Does language guide event perception? Evidence from eye movements

Anna Papafragou[1,*], **Justin Hulbert**[2], and **John Trueswell**[3]

[1]University of Delaware

[2]University of Oregon

[3]University of Pennsylvania

Abstract

Languages differ in how they encode motion. When describing bounded motion, English speakers typically use verbs that convey information about manner (e.g., *slide, skip, walk*) rather than path (e.g., *approach, ascend*), whereas Greek speakers do the opposite. We investigated whether this strong cross-language difference influences how people allocate attention during motion perception. We compared eye movements from Greek and English speakers as they viewed motion events while (a) preparing verbal descriptions, or (b) memorizing the events. During the verbal description task, speakers' eyes rapidly focused on the event components typically encoded in their native language, generating significant cross-language differences even during the first second of motion onset. However, when freely inspecting ongoing events, as in the memorization task, people allocated attention similarly regardless of the language they speak. Differences between language groups arose only after the motion stopped, such that participants spontaneously studied those aspects of the scene that their language does not routinely encode in verbs. These findings offer a novel perspective on the relation between language and perceptual/cognitive processes. They indicate that attention allocation during event perception is not affected by the perceiver's native language; effects of language arise only when linguistic forms are recruited to achieve the task, such as when committing facts to memory.

Keywords

spatial language; language and thought; visual world; eyetracking; motion; Greek; cross-linguistic differences

1. Introduction

How do humans talk about the visual world? In an obvious sense, what we talk about is limited by constraints on how we see the world, including basic biases affecting how we conceptualize objects and events. Most theories of cognition and language assume that core aspects of the human perceptual and cognitive machinery are universal: given innate maturational properties

***Corresponding author.** Address: Dept. of Psychology, University of Delaware, 108 Wolf Hall, Newark, DE 19716. Email: papafragou@psych.udel.edu. Phone: (302) 831 2752.

of the human brain and typical experiential input, our perception and conception of objects and events is expected to be largely the same across individuals regardless of the language (or languages) learned during childhood. Under this view, these core systems generate nonlinguistic event and object representations that are shared by members of different linguistic communities and form the starting point for the generation of event and object descriptions in language (Gleitman & Papafragou, 2005; Jackendoff, 1996; Levelt, 1989; Miller & Johnson-Laird, 1976; Pinker, 1989).

Despite the broad appeal of this position, the transition from event conceptualization (the nonlinguistic apprehension of the main aspects of an event) to sentence planning (the mobilization of structural/lexical resources for event encoding) to speech execution has remained a mystery, largely because the particulars of nonlinguistic event representations have been notoriously hard to specify (Bock, Irwin & Davidson, 2004; cf. Jackendoff, 1996; Lashley, 1951; Paul, 1886/1970; Wundt, 1900/1970). [1] It is only recently that researchers have begun to get an experimental foothold into exploring the relationship between event apprehension and event description, with these findings showing a surprisingly tight temporal coupling between these two interrelated processes. In the first experiment to explore the temporal interface between language production and event comprehension, Griffin and Bock (2000) recorded speaker's direction of gaze as they visually inspected and described static line drawings of simple actions (e.g., a picture of a girl spraying a boy with a garden hose) that could be described with either an active or a passive sentence. Analysis of the eye movements in relation to active/passive linguistic choices led to the conclusion that there exists an initial rapid-event/gist extraction stage (event apprehension) that is temporally dissociable from any linguistic planning stage. However, further eyetracking studies using picture description tasks have shown that these apprehension and linguistic formulation processes overlap temporally to a considerable extent: initial shifts of attention to event participants predict which participant will be mentioned first in the sentence (Gleitman, January, Nappa & Trueswell, 2007). Taken together, these and related findings show that preparing to speak has rapid differential effects on how people allocate visual attention to components of a scene: if people need to talk about what they see, they immediately focus on aspects of scenes which are relevant for purposes of sentence planning (for recent reviews, see Meyer & Lethaus, 2004; Griffin, 2004; Bock et al., 2004; Altmann & Kamide, 2004).

An important (but little studied) aspect of the interface between event apprehension and language production lies with the fact that there are considerable cross-linguistic differences in how fundamental event components are segmented and packaged into sentences (e.g. Talmy, 1985, 1991). This cross-linguistic variation raises two interrelated questions. First, how do speakers of different languages manage to attend to different aspects of the visual world and integrate them into linguistic structures as they plan their speech? Given the within-language evidence sketched above, it is reasonable to expect that language-specific semantic and syntactic requirements create different pressures on the on-line allocation of attention as speakers of different languages plan their speech. Accordingly, some commentators have proposed that such language-specific demands on the formulation of messages become automatized (at least in adult speakers) and shape the preparation of utterances even before the activation of specific lexical items (Bock, 1995; Levelt, 1989). This idea, otherwise known as "thinking for speaking" (Slobin, 1996, 2003), can be summarized as follows: Languages differ in a number of dimensions, most importantly in the kind of semantic and syntactic distinctions they encode (e.g., morphological markers of number, gender, tense, causality, and so on). Mature (i.e., adult) speakers know by experience whether their language requires such categories or not, and select the appropriate information in building linguistic messages. Thus

---

[1]One exception has been the successful investigation of the cognitive and neurological underpinnings of the production of single words (Dell, Schwarz, Martin, Saffran & Gagnon, 1997; Garrett, 1988; Levelt, 1989; Levelt, Roelofs & Meyer, 1999).

language-specific requirements on semantic structure become represented in the procedural knowledge base of the mechanism responsible for formulating event representations for verbal communication (see Levelt, 1989). This view predicts that there should be early effects on attention allocation in speakers of different languages as they prepare to describe an event.

