participants were looking at the character that they ultimately chose to be the first mentioned character for that utterance (N1) vs. the second mentioned character (N2). For example, on a trial for which the participant uttered *A cat and a dog are growling*, looks to the cat were classified as N1 looks whereas looks to the dog were classified as N2 looks. If the participant had instead said *A dog and a cat are growling*, looks to the dog would be N1 looks and the cat N2 looks. Fig. 11A plots N1 and N2 looks relative to utterance onset across all stimulus types and separately for each construction type (Conjoined NP, Symmetrical, Active/Passive, and Perspective verbs; see Figs. 11B–E). Indeed, as reported by Griffin and Bock (among others), this analysis revealed strong effects of word order at utterance onset. That is, just prior to mentioning a character, participants fixate that character.

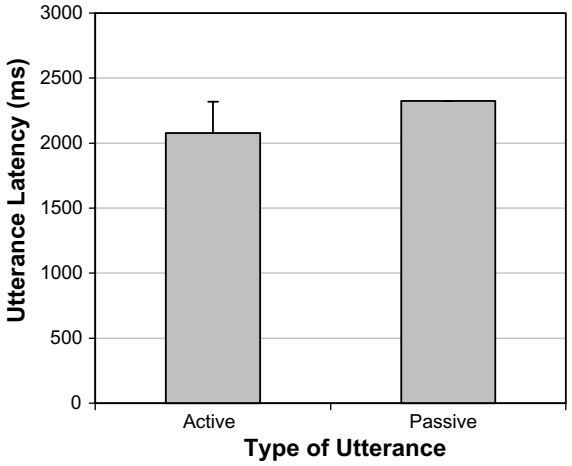

Fig. 10. Utterance latencies in ms for Active/Passive items in Experiment 2. Error bar indicates 95% CIs for the pairwise comparison and by convention is placed on the lower of the two data series.

This pattern holds for all stimulus types. Data of this sort are consistent with a range of existing theories of sentence production but do not address claims of an early nonlinguistic gist-extraction stage. To address these latter issues, we now turn to effects of attention capture on both eye position and utterance choice.

Effects of early fixation on word order. Our production data indicate that the attention capture manipulation

[8] Trials with RTs longer than 5000 ms (4 Trials) were dropped, as were trials that were uncodable or had experimental error (32 trials). Moreover, entire participants had to be dropped from the analysis because they had one or zero observations in Passive Condition (leaving just 21 Subjects total). As such, only a Subject ANOVA was performed, resulting in a main effect of structure choice on RT ($F(1,20) = 4.60$, $p = .04$).