A second issue raised by cross-linguistic variation is whether core aspects of event apprehension itself could be substantially shaped by the properties of one's native language *even in contexts where no explicit communication is involved*. According to this stronger hypothesis, the rapid mobilization of linguistic resources for language production ('thinking for speaking') may affect other, non-linguistic processes which interface with language, such as perception, attention and memory. For instance, of the many ways of construing a scene, those that are relevant for linguistic encoding may become 'privileged' both in terms of on-line attention allocation and for later memory and categorization (Whorf, 1956; and for recent incarnations, Gumperz & Levinson, 1996; Levinson, 1996). Unlike the universalist view outlined earlier, according to which conceptual/perceptual and linguistic representations are distinct and dissociable, this linguistic relativity position assumes that these levels of representation are essentially isomorphic (Pederson, Danziger, Wilkins, Levinson, Kita, & Senft, 1998). Language, by virtue of being continuously used throughout one's life, can thus come to affect an individual's "habitual patterns of thought" (Whorf, 1956) by channeling the individual's attention towards certain distinctions and away from others. Ultimately, "[t]he need to output language coded in specific semantic parameters can force a deep-seated specialization of mind" (Levinson, 2003, p. 291) — and linguistically encoded semantic-conceptual distinctions might create striking (and permanent) cognitive asymmetries in speakers of different languages (for further discussion, see Levinson, 2003; Majid, Bowerman, Kita, Haun & Levinson, 2004; cf. also Boroditsky, 2001; Lucy, 2001; McDonough, Choi & Mandler, 2003).

These cross-linguistic issues raise two different ways in which language might guide attention during event perception - the first more modest, the second deeper and more controversial.[2] Nevertheless, both these hypotheses have proven hard to evaluate: Very little is known yet about how sentence generation interacts with event apprehension in speakers of different languages, and the workings of 'thinking for speaking' have not been demonstrated cross-linguistically. Furthermore, cross-linguistic work on the linguistic relativity hypothesis has focused on off-line cognitive tasks and has not adequately addressed potential effects of language-specific structures on on-line event apprehension (see the reviews in Bowerman & Levinson, 2001;Gentner & Goldin-Meadow, 2003). In both cases, since event apprehension happens very quickly (Griffin & Bock, 2000;Dobel, Gumnior, Bölte & Zwitserlood, in press), studying how event perception and language make contact cross-linguistically requires methodologies that track how perceivers extract linguistic and cognitive representations from dynamic events as these events unfold rapidly over time.

Here we pursue a novel approach to cross-linguistic event conceptualization using an eye-tracking paradigm. Specifically, we investigate the language-thought relationship by monitoring eye movements to event components by speakers of different languages. Since eye fixations approximate the allocation of attention under normal viewing conditions, people's eye movements during event apprehension offer a unique window onto both the nature of the underlying event representations and the link between those representations and language. Our approach is two-fold: First, we look at how speakers of different languages visually inspect dynamic events as they prepare to describe them. Under these conditions, cross-linguistic

---

[2]These two ways in which language can affect cognitive processes are closely linked, since cross-linguistic (relativistic) effects of language structure on event perception depend on prior specialization of the sentence production mechanism to within-language encoding demands.

differences should impact the on-line assembly of event representations if 'thinking for speaking' affects sentence generation cross-linguistically. Second, we compare shifts in attention to event components during a nonlinguistic (memory) task to test whether nonlinguistic processes of event apprehension differ in speakers of different languages in accordance with their linguistic-encoding biases.

Our empirical focus is events of motion (e.g., sailing, flying. We chose motion events for two main reasons. First, motion scenes are concrete, readily observable and easily manipulated and tested; second, the expression of motion is characterized by considerable cross-linguistic variability. This variability includes asymmetries in *what* languages choose to encode (e.g., semantic information about the components of a motion event), but mostly *how* languages encode this information (e.g., in syntactic configurations vs. the lexicon; in verbs vs. adverbial modifiers), and *how often* the information is used (whether it is an obligatory grammatical category, a robust typological tendency, or an infrequent occurrence). Our focus is on a well-documented difference in both how and how often motion paths and manners are expressed across languages (Section 1.1) that has been linked to potential effects on non-linguistic event processing (Section 1.2).

1.1 Motion paths and manners cross-linguistically

All languages typically encode the path, or trajectory (e.g., *arrive, ascend*), and the manner of motion (e.g., *skate, fly*), but differ systematically in the way path and manner are conflated inside sentences. *Manner* languages (e.g., English, German, Russian, and Mandarin Chinese) typically code manner in the verb (cf. English *skip, run, hop, jog*), and path in a variety of other devices such as particles (*out*), adpositions (*into the room*), or verb prefixes (e.g., German *raus-* 'out'; cf. *raus-rennen* 'run out'). *Path* languages (e.g., Modern Greek, Romance, Turkish, Japanese, and Hebrew) typically code path in the verb (cf. Greek *vjeno* 'exit', *beno* 'enter', *ftano* 'arrive/reach', *aneveno* 'ascend', *diashizo* 'cross'), and manner in adverbials (*trehontas* 'running', *me ta podia* 'on foot', g*rigora* 'quickly').[3] The Manner/Path distinction is not meant to imply that the relevant languages lack certain kinds of verbs altogether (in fact, there is evidence that all languages possess manner verbs, though not path verbs; Beavers, Levin, & Wei, 2004). But the most characteristic (i.e., colloquial, frequent, and pervasive) way of describing motion in these languages involves manner and path verbs respectively.