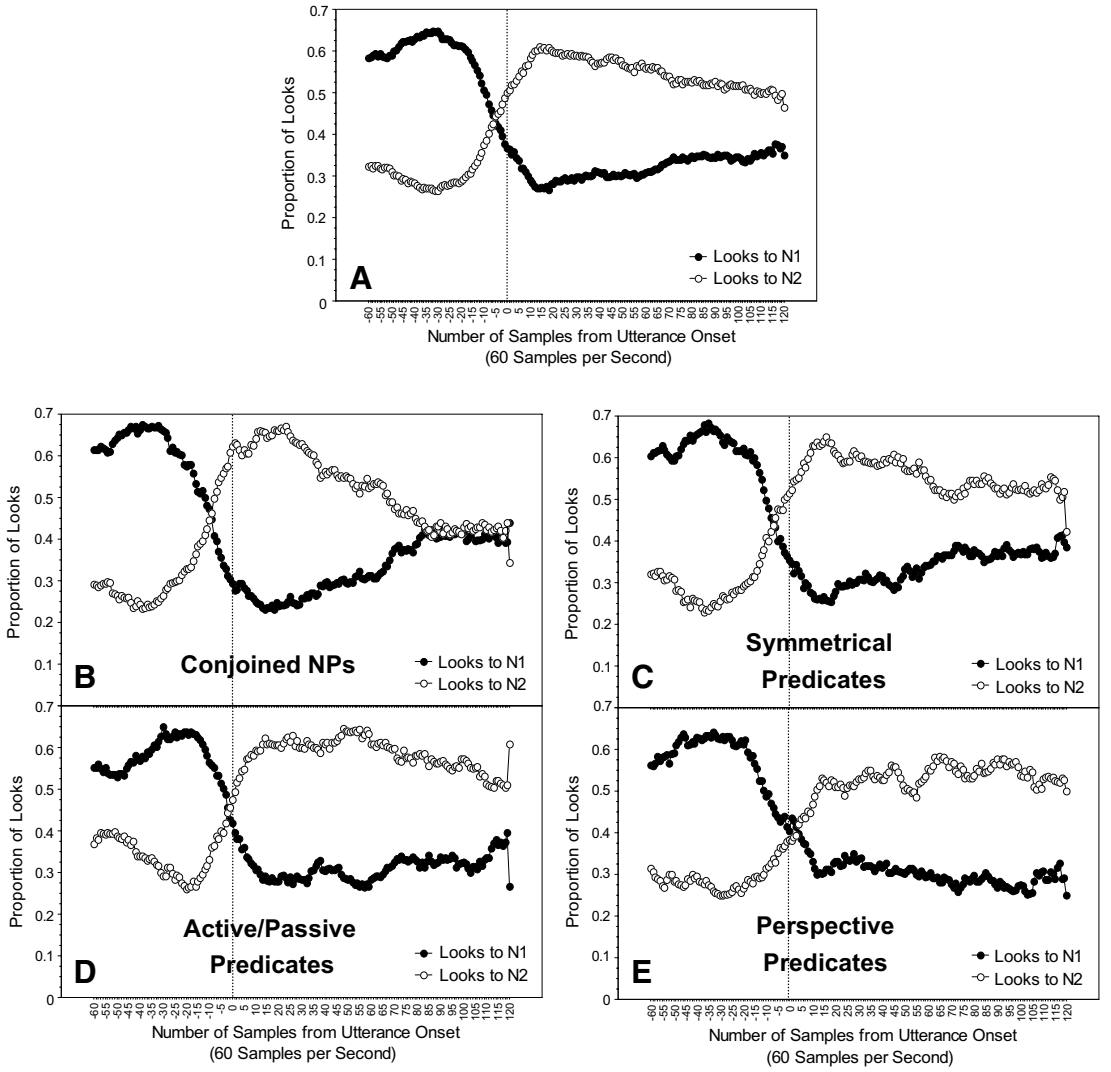

Fig. 11. Eye movement patterns relative to utterance onset from Experiment 2: changes in viewing across successive 17 ms intervals, beginning one second prior to utterance onset (indicated by the dashed vertical line), and continuing for two seconds thereafter, collapsed across all Stimulus Types (A), and plotted separately for Conjoined NP items (B), Symmetrical Predicate items (C), Active/Passive items (D)—all of which utilized attention capture priming—and PSP items (E), which did not manipulate participants' attention.

affects *word order choice*: Cued characters are more likely to be promoted to early positions in participants' utterances describing such scenes. The eye movement results indicate that attention capture also influences *early eye movements*: People are likely to look first at the cued character. An important question therefore becomes whether it is these early movements to characters or the attention capture of characters (or both) that influences word order choice.

To examine this question, we plotted our data relative to image onset (Fig. 12) rather than relative to utterance onset. Fig. 12A collapses across all of the stimulus types: Conjoined NP, Symmetrical, Active/Passive and Perspective verbs (which, recall, received no attention capture manipulation). Figs. 12B through E plot the data separately for each stimulus type.

Overall (Fig. 12A) there is a strong and early relationship between referent choice and eye movements: The character that is ultimately mentioned first by the speaker is also likely to be looked at early in the display of the image. This pattern is similar to that reported by Griffin and Bock (2000). However, our effect appears much earlier than Griffin and Bock's: During the first 200 ms of the display, looks to N1 and N2 diverge. This difference is unexpected according to the approach of Griffin and Bock (2000) and Bock et al. (2004), which posits an initial gist-extraction stage that is unrelated to linguistic planning.