The manner/path asymmetry in verb use is made more salient by the following compositionality restriction: while in Manner languages manner verbs seem to compose freely with different kinds of path modifiers, in many Path languages manner verbs, with some exceptions, do not appear with resultative phrases to denote bounded, culminated motion (e.g., Aske, 1989; Cummings, 1996; Slobin & Hoiting, 1994; Stringer, 2003). Thus the compact way of expressing motion in the English example in (1) is not generally available in Greek; the PP in (2) can only have a locative (not a resultative/directional) interpretation:

```
(1)    English      The bird   flew          to its nest.      FIGURE
MANNER V       PATH MODF(2)   Modern Greek      To puli   petakse    sti
folia tu.       'the bird  flew      to its nest'      FIGURE      MANNER
V       PATH
```

---

[3]This distinction is inspired by, though not exactly equivalent to, Talmy's distinction between Satellite-framed and Verb-framed languages (e.g., Talmy, 1991). Several commentators have pointed out that more complex constellations of morpho-lexical patterns actually underlie Talmy's typology, such that 'mixed' patterns can be found (e.g., the presence of serial or compounding verbs of the type $V_{MANNER}+V_{PATH}$ in Korean; Beavers et al., 2004; see also Kopecka, 2006; Noonan, 2003; Slobin, 2004; Wienold, 1995; Zlatev & Yangklang, 2004).

The roots of this constraint lie in the morphosyntax of telicity: Greek disallows non-verbal resultatives, or 'secondary predication', with only a few exceptions (for discussion, see Folli & Ramchand, 2002; Horrocks & Stavrou, 2007; Napoli, 1996; Snyder, 2001; Washio, 1997). In order to convey the bounded event in (1), Greek needs to either switch to a path verb and optionally encode manner in a modifier such as a present participle as in (3a), or break down the event into two separate clauses with a path and manner verb as in (3b). Since both options in (3) are dispreferred/'marked', the default way of encoding this event would be (3a) without the manner participle:

```
(3)    a.  To puli    bike     sti folia tu  (petontas).          'the
bird    entered   into its nest (flying)'         FIGURE      PATH
PATH GROUND   (MANNER)     b. To puli    petakse   ke   bike    sti
folia tu.        'the bird   flew    and   entered  into its
nest'        FIGURE    MANNER        PATH      PATH GROUND
```

In effect, then, the resultative frame constraint intensifies the manner/path asymmetry in the use of motion verbs across Manner and Path languages. Notice that this constraint does not affect the use of manner of motion verbs to describe simple, unbounded events (*The bird flew/ To puli petakse* are fine in English and Greek respectively).

1.2 Nonlinguistic representations of motion events

The typological differences outlined in the previous section affect how speakers habitually talk about motion. These differences are already in place as early as three years of age in Path vs. Manner languages (Allen, Özyürek, Kita, Brown, Furman, Ishizuka & Fujii, 2007; Slobin, 1996, 2003; cf. Naigles, Eisenberg, Kako, Highter & McGraw, 1998; Papafragou, Massey & Gleitman, 2002, 2006) and affect conjectures about the meanings of novel verbs in both children and adults (Naigles & Terrazas, 1998; Papafragou & Selimis, 2007). These lexicalization patterns have been hypothesized to be a powerful source of 'thinking for speaking' effects, since they represent different ways in which languages allocate their grammatical/lexical resources to a common semantic domain (Slobin, 1996). On this view, the specialization of the event conceptualization phase for language-specific properties can be triggered not only by obligatory grammatical categories (Levelt, 1996) but also by systematic typological patterns in mapping semantics onto lexico-syntactic classes.

There are independent reasons to assume that verb encoding biases, in particular, might exercise especially powerful pressures on message preparation: verbs determine the argument structure of a sentence so they can be particularly central to the process of preparing event descriptions. Thus, even though the same kind of semantic information is conveyed in the English example in (1) and its Greek equivalent that includes a manner modifier in (3a) (i.e., both languages refer to path and manner), the distribution of information into morphosyntactic units should drive differences in early allocation to manner when speakers of the two languages prepare to describe the relevant motion event (with English speakers attending to manner earlier for purposes of selecting an appropriate verb). Notice that the crucial difference lies not only in the order of mention of path and manner information, but also in the placement of the information inside or outside the sentential verb.

Could such lexicalization preferences 'percolate' into conceptual event encoding more broadly? Some commentators have taken this stronger position. For instance, Slobin has proposed that manner in speakers of Manner languages is a "salient and differentiated conceptual field" compared to speakers of Path languages, with potential implications for how manner of motion is perceived/attended to online and remembered (Slobin, 2004; cf. Bowerman & Choi, 2003). In a related context, Gentner and Boroditsky (2001, p. 247) have

noted that "verbs … —including those concerned with spatial relations — provide framing structures for the encoding of events and experience; hence a linguistic effect on these categories could reasonably be expected to have cognitive consequences". These views fit with a more general expectation that linguistic encoding categories are imposed upon incoming perceptual spatial information so that speakers might later describe spatial configurations using the resources available in their native language (Gumperz & Levinson, 1996; Levinson, 1996).

Some earlier studies have cast doubt on these relativistic claims. Papafragou, Massey and Gleitman (2002) examined motion event encoding in English- and Greek-speaking adults and young children through elicited production tasks and compared these results to memory for path/manner or use of path/manner in event categorization. Even though speakers of the two languages exhibited an asymmetry in encoding manner and path information in their verbal descriptions, they did not differ from each other in terms of classification or memory for path and manner (cf. also Papafragou, Massey & Gleitman, 2006). Similar results have been obtained for Spanish vs. English (see Gennari, Sloman, Malt & Fitch, 2002; cf. also Munnich, Landau & Dosher, 2001). However, one might object that tasks and measures employed in these studies bring together many known subcomponents of cognition (perception, memory encoding, memory recall) and thus may fail to adequately identify possible linguistically-influenced processes. According to the attention-based hypothesis we are considering, linguistic encoding preferences might affect the earliest moments of event perception. Specifically, the online allocation of attention while perceiving an event might be shaped by those aspects of the event that are typically encoded in the observer's native language, especially in verbs. This preference might emerge irrespective of other (possibly language-independent) biases in processing rapidly unfolding events. It seems crucial therefore to explore the time course of spatial attention allocation during event perception in both linguistic and nonlinguistic tasks.