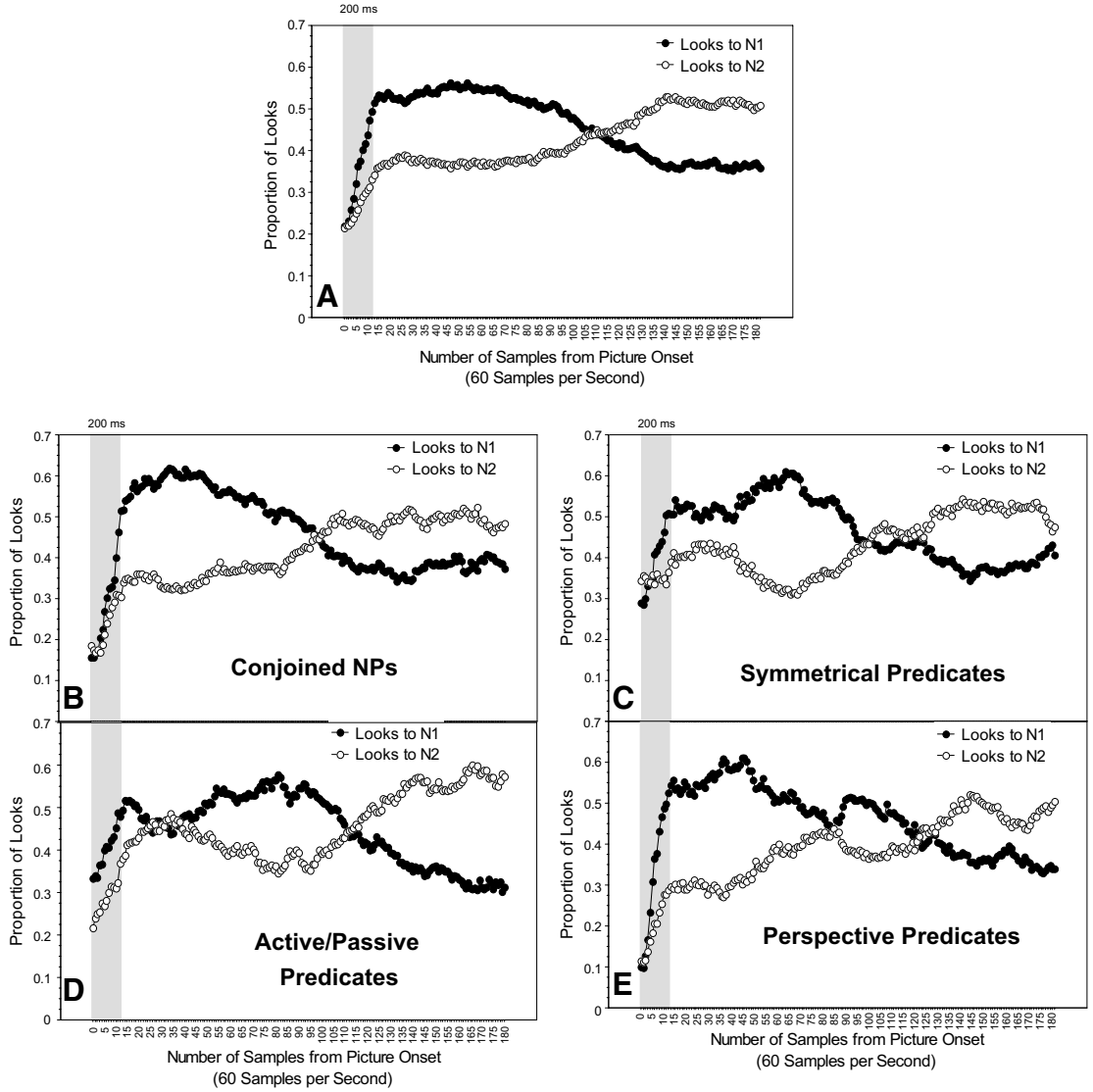

Fig. 12. Eye movement patterns relative to picture onset from Experiment 2: changes in viewing across successive 17 ms intervals from picture onset through three seconds thereafter, collapsed across all Stimulus Types (A), and plotted separately for Conjoined NP items (B), Symmetrical Predicate items (C), Active/Passive items (D)—all of which utilized attention capture priming—and PSP items (E), which did not manipulate participants' attention.

Statistical analyses of these data support the hypothesis that there are linguistic planning influences during the initial 200 ms of scene apprehension. In particular, we calculated the proportion of time spent looking at the N1 and N2 characters during the first 200 ms of image display. Participant and Item means of this measure were entered into separate ANOVAs, which included List (4 lists), Character Type (N1 vs. N2) and Stimulus Type (Conjoined NP, Symmetrical, Active/Passive and Perspectives) as factors. These data are presented in Table 5. Consistent with Fig. 12A, we found a reliable effect of Character type during this first 200 ms time window such that N1 looks were greater than N2 looks (mean proportion of time during this early time window that participants spent viewing the first-mentioned scene character was 0.33, and the mean proportion of time spent viewing the second-mentioned scene character was 0.26, 95% CI = ±0.04). A main effect of Stimulus Type was also found (mean proportion of time during this early time window that participants spent viewing either scene character was 0.24, 0.36, 0.34 and 0.25 for Conjoined NP, Symmetrical, Active/Passive, and Perspective items, respectively), but interestingly, this early effect did not interact with Stimulus Type.

Table 5
Statistical tests for Experiment 2: ANOVAs on the proportion of time spent looking at a character within the first 200 ms of scene viewing, with Character Type (First-Mentioned vs. Second-Mentioned) and Stimulus Type (Conjoined NP, Symmetrical, Active/Passive, and Perspective Predicates) as factors

	df_1	F_1	P_1	df_2	F_2	p_2	$\min F$	$df_{\min F}$	$p_{\min F}$
All Stimulus Types									
Character Type	1, 32	19.97	<.01	1, 43	6.36	.02	4.977	1, 66	.03
Stimulus Type	3, 96	52.10	<.01	3, 43	3.17	.03*	2.988	1, 48	<.01
Interaction	3, 96	2.11	.10	3, 43	0.73	>.05	0.542	3, 74	.35
Perspective Predicates									
Character Type	1, 32	19.24	<.01	1, 10	1.14	>.05	1.076	1, 11	.32

* Indicates the effect was only marginally significant when analysis was done on log-odds ratio.

Table 6
Statistics for effects of first looks on word order in Experiment 2 for Perspective Predicates: ANOVAs on the mean proportion of utterances beginning with the Preferred Subject, with first look (to the Preferred Subject vs. not to the Preferred Subject) as a factor

	df_1	F_1	p_1	df_2	F_2	p_2	$\min F$	$df_{\min F}$	$p_{\min F}$
First Look	1, 32	25.03	<.01	1, 11	3.54	.09	3.101	1, 14	.10

Indeed, as can be seen in the separate plots for each Stimulus Type (Figs. 12B–E), there is an N1 preference regardless of type of stimulus.

Uncued Perspective Predicates. Particularly striking is that we even see this N1 preference in the uncued Perspective Predicates (*chase/flee*), which did not involve attention capture. This condition is most like the one used in Griffin and Bock (2000), who did not use an attention capture method. Unlike Griffin and Bock, however, we do observe an early effect of linguistic choice: N1 looks were greater than N2 looks even during the first 200 ms, though the effect was significant only in the participant analysis and not the item analysis (examining only the Perspective Predicate items, mean proportion of time during this early time window that participants spent viewing the first-mentioned scene character was 0.31, and the mean proportion of time spent viewing the second-mentioned scene character was 0.22, 95% CI = ±0.02).