To test whether cross-linguistic differences affect the way language users direct attention to manner of motion, we recorded eye movements from native Greek and English speakers as they watched unfolding motion events (both bounded and unbounded) while performing either a linguistic or a nonlinguistic task. Specifically, trials consisted of watching a three second video of an unfolding event, which then froze on the last frame of the video. At that time, participants had to either describe the event or study the image further for a later memory test. It is well known that the allocation of attention during scene perception depends upon which aspects of the scene are deemed important to achieve the task (e.g., Triesch, Ballard, Hayhoe & Sullivan, 2003). Given this, we carefully constructed our animated visual stimuli so that linguistically relevant manner and path information could be easily defined as distinct regions spatially separated from each other. If cross-linguistic differences affect the allocation of attention during event perception generally, we would expect English speakers to be more likely than Greek speakers to focus on manner information early and consistently in both the linguistic and nonlinguistic task. But if event perception is independent of language, we should see differences between English and Greek speakers only in the linguistic task (and only in bounded events, where the two languages differ in terms of the information typically encoded in the verb): such differences in attention allocation should be consistent with preparatory processes for sentence generation ('thinking for speaking'). During tasks that do not require description of the event (as in the nonlinguistic task), Greek and English speakers should behave similarly in terms of attention allocation.

Finally, it should be noted that a third viable outcome exists. During the dynamic unfolding of each event, it is possible that no differences between English and Greek speakers will be observed, even in the linguistic task — i.e., even when participants know they will have to describe the movie when it stops. The characteristic effects of sentence planning on eye

movement patterns might only occur just prior to speech production. Such an outcome would be inconsistent with the thinking-for-speaking hypothesis and would be most consistent with a strong universalist view. Prior studies of event description have found very early effects of linguistic-preparation on eye movements (Griffin & Bock, 2001; Gleitman et al., 2007), even when participants had to wait a fixed amount of time before describing the event (Griffin & Bock, 2001). However, these past studies used static pictures of events rather than dynamic displays. It is possible that the demands of interpreting an unfolding event in real-time preclude the attentional prioritization of language-relevant event components, i.e., thinking-for-speaking.

2. Experiment

2.1 Participants

Seventeen native English speakers and seventeen native Greek speakers participated in the experiment. The English speakers were Psychology undergraduates at the University of Pennsylvania and received course credit for participation. The Greek speakers were students or junior faculty at various Universities in the Philadelphia area and were paid $8 for participation. Data from three additional participants showed severe track loss in the eye-tracking records and were discarded.

2.2 Stimuli

Test items consisted of 12 short clip art animations depicting motion events. Each animation lasted three seconds ending with a beep. Motion lasted throughout the period of three seconds. The final frame of the animation then remained on the screen for two additional seconds.

All test stimuli involved instrumental motions (e.g., skating, sailing, skiing) performed by animate agents (humans or animals), which allowed us to easily define a spatial region associated with the manner of motion (i.e., the instruments themselves, e.g., the skates, the sail, the skis). For our (linguistic) purposes, manner of motion involving vehicles/instruments belongs to the same category as spontaneous manner of motion. In Greek, both are expressed by w*ith* (Gr *me*) prepositional phrases and, when encoded in verbs, their distribution is subject to the same syntactic constrains concerning bounded vs. unbounded events. We did not use spontaneous motion events such as running or jumping: since manner apprehension for these events requires global appreciation of the motion features (e.g., people can discriminate such events from simple point-light displays; Johansson, 1973), it would have been difficult to identify specific 'manner' regions for running or jumping in the clips. We also avoided demarcated visual paths in our stimuli (e.g., a winding road) since pilot data have shown that people do not spend time looking at intermediate points of such paths but focus attention at the beginnings and endpoints of motion trajectories.

Each test event had two variants, one bounded (e.g., Fig. 1a, a man skating to a snowman), the other unbounded (e.g., Fig. 1b, a man skating). In the bounded events, the moving character always followed a smooth and predictable path, such that the initial heading of this character was toward the spatial region occupied by the goal (e.g., the snowman). Motion paths included a variety of spatial relations (e.g., *to, into, over, under*). The relative position of the goal object (left-right, above-below the agent) was counterbalanced. All goal objects were inanimate. Unbounded motion videos were identical to bounded videos except that the goal object was removed.

We also created 12 filler events that did not involve pure motion (e.g., a woman knitting; see Appendix for descriptions of the full sets of test and filler items). A presentation list was generated such that six of the test items depicted bounded events and six depicted unbounded events. These trials were randomly intermixed with filler trials. A second list was generated

from the first list by swapping the conditions of the test trials (bounded/unbounded). Two additional reverse-order lists were also generated.

2.3 Procedure

Participants were randomly assigned to either the Linguistic or the Nonlinguistic condition and to one of the four presentation lists. In both conditions, participants were informed that they would watch a set of animated clips on a computer screen showing a person or an animal do something. They would hear a beep at the end of each animation followed by two seconds in which the last frame would remain on the screen. In the Linguistic condition, participants were asked to describe the clips freely after they heard the beep. In the Nonlinguistic condition, they were told to watch each video and further inspect the image after the beep. Nonlinguistic participants were also told that, after viewing all the clips, they would see a series of still images corresponding to each clip, and they would have to indicate whether the image was the same or different from the clip they had seen earlier.

All participants received the memory test at the end of the experiment (i.e., it was a surprise task for the Linguistic participants). During the memory test, participants saw a set of still images each of which was extracted from the mid-point of each of the original events. For filler events, the images were the same as the prior events. For test items, half of the bounded and half of the unbounded trials were altered and half remained the same. Altered bounded trials consisted of the still image extracted from its unbounded variant, and vice versa. In practice, this meant removing the goal from the bounded motion events (e.g., people who had seen the clip of a man skating to a snowman were shown an image of the same man simply skating), or introducing the goal when the original event was unbounded (e.g., people who had seen the clip of a man skating were shown an image of the same man skating to a snowman). Accuracy of responses was recorded.

In preparation for the memory task, participants in the Nonlinguistic condition went through two practice trials with non-motion events, one of which was the same and the other different from the original, before they viewed the set of 24 events.

Experiments were conducted in the participants' native language. For the Greek participants, the experimenter was a native Greek speaker (the first author) and instructions were in Greek throughout. The experiment was preceded by greetings and casual conversation in Greek between experimenter and participants.

Eye movements were recorded using an ISCAN ETL-500 remote table-top eyetracker trained on the participant's right eye. Stimuli were displayed on a 15-inch monitor at a 22-inch viewing distance. No fixation cross was used prior to the start of each animation. A computer analyzed the eye image in real-time, superimposing the horizontal and vertical eye position of the direction of gaze onto the video image of the stimuli. The stimuli video and the superimposed eye position, along with all auditory events, were recorded to tape using a frame-accurate digital video recorder (a SONY DSR-30). The videotape was later examined frame-by-frame by a trained coder, who coded looks to pre-determined critical components of each event. Utterances were transcribed and coded by the first author, a bilingual Greek-English speaker.

2.4 Coding

2.4.1 Linguistic (production) data—For each utterance, we coded the main verb of the utterance (Path, Manner or Other): Path verbs encoded the trajectory of the moving agent and/ or the relationship between the agent and a reference object (e.g., English *enter, exit, cross, pass, approach*, and their Greek counterparts *beno, vjeno, diasxizo, perno, plisiazo*); Manner verbs encoded details such as the speed, gait, rate or means of motion (e.g., English *fly, sail,*

drive, float, and their Greek equivalents *peto, pleo, odigo* and *eorume* respectively); Other verbs did not belong to either of the two categories. We also coded the types of event modifiers included (Path or Manner), as well as the surface order of verbs and modifiers in the utterance. Path modifiers included adverbials, particles and prepositional phrases denoting directional information (*down, across the ice rink, into a cave*). We included here a few locatives (*on a rink*) which, even though not directly encoding direction, can be used to imply it. We did not include ground objects for Greek Path verbs as separate Path modifiers (e.g., *kateveni tin plajia* 'goes-down the slope'). Manner modifiers included adverbials, gerunds and prepositional phrases (*flying, with skates*) as well as some locative phrases that mentioned the instrument (e.g., *in a car, in a hot air balloon, on a boat*).

2.4.2 Eye movement data—For coding participants' eye movements, we defined path regions on the basis of path endpoints, since that gave an especially clear criterion for paths. Pilot work revealed that one common eye movement around the start of animation was to 'project the path' of motion to the goal (where the moving character was heading). Thus, Path looks were defined as looks to this path-endpoint region. Path looks in the unbounded condition were defined the same way: as path-projecting looks to where the goal would have been if present.

We also classified instrument regions as depicting manner, since they were spatially separated from path endpoint objects and (to a certain extent) from the agents in the events. Eye movements to these instruments were therefore coded as Manner looks. Finally, we also coded looks to moving characters (Agents). Regions for Agents were distinct from regions for Manners (e.g., in the skating example of Fig.1, the area of the feet that included the skates was coded as Manner, and the rest of the body was coded as Agent).

We did not interpret looks to the trajectory of the Agent (i.e., the continuous region of space through which the Agent traveled) as indicative of the encoding of path information. The first reason is practical: people rarely fixate empty (objectless) regions of space. The second reason is theoretical: linguistic (and conceptual) encoding of a path is achieved via a reference point or reference points (e.g., a source, a goal, or a frame). In our examples, goals were prominent reference points in the bounded event examples of our stimuli and (as we shall discuss) were frequently mentioned by speakers to communicate path information.

We also chose not to include looks to the Agent when calculating looks to manner of motion. Visual inspection of an agent is done for many reasons (e.g., assessing if this entity is a central character in the event, as well as assessing the posture and internal motion of the character, which indeed are related to calculating manner of motion). The result is that the meaning of such looks (to an experimenter) is at best ambiguous. Moreover, our moving images, due to technical limitations, were always 'sliding' clipart images rather than animated characters with internal motion. In this situation, inspection of an agent to assess manner of motion would be limited to assessing the posture of the character. We therefore decided to construct spatially distinct components to the manner of motion, looks to which we felt provided fairly unambiguous evidence of a perceiver encoding a key aspect of the manner. Nevertheless, looks to Agents are reported below for completeness.