We also examined the effects of first fixation on first mention (N1 vs. N2), following the analyses reported in Griffin and Bock. These data are presented in Table 6. And again, we find that a character was more likely to be the Subject of a Perspective Predicate sentence when gaze was directed to that character first; the mean proportion of Perspective Predicate utterances beginning with the Preferred Subject was 0.80 when the Preferred Subject was first-fixated and was only 0.54 when the Preferred Subject was *not* first-fixated, 95% CI = ±0.10, with the effect being significant by participants and marginal by items. In both tests, we believe the reason for the weakness of the item analyses is the sparseness of condition data when divided based on participant behavior: Some items were strongly biased toward one alternative (again, see baseline rates in Appendix B).

Taken together, these eye-movement findings indicate that *where* people look first during scene description is related to *what* is mentioned first. We, by experimental artifice, can change the probabilities of where participants look first (via attention capture) and this correspondingly changes the probability of what the participants mentioned first. Moreover, even when procedurally we don't exert an external influence on looking preferences (as in the Perspective Predicate manipulations), the finding is that first looks are related to linguistic choice.

Summary

The results of Experiment 2 strongly support the claim that scene apprehension and linguistic planning temporally overlap from the onset of both processes. There does not appear to be a period of time in which eye movements are dissociable from linguistic factors during scene description tasks.

General discussion

Across two experiments, we found that manipulation of a speaker's initial visual attention toward a character in a scene has a reliable effect on word order choice, regardless of the type of linguistic alternation tested: Depicted characters that are looked at first are more likely to be mentioned first. This effect was found not only for utterances whose word-order alternatives contrast Figure vs. Ground interpretation of depicted characters (Active/Passive, Perspective, and Symmetricals) but also for utterance alternatives that do not make this contrast (Conjoined NP Pairs). Inter-

estingly, the magnitude of these visual-attentive effects was not noticeably different across types, despite the presence in some cases of strong conceptual and linguistic preferences for particular word orders. Participants were sufficiently influenced by visual-attentive factors to choose ordinarily disfavored conceptualizations of the scene or ordinarily disfavored linguistic constructions, or sometimes both. For instance, these effects held even when this Subject assignment required the use of a different and less favored verb (*flee* rather than *chase*) or a different and less favored sentence structure (Passive rather than Active). These results are very much in line with prior results of Tomlin (1997), although here we show the surprising generality of this visual-attentive effect across a broad range of event and construction types. This effect holds despite the fact that participants in our experiments were unaware that their direction of gaze was being manipulated whereas Tomlin's participants were implicitly instructed as to which character the images were "about" by being told which one to fixate on and were cued as to this character throughout the trial by a pointing arrow.

Detailed analyses of the time course of eye position (Experiment 2) showed that looking to a character during the first 200 ms after image onset reliably predicted the speaker's tendency to mention this character first. This pattern was observed even for uncued conditions (using Perspective Predicate items); that is, endogenous shifts in attention at image onset partially predicted speakers' word order variation, contrary to earlier results of Griffin and Bock (2000) who used a mixture of Perspective verbs (one of the three item types that they coded as "Active/Passive") and Active/Passives (the other two of their three Active/Passive types).

Implications

It is difficult to imagine an account of these new results that preserves the Bock and colleagues' (Bock et al., 2004, 2003; Griffin & Bock, 2000) notion of an initial gist-extraction stage of processing in which linguistic processes are not at work as well. The findings support language production theories that allow for considerable temporal overlap of the processes related to scene *apprehension* and linguistic *formulation*. We find no evidence for an initial visual apprehension stage during which the language processing system is disengaged or turned off. Participants knew that they were going to describe simple pictures, and apparently both their visual-perceptual and linguistic systems were prepared for this task from the outset, working out "what to see" and "what to say" about the world in an integrated interactive fashion. This doesn't mean that computations in the brain are somehow mysteriously instantaneous. On the contrary, precisely because computations unfold over time (both visual-apprehension computations and linguistic-formulation computations), efficiency is gained by letting these computational systems work in parallel and in concert on temporally accumulating information.

Indeed, the simplest account assumes that initial shifts in attention exert an immediate influence on lemma accessibility precisely when apprehension processes are still unfolding. For example, if attention is drawn to a dog in a scene where a dog is chasing a man, increased activation of the lemma associated with this entity (the dog) could lead to increased availability of the corresponding lexeme (the phonological form, /dog/). More available (accessible) lexical entries will tend to be mentioned first, so this increased availability could lead to an increased tendency to place *the dog* at the beginning of the participant's description (in this case, in the sentential Subject role).

This lemma-activation account, which we suspect is the correct one, is compatible with most current theories of language production and is reminiscent of earlier work of Bock (1986), who found that the semantic priming of a to-be-uttered constituent promoted it to an earlier position in the speaker's utterance. It was found, for example, that the use of a lexical prime (*thunder* vs. *worship*) influenced whether speakers went on to describe a picture as *Lightning is striking a church* or *A church is being struck by lightning.* Note, however, that the present 'priming' effects were instigated by a non-linguistic cue (an attention-capturing flash). This suggests that when people are engaged in a task of describing visual input, linguistic representations are immediately triggered from perceptual input regardless of how far along into the apprehension process the perceiver has gotten—regardless, that is, of whether he/she has apprehended *chasing or* any other relational property of the world.

Despite the apparently straightforward account that we just offered, it is not as if there are no alternatives that remain at least partly viable. Imagine for example an initial nonlinguistic gist-extraction stage of processing that primes a particular interpretation of the scene as a whole, rather than just priming individual scene elements and their corresponding lexical entries. On this interpretation, attentionally focusing the dog in a scene in which a dog is chasing a man would encourage the speaker to view this scene as one in which chasing rather than fleeing is happening. This interpretation of our findings comports well with the view of Bock (1995), where an early gist-extraction stage precedes linguistic planning, during which the relationship between entities involved in an event is assessed. This account is also more akin to the explanation given for attentional cueing effects when viewing ambiguous figures (i.e., Figure/Ground, wife/mother-in-law images)—in these cases there are no lemmas to account for priming effects more elementally. The only explanation for these effects is one in which certain features lend themselves more to one interpretation (e.g. mother-in-law) than the other (e.g. wife), and focusing these features promotes the corresponding interpretation. If our attention capture effects are

exerting their influences similarly, one can think of these stimuli as ambiguous scenes, which can be viewed in one of two Figure–Ground ways. For example, in our dog-chasing-man scene, either the dog or the man can be viewed as the Figure, and the surrounding information will thus serve as the background. Attentional focus on one of these scene characters will encourage one viewing of the scene or the other, and this will have ensuing effects on the way the scene is described.

Note however that this explanation has difficulty explaining the full data pattern reported here. Effects of attention on word order were observed for Conjoined NP items, in which the ordering of NPs does not reflect the extraction or communication of Figure–Ground information. Both entities are in Subject position and therefore have the status of Figure in the message. It seems likely then that lemma activation is at least at play in producing the Conjoined NP word orderings. Indeed, Conjoined NP items were selected for our studies precisely because they exhibit no alterations in Figure–Ground assignment. The finding that cued elements tend to occur first within Conjoined NPs implicates the simpler account given above in which visual apprehension and linguistic formulation are evolving together during the planning and execution of a speech event.

Our interpretation of the findings does not preclude effects of Figure–Ground assignment. We simply see this as one of several factors contributing to word order choices. Indeed, a closer examination of the Symmetrical items provides suggestive evidence for this view, namely that lemma accessibility *and* Figure–Ground assignment provide independent but additive effects on the probability that an entity will be mentioned first in an utterance. For symmetrical predicates, the ordering of NP elements when an SVO structure is produced (as in, *The man/woman is shaking hands with the woman/man*) conveys solely the Figure–Ground relationship a speaker has selected to communicate. However, symmetrical predicates were also sometimes produced using Conjoined NP constructions, in which NP ordering does not affect Figure–Ground assignment (as in, *The man and the woman are shaking hands*). Thus SVO symmetrical items are particularly useful for determining the significance of any role for holistic aspects of scene apprehension in word order choices whereas Conjoined NP symmetrical predicates are useful for assessing effects of non-holistic (lemma accessibility) effects on ordering. A finding of comparable levels of priming effects within each of these structures would lend itself to a simple lemma activation account of initial attention, since lemma activation should be contributing comparably in each of these cases. Instead, we found higher proportions of cued elements appearing first in SVO structures (cued elements appeared as the sentential Subject for these trials 75% of the time) than in Conjoined NP structures (cued elements appeared as the first-mentioned noun in the Conjoined NP Subject 63% of the time). This finding suggests a deeper involvement of priming in visual-apprehension processes. That is, both Figure–Ground assignment and lemma accessibility are at work to influence word order choices.

Comparison to previous findings

In Experiment 2, visual-attentive effects on word order were observed even when the attention capture cue was not administered (i.e., the uncued condition using Perspective Predicates): Initial self-generated shifts in attention predicted speakers' word order variation, contrary to earlier results of Griffin and Bock (2000). We suspect that this difference stems from the limited scope of the Griffin and Bock (2000) materials. In particular, their observations were drawn from only three stimulus types: One Perspective Predicate item and two Active/Passive items, which were the only items that showed some word order flexibility in speaker's productions. Griffin and Bock's (2000) failure to find early visual-attention effects on word order choice may be due to the lack of statistical power offered from this small sample, or perhaps idiosyncratic properties of these three experimental items. It is also the case however that our own experimental exploration of uncued attentional effects was limited; we only examined Perspective Predicates and not the full range of other stimulus types discussed here. Word order changes for Perspective Predicates do not require the use of the disfavored passive structure (which was required for two of the three items used in Griffin & Bock, 2000). Thus it is possible that the endogenous attention effects that we observed here will only hold up for variations that are not dispreferred structurally. This explanation seems unlikely, however, given that our exogenous manipulations of attention were found to influence word order in Active/Passive pairs.