2.5 Results and Discussion

2.5.1 Linguistic data—Results from the production task are presented in Table 1. Verbal descriptions in the Linguistic task confirmed what is already known about English and Greek verb usage. For bounded motion events, English speakers were much more likely to produce manner verbs than Greek speakers (72% vs. 36% of responses respectively) while, for unbounded events, the two languages did not differ as widely (76% of all responses contained manner verbs in English vs. 54% in Greek). More specifically, in English, the most typical

pattern of responses for unbounded events was a bare manner verb (see (4a)), occasionally with additional manner or path modifiers (see (4b); despite the absence of a goal, many speakers in both languages added path information such as *along a river*, or *across the screen* to their description of unbounded motion scenes). In Greek, there was a similar preference (see (5a–b)), even though overall the use of manner verbs was slightly lower than in English:

```
(4)    a.  A man        is skating.          FIGURE      MANNER V       b.
A rabbit    is sailing (in a boat)  (along a river).         FIGURE
MANNER V    (MANNER MODF) (PATH MODF)(5)    a.  Enas anthropos  kani
patinaz.          'a man        does skating'        FIGURE
MANNER V      b.  Enas lagos   plei      (se mia varka)  (kata mikos enos
potamu).          'a rabbit    is floating (in a boat)      (along a
river)'        FIGURE        MANNER V     (MANNER MODF)   (PATH MODF)
```

Bounded events in English also overwhelmingly elicited a manner verb plus a path modifier (see (6a)), occasionally with an additional manner modifier (see (6b)). In Greek, however, bounded events were much less likely to elicit a manner verb; in fact, Greek speakers were equally likely to produce either a path or a manner verb (with accompanying modifiers, see (7a–b)) for events with a culminated path:[4]

```
(6)    a.  A man    is skating  to a snowman.          FIGURE    MANNER V
PATH MODF      b.  A rabbit  in a boat    is sailing under a bridge.
FIGURE    MANNER MODF   MANNER V   PATH MODF(7)    a.  Enas anthropos
plisiazi      ena xionanthropo   me patinia.      'a    man        is
approaching  a  snowman       with skates'       FIGURE        PATH
V        GROUND           MANNER MODF     b.  Enas lagos      plei
(se mia varka)  pros mia gefira.          'a rabbit      is floating (in
a boat)     towards a bridge'      FIGURE        MANNER V     (MANNER
MODF)    PATH MODF
```

The Greek-English verb difference confirms the typological asymmetry documented in prior studies of motion (Gennari et al., 2002; Naigles et al., 1998; Papafragou et al., 2002, 2006; Papafragou & Selimis, 2007; Slobin, 1996; Talmy, 1985).

2.5.2 Eye movement data—Given the theories sketched above, we were especially interested in whether Greek and English participants differed in how they prioritized the collection of manner- and path-relevant visual information during the perception of unfolding motion events. In particular, how did participants prioritize looks to the path endpoint region vs. the manner region? As we show below, our results indicate that Greek and English speakers do indeed allocate attention to these regions differently in accordance with their language, but only in the Linguistic task and only when the languages differ in terns of the kind of information is encoded in the verb (i.e., in the case of bounded motion events).

Before we proceed, we need to establish that participants were consistent when it came to commencing their verbal description in the Linguistic condition. This is important because we expect participants in this condition to inspect events while planning their linguistic productions. We coded the latency to the beginning of each utterance from the onset of the beep for a subset (n=13) of the participants in the Linguistic condition. The median latency

---

[4]Previous work reports a strong preference for path verbs when Greek speakers describe bounded events (Papafragou et al., 2002; Papafragou & Selimis, 2007). It is not clear why we do not see this preference in the present data. Potential reasons might involve the specific stimuli we used, the fact that our native Greek speakers were also fluent in English, or some other factor.

was 1267 ms. The shortest latency was 667 ms, and on the vast majority (88%) of trials people were speaking within 2000 ms of the beep onset. Thus, with respect to interpreting the eye movements in the time-course graphs that follow, essentially everyone was preparing to talk during the first second after the beep, and most participants started talking toward the second half of the 2nd second. The tightness in the distribution of utterance onsets is consistent with our claim that people were preparing a linguistic message during the perception of the unfolding event.

2.5.2.1 Attending to manner vs. path over time: Figure 2 and Figure 3 plot the difference in manner and path-endpoint looks on a frame-by-frame basis from motion onset for both bounded and unbounded events. Beginning with bounded events (Fig. 2), participants' behavior in the Linguistic condition showed effects of speech planning during event perception. Specifically, the strong typological asymmetry in motion verb content was reflected in attention allocation as soon as motion began, with Greek speakers being much more likely to look initially to the path endpoint (e.g., the snowman) compared to English speakers. After about a second and a half, Greek speakers turned their attention to manner (e.g., the skates), while English speakers focused on the path-endpoint, presumably as a result of preparing relevant post-verbal modifiers in each language. These eye movement patterns were repeated after the beep while people were describing aloud bounded events.[5]

Remarkably, in the Nonlinguistic task (Figure 2b), eye movements were nearly identical for both language groups as bounded events unfolded over time. As the figure shows, interrogation of these events generated nearly identical shifts in attention: both language groups exhibited an overall preference for path-endpoints over instrument regions. This overall preference for inspecting path endpoints could reveal a principled asymmetry in event apprehension; endpoints and other stable reference objects are necessary for defining a motion event (Miller & Johnson-Laird, 1976; Talmy, 1985; Jackendoff, 1996). This viewing preference could also be due to the specific visual characteristics of our displays: for instance, path-endpoint scoring regions tended to be larger than manner regions. (Preference for path-endpoints cannot be explained as an artifact of looking at animates more than inanimates, since all goal objects in our materials were inanimate.)

Differences did emerge late though, at the end of each Nonlinguistic trial after the event had finished and participants began studying the last frame for later recollection: Greek speakers were equally likely to focus their attention on either the manner or the path region whereas English speakers were concerned about studying the path endpoint. A plausible explanation of these differences is that the two populations constructed linguistic representations of the scenes before committing perceptions to memory. Recall that Greek speakers when describing these events were equally likely to offer either a manner or a path verb in their motion descriptions, while English speakers typically preferred manner verbs: their late looking patterns in the two languages therefore seem to reflect exactly those components of motion which are not typically encoded in the main verb. We return to the significance of this finding in our discussion of the memory data (Section 2.5.3).