Differences also exist in the visual stimuli used across these two studies. Notably, Griffin and Bock (2000) used black-and-white line drawings whereas the present study used full-color clip-art renderings of actions. It is possible that these more salient stimuli allowed participants in the present study to apprehend the event more rapidly as compared to participants in the Griffin and Bock (2000) study. This would essentially preserve the Griffin and Bock (2000) account by assuming that event apprehension occurs during an even shorter time window in the present studies (e.g., the first 100 ms rather than the first 300 ms). Such an account however becomes essentially indistinguishable from our own in terms of experimental predictions and assumes the tight temporal relationship between visual apprehension and linguistic formulation advocated here.

Indeed, there is mounting evidence that linguistic elements (names for objects, events, etc.) are triggered at astonishing speeds during visual perception (specifically, and perhaps crucially, during description tasks). For

instance, Morgan and Meyer (2005) asked participants to describe pairs of visually presented objects. After fixation of the first object, but prior to fixating the second object, this second object was sometimes replaced by a different object. As expected, gaze durations and naming latencies of this second object were speeded when the initial version of this object had been the same as the final version (identity priming) as compared to previewing a different unrelated object. However, facilitation of gaze durations and naming latencies were also observed as compared to an unrelated object when the preview version was an object whose name was a homonym of the target object (e.g., a baseball bat changing to a flying-mammal-bat). This study (see also Pollatsek, Rayner, & Collins, 1984, for similar findings) suggests that parafoveally viewed objects that are about to be fixated have already triggered not only conceptual representations, but also lexical-semantic and lexical-form representations.

Rapid gist extraction

The Morgan and Meyer (2005) findings raise an important question regarding exactly how much parafoveal visual information is processed during a given fixation within a scene and how deeply and abstractly this information is processed so as to generate task relevant information on the fly. Vision researchers are only now beginning to explore this question for 'complex' images like the ones used in the present studies (and more complex but richer photographic images, Brockmole & Henderson, 2006; Brockmole, Castelhano, & Henderson, 2006; Henderson, 2003).

The burgeoning literature on rapid gist extraction suggests that some surprisingly abstract information is rapidly extracted from both foveal and extra-foveal regions within the first 100 ms of image onset (e.g., Oliva, 2005; Potter, 1976; Potter, Staub, & O'Connor, 2004). For instance, this work shows that participants can detect basic level scene categories (e.g., find a mountain scene) from RSVP of pictures, each presented at very fast rates (100–250 ms). This work however, has not yet examined systematically the content of these representations (e.g., Underwood & Green, 2003). For example it is not known the extent to which detailed relational representations (e.g., a man is chasing a dog) contribute to scene gist. Moreover, the work to date suggests that rapid scene categorization may reflect computations over low-frequency texture gradient information (Oliva & Torralba, 2006; Renninger & Malik, 2004). Such texture information may not be sufficient to categorize events of the sort used here. Nevertheless, we see this as an open issue that requires further experimental research. If such computations are possible in such brief periods of time, the finding would only serve to highlight in the current research that visual apprehension and linguistic formulation in scene description are tightly coupled, rapid and not dissociable in a stage-like fashion at any time scales thus far explored behaviorally.

Eye gaze, joint attention, and the child's discovery of the referents of words

Although acquisitional issues are not the central topic of this investigation, our findings do have potential implications for the way adults interpret language and the way children learn it. Both adults and children use eye-gaze cues to establish communicative alignment; adults select intended referents with increased alacrity when gaze cues are available to guide referential processing (Hanna & Brennan, 2004) and can infer the meaning of a nonce verb when gaze cues to its intended sentential Subject are provided (Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Trueswell, Nappa, Wessel, Gleitman, & March, 2007.). Adults with access to the gaze direction of their communicative partner (once the referential domain has been established) are likely to omit otherwise necessary disambiguating information, for example simply saying, "the square," in the presence of many squares (Brown-Schmidt, 2006). Moreover, young language learners are quite adept at taking visual perspective in object labeling tasks; by the time they're 18 months old, young children will inspect a speaker's attentional state upon hearing a novel label, even when an obvious candidate object (a novel toy, which they've never seen before) is present and salient (Baldwin, 1993). Since our findings demonstrate the intensely predictive nature of gaze direction on utterance formation (initial fixations affect linguistic choice, and there is a tight time link between gaze to mentioned objects and their mention), this provides the potential to explore how gaze cues might be used in much more complex and transient environments than referential resolution and object-labeling tasks. Adults may be using this information rapidly and expediently to arrive at increased communicative alignment, and children may be able to utilize the caretaker's gaze direction patterns in complex language-learning tasks such as verb learning and syntactic interpretation.

Closing remarks

At their most general, the studies presented here tried to contribute to understanding of how the processes of conceptualizing the world and linguistically describing it exert mutual and often simultaneous influences. If we are right, the unconscious, rapid, and incremental speech machinery is not wholly or even predominantly conception first and speech only thereafter and in consequence; rather, the representations constructed by the visual-attentive and linguistic-conceptual systems may be integrated all along the line.