Analyses of Variance (ANOVAs) were conducted on looking times to the critical components of the event during each of five separate 1-second intervals beginning from video onset for

---

[5]One might find the correspondence between looking patterns and verbal descriptions in the Linguistic condition less than straightforward. After all, English speakers seem to prefer to look at Path compared to Manner early on (the Path-Manner difference score in Fig.2A is greater than 0), despite the fact that they overwhelmingly respond with manner verbs; the same general bias for Path emerges in the Greek data, despite the fact that Greek speakers are equally likely to use manner and path verbs in bounded events. In interpreting the data from Fig.2A, one needs to bear in mind that there is a general bias for Path looks in both language groups: this bias appears again in the Nonlinguistic condition (see Fig.2B and discussion below). However, clear cross-linguistic differences emerge over and above this general bias; moreover, these differences are tightly coupled with cross-linguistic differences in preference for path vs. manner verbs.

both the Linguistic and Nonlinguistic tasks. ANOVAs had two factors: Language Group (Greek vs. English) and Region-Fixated (Pathendpoint vs. Manner). (All ANOVAs were conducted on arcsin transformations of looking time because the untransformed measure is bounded between 0 and 1 sec - see Trueswell, Sekerina, Hill and Logrip, 1999.) For the Linguistic task, a significant interaction between Language Group and the Region-Fixated emerged during the first second of animation (F1(1,16)=4.01, p = 0.03, η_p^2 = 0.20; F2(1,11) =10.14, p = 0.004, η_p^2 =0.48), such that Greek speakers fixated the path-endpoint (e.g., the snowman) more than English speakers (F1(1,16)=4.59, p = 0.02, η_p^2 = 0.22; F2(1,11)=9.36, p = 0.006, η_p^2 = 0.46). During the third second, a similar interaction obtained (F1(1,16)=3.43, p = 0.04, η_p^2 = 0.18; F2(1,11)=7.40, p = 0.01, η_p^2 = 0.40), such that English speakers fixated the path-endpoint region more than Greek speakers (F1(1,16)=5.30, p = 0.01, η_p^2 = 0.25; F2(1,11) =4.86, p = 0.02, η_p^2 = 0.44).

For the Nonlinguistic task, Region-Fixated (Path vs. Manner) did not reliably interact with Language Group until the fourth second, i.e., immediately after the animation froze (F1(1,14) =6.31, p = 0.01, η_p^2 = 0.31; F2(1,11)=5.86, p = 0.01, η_p^2 = 0.35), such that English speakers studied path-endpoints more than the Greek speakers (F1(1,14)=4.45, p = 0.02, η_p^2 = 0.24; F2 (1,11)=6.92, p = 0.01, η_p^2 = 0.39). The same statistical pattern held for the fifth second: a reliable interaction (F1(1,14)=9.66, p = 0.003, η_p^2 = 0.41; F2(1,11)=15.59, p = 0.0005, η_p^2 = 0.59) again with English speakers studying path-endpoints reliably more than Greek speakers (F1(1,14)=9.20, p = 0.003, η_p^2 = 0.40; F2(1,11)=8.28, p = 0.006, η_p^2 = 0.43).

As discussed already, unbounded events did not yield wide differences in linguistic descriptions. Accordingly, for these events, there were no major differences in eye movements between English and Greek speakers in either the Linguistic or the Nonlinguistic task (Fig. 3). ANOVAs identical to those described for bounded events were also conducted here. Region-Fixated did not interact significantly with Language Group during any of the five 1-second intervals.[6]

2.5.2.2 Early vs. global effects on attention in bounded events: Given the cross-linguistic differences observed above for bounded events, we were interested in whether these differences reflected differences in how manner vs. path information uptake was initially prioritized. To investigate this, we analyzed for each trial whether participants looked first to path-endpoint or manner region, ignoring all other regions (e.g., the agent, the background, etc.). For example, a sequence of looks to the agent, the path-endpoint and then the manner (instrument) would be coded as a 'Path First' look. Results are plotted separately for the Linguistic and Nonlinguistic condition in Figure 4. As shown in the figure, there was again a general bias to inspect paths first, in both the Linguistic (F1(1,16)=33.92, p =0.0005, η_p^2 =0.68; F2(1,11) =11.14, p = 0.006, η_p^2 = 0.50) and the Nonlinguistic condition F1(1,14)=9.45, p = 0.006, η_p^2 = 0.40; F2(1,11)=11.61, p = 0.006, η_p^2 = 0.51). However, there was an interaction between First Look Type (Path/Manner) and Language in the Linguistic condition, significant by subjects and marginal by items (F1(1,16)=5.50, p = 0.03, η_p^2 = 0.26; F2(1,11)=3.61, p = 0.08, η_p^2 = 0.25), such that Greek speakers were more likely than English speakers to arrive first at the path-endpoint by subjects but not by items (F1(1,16)=7.18, p = 0.01 , η_p^2 =0.31; F2(1,11)

---

[6]We considered the possibility that a difference in ease of naming of the Goal object in the two languages might affect our results. Since harder-to-name objects elicit longer looks from speakers (Griffin, 2001), such a difference might explain some of the cross-linguistic differences in the eye movement data for bounded events. To exclude this possibility, we obtained an indirect measure of ease of lexicalization for the Path-endpoint objects by looking at how consistent speakers within each language were in naming these objects. Goal objects in bounded events were named in 67% of the English responses and 77% of the Greek responses. Within this group of responses, we assigned a consistency score to each bounded event in the two languages; the score corresponded to the maximal proportion of responses that used the exact same word (e.g., "snowman") for the Goal object in that event. The mean consistency score for English was 97.7 and for Greek 93.1. This inter-speaker convergence confirms that the difference in eye movement patterns between bounded and unbounded events cannot be due to English-Greek differences in lexicalizing the Goal objects.