Appendix A

Baseline rates of using the Preferred vs. the Dispreferred Verb in the norming of the Perspective Verb Pair (PSP) items and of starting an utterance with Character A vs. Character B in the Conjoined Noun Phrase (CNP) items

Perspective Verb Pair (PSP) item	Preferred Verb	Dispreferred Verb	Conjoined Noun Phrase (CNP) item	Character A first-mentioned	Character B first-mentioned
Buy/sell	Sell (13/21)	Buy (8/21)	Biking	Turtle (13/21)	Dog (8/21)
Chase/flee (dog/man)	Flee (12/21)	Chase (9/21)	Dancing	Fish (12/21)	Bear (9/21)
Chase/flee (rabbit/elephant)	Chase (15/21)	Flee (6/21)	Eating	Panda (10/21)	Koala (11/21)
Eat/feed (puppies/dog)	Feed (16/21)	Eat (5/21)	Growling	Cat (11/21)	Dog (10/21)
Eat/feed (child/mother)	Feed (20/21)	Eat (1/21)	Juggling	Elephant (11/21)	Seal (10/21)
Give/receive	Give (15/21)	Receive (6/21)	Jumping	Cat (9/21)	Frog (12/21)
Listen/talk (office)	Talk (16/21)	Listen (5/21)	Playing cards	Pig (12/21)	Dog (9/21)
Listen/talk (phone)	Talk (4/21)	Listen (6/21)	Playing horns	Snail (10/21)	Rhino (11/21)
Perform/watch (singer)	Perform (14/21)	Watch (7/21)	Rowing	Bear (11/21)	Snowman (10/21)
Perform/watch (speaker)	Perform (18/21)	Watch (3/21)	Skating	Monkey (12/21)	Rabbit (9/21)
Win/lose (boxing match)	Win (20/21)	Lose (1/21)	Swinging	Elephant (13/21)	Monkey (8/21)
Win/lose (race)	Win (10/21)	Lose (5/21)	Waiting	Deer (10/21)	Penguin (11/21)

Appendix B

Baseline rates of starting an utterance with Character A vs. Character B in the norming of the Symmetrical Predicate (SP) items and of starting an utterance with the agent (Active Structure) vs. the Patient (Passive Structure) in the norming of the Active/Passive (A/P) items

Symmetrical Predicate (SP) Item	Character A first-mentioned	Character B first-mentioned	Active/Passive (A/P) Item	Agent first-mentioned (Active Sentence)	Patient first-mentioned (Passive Sentence)
Argue	Batter (13/21)	Umpire (8/21)	Electrocute	Woman (15/21)	Man (6/21)
Crash	Car (10/21)	Truck (11/21)	Videotape	Man (17/21)	Bear (1/21)
Dance	Woman (12/21)	Man (8/21)	Fire (from a job)	Boss (17/21)	Employee (4/21)
Fight (boxing)	Cat (19/21)	Mouse (1/21)	Hit	Boy (20/21)	Man (1/21)
Fight (tug-of-war)	Man (18/21)	Dog (3/21)	Kick (out the door)	Man (18/21)	Boy (3/21)
Hug	Mother (18/21)	Daughter (3/21)	Lick	Cat (20/21)	Dog (1/21)
Kiss	Kangaroo (12/21)	Cat (9/21)	Lift	Father (19/21)	Son (1/21)
Marry	Bear (10/21)	Cat (8/21)	Scold	Mother (18/21)	Child (2/21)
Shake Hands	Construction Worker (8/21)	Policeman (13/21)	Shoot (squirt gun)	Boy (16/21)	Girl (1/21)
Talk (face-to-face)	Nurse (11/21)	Doctor (10/21)	Splash	Boy (20/21)	Girl (1/21)
Talk (phone)	Woman (10/21)	Man (7/21)	Step On	Foot (18/21)	Bug (2/21)
Touch	Girl (14/21)	Alien (5/21)	Throw	Father (19/21)	Child (2/21)

References

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418.

Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics, 28*, 241–248.

Bock, J. K. (1986). Meaning, sound, and syntax: lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575–586.

Bock, J. K. (1987). Coordinating words and syntax in speech plans. In A. Ellis (Ed.). *Progress in the psychology of language* (Vol. 3). London: Erlbaum.

Bock, J. K. (1995). Sentence production: from mind to mouth (2nd ed.. In J. L. Miller & P. D. Eimas (Eds.). *Speech, language, and communication. Handbook of perception and cognition* (Vol. 11, pp. 181–216). San Diego, CA, US: Academic Press.

Bock, J. K., Irwin, D. E., & Davidson, D. J. J. (2004). Putting first things first. In F. Ferreira & M. Henderson (Eds.), *The integration of language, vision, and action: Eye movements and the visual world* (pp. 249–278). New York: Psychology Press.

Bock, J. K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language, 48*, 653–685.

Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*, 1–39.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in backward visual masking: a comparison, review, and update. *Perception & Psychophysics, 62*, 1572–1595.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13*, 99–108.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: gobal and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 699–706.

Brown-Schmidt, S. (2006). Language processing in conversation. (Doctoral dissertation, University of Rochester, 2005). Dissertation Abstracts International: Section B: The Sciences and Engineering, 66 (9-B), 5079.

Cooper, W. E., & Ross, J. R. (1975). World order. In R. E. Grossman, L. J. San, & T. J. Vance (Eds.), *Papers from the parasession on functionalism* (pp. 63–111). Chicago: Chicago Linguistic Society.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*, 547–619.