=2.56, p = 0.13, $\eta_p^2 = 0.19$), and English speakers were more likely than Greek speakers to arrive first at the manner region, though only marginally so (F1(1,16)=3.73, p = 0.7 , η_p^2 =0.19; F2(1,11)=4.55, p = 0.05, $\eta_p^2 = 0.29$). No such interaction emerged in the Nonlinguistic condition. This pattern is consistent with the hypothesis that attention allocation at the earliest stages of event apprehension is affected by linguistic encoding preferences but only when language is needed to complete the task.

Interestingly however, it is not the case that Greek and English speakers differ in how much time they spend overall inspecting path or manner information during the three second animation. This is illustrated by calculating the proportion of time spent looking at these two regions of interest (see Figure 5). As the figure shows, participants spent more time looking at paths than manners in both the Linguistic (F1(1,16)=19.71, p = 0.0005, $\eta_p^2 = 0.55$; F2(1,11) =4.47, p = 0.05, $\eta_p^2 = 0.29$) and the Nonlinguistic condition (F1(1,14)=28.45, p = 0.0005, $\eta_p^2 = 0.67$; F2(1,11)=23.35, p = 0.0005, $\eta_p^2 = 0.68$). Looking time however did not interact with Language for either task (see Fig.6).

Taken together, these results show that overall attention allocation in event perception does not differ across languages, even when participants are engaged in a linguistic task: what does differ is *when* people look to particular regions, i.e. which information is prioritized, during sentence planning. Specifically, as motion events are apprehended, the regions which attract participants' attention early show evidence for language-dependent preferences for encoding motion information in the verb (e.g., the where vs. how of motion in Greek and English verbs respectively).

2.5.2.3 Early vs. late reference to manner: We have been assuming throughout that the eye movement patterns observed for bounded events in the Linguistic condition reflect a deep semantic difference in how verb systems of different languages encode such events, as well as the order in which verb information appears in a language. An alternative, simpler explanation is that the order in which semantic information is mentioned when formulating appropriate descriptions in English or Greek is sufficient to explain these results without a special role for verb selection. We know that surface sentential order affects planning looks in production (Griffin, 2001). It might be then that cross-linguistic differences in looking patterns are solely the result of earlier encoding of manner information in English (verb position) compared to Greek (post-verbal modifiers).

To explore this possibility, we analyzed manner descriptions in a way that differs from our coding scheme in Table 1. Specifically, we classified manner information in bounded events as being encoded *early* in the sentence if it appeared (a) before the verb, either in the NP-subject (e.g., *a skater*) or a post-nominal modifier (*a man with skates*), or (b) in the main verb (*skated*). (Table 1 does not include manner encoding in sentential subjects and does not take into account the surface position of manner modifiers.)

Interestingly, we found that manner was actually encoded early in 72% of the Greek descriptions, and in 90% of the English descriptions. According to the order-of-mention hypothesis, Greek and English speakers should be equally likely to attend to manner early on during event apprehension for this set of trials. (Notice that all cases of early mention of manner are cases where manner is expressed before path-endpoint.) But if the eye movement patterns are mainly determined by the sort of semantic information preferentially encoded in the main verb, the English-Greek difference in terms of early attention to manner should persist in these trials.

For each of these early-manner trials, we analyzed whether participants looked first to path-endpoint or manner region, ignoring all other regions (e.g., the agent, the background, etc.). In

the English-speaking group, first looks were directed to path-endpoints 69% of the time and to manner regions 31% of the time. In the Greek-speaking group, first looks fell on path-endpoints 87% of the time and on manner regions 13% of the time. We conclude that the path-manner difference in eye gaze patterns in the Linguistic condition (bounded events) persists even when both language groups mention manner information early in the sentence (and place it before path-endpoint information). Thus the eye movement patterns seem to depend not only on a general preference to look to whatever information speakers need to describe the scenes in the order in which they need the information but also by deeper sentence planning processes driven by verb selection.

2.5.2.4 Attending to agents: Thus far we have only focused on the contrast between looks at the path-endpoint vs. the manner of motion. At any given point, however, increased looks at a certain region necessarily result in reduced looks to other regions. So the observed differences between English and Greek speakers in attending to manner and path-endpoint components of motion events could be related to differences in attending to the agent (i.e., the moving figure). To exclude this possibility, we analyzed the proportion of time spent looking to the agent region across all conditions. Results are presented in Figure 6 and Figure 7. As shown in the figures, English and Greek speakers spent almost identical amounts of time looking at the agent. This conclusion was supported by a series of ANOVAs performed on this set of data. For each 1-second segment shown in the Figure 6 and Figure 7, subject and item ANOVAs were conducted separately for the bounded and unbounded events, with Language (Greek, English) and Task (Linguistic, Nonlinguistic) as factors. These analyses revealed no main effects of Language or interactions between Language and Task with the exception of one time window; during the fifth second of the bounded events, the item ANOVA but not the subject ANOVA revealed a main effect of Language ($F1(1,30)=2.01$, $p = .17$; $F2(1,11)=5.32$, $p < .05$). It is interesting to note that the analysis did reveal some effects of task; looks to the agent were greater in Linguistic as compared to Nonlinguistic tasks regardless of Language in the first second of viewing the events (all p's <.05). The effect persisted during the second segment but only for the bounded events (p's < .05). Again, when people had to describe events they showed rapid effects of sentence planning, with participants preferentially looking at the subject/agent as soon as the event begins unfolding. Since both Greek and English produced SVO sentences, no interactions with Language would be expected, and none were observed. A similar pattern of rapid extraction of information relevant for sentential subjects has been observed elsewhere in the literature (e.g., Griffin & Bock, 2000).

2.5.3 Memory data—Five participants (2 English speakers, 3 Greek speakers) were dropped from analysis of the memory test because of a strong