Enns, J. T., & DiLollo, V. (2000). What's new in visual masking. *Trends in Cognitive Sciences, 4*, 345–352.

Enns, J. T., & DiLollo, V. (1997). Object substitution: a new form of masking in unattended visual locations. *Psychological Science, 8*, 135–139.

Ellis, S. R., & Stark, L. (1978). Eye movements during the viewing of Necker cubes. *Perception, 7*, 575–581.

Fenk-Oczlon, G. (1989). Word frequency and word order in freezes. *Linguistics, 27*, 517–556.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth. *Lingua, 92*, 333–375.

Flores d'Arcais, G. B. (1975). Some perceptual determinants of sentence construction. In G. B. F. d'Arcais (Ed.), *Studies in perception: Festschrift for Fabio Metelli* (pp. 344–373). Milan, Italy: Martello-Giunti.

Gale, A. G., & Findlay, J. M. (1983). Eye movement pattern in viewing ambiguous figures. In R. Groner, C. Menz, D. F. Fischer, & R. A. Monty (Eds.), *Eye Movements and Psychological Functions* (pp. 145–168). Hillsdale, NJ: Lawrence Erlbaum Association.

Georgiades, M. S., & Harris, J. P. (1997). Biasing effects in ambiguous figures: removal or fixation of critical features can affect perception. *Visual Cognition, 4*, 383–408.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development, 1*, 23–64.

Gleitman, L. R., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts. *Cognition, 58*, 321–376.

Goldman-Eisler, F., & Cohen, M. (1970). Is N, P, and PN difficulty a valid criterion of transformational operations? *Journal of Verbal Learning and Verbal Behavior, 9*, 161–166.

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

Hanna, J. E., & Brennan, S. E. (2004). Using a speaker's eyegaze during comprehension: a cue both rapid and flexible. In *Proceedings from CUNY '04: The 17th annual CUNY conference on sentence processing*. College Park, MD.

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

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt onset in capturing attention. *Perception & Psychophysics, 43*, 346–354.

Kelly, M. H. (1986). On the selection of linguistic options (Doctoral dissertation, Cornell University, 1986). Dissertation Abstracts International, 47(8-B), 3527.

Kelly, M. H., Bock, J. K., & Keil, F. C. (1986). Prototypicality in a linguistic context: effects on sentence structure. *Journal of Memory and Language, 25*, 59–74.

Lakusta, L., & Landau, B. (2005). The importance of goals in spatial language. *Cognition, 96*, 1–33.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: John Wiley and Sons.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation (ACL-MIT Press series in natural-language processing)*. Cambridge, MA: The MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

MacDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word order and world order: semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology, 25*, 188–230.

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 168–180.

Morgan, J. L., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple object naming. *Journal of Experimental Psychology: Language, Memory, and Cognition, 31*, 428–442.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research: Visual perception, 155*, 23–36.

Oliva, A. (2005). Gist of a scene. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 251–256). Academic Press, Elsevier.

Osgood, C. E., & Bock, J. K. (1977). Salience and sentencing: some production principles. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory* (pp. 89–140). Hillsdale, NJ: Erlbaum.

Paul, H. (1970). The sentence as the expression of the combination of several ideas. In Blumenthal, A. L., (Trans.), *Language and psychology: Historical aspects of psycholinguistics* (pp. 20–31). New York: Wiley (Original work published 1886).

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113*, 426–442.

Pomplun, M., Ritter, H., & Velichkovsky, B. M. (1996). Disambiguating complex visual information: towards communication of personal views of a scene. *Perception, 25*, 931–948.

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

Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 478–489.

Renninger, L. W., & Malik, J. (2004). When is scene recognition just texture recognition? *Vision Research, 44*, 2301–2311.

Rosenblood, L. K., & Pulton, T. W. (1975). Recognition after tachistoscopic presentations of complex pictorial stimuli. *Canadian Journal of Psychology, 29*, 195–200.

Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: a crosslinguistic study of word order and inflections. *Cognition, 12*, 229–265.

Talmy, L. (1978). Figure and ground in complex sentences. In J. H. Greenberg (Ed.). *Universals of Human Language Syntax* (Vol. 4, pp. 25–49). Stanford: Stanford University Press.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Tomlin, R. (1997). Mapping conceptual representations into linguistic representations: the role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162–189). Cambridge: Cambridge University Press.

Trueswell, J. C., Nappa, R., Wessel, A. & Gleitman, L. *Social and linguistic contributions to verb learning*. To be presented at the annual meeting of the Society for Research in Child Development, March, 2007, Boston, MA.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Underwood, G., & Green, A. (2003). Processing the gist of natural scenes. *Cognitive Processing, 4*, 119–136.

VanRullen, R., & Thorpe, S. J. (2001). Rate coding vs temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation, 13*, 1255–1283.

Vecera, S. P., Flevaris, A. V., & Filapek, J. C. (2004). Exogenous spatial attention influences figure–ground assignment. *Psychological Science, 15*, 20–26.

Wundt, W. (1970). The psychology of the sentence. In Blumenthal, A. L. (Trans.) *Language and psychology: Historical aspects of psycholinguistics* (pp. 20–31). New York: Wiley (Original work published 1900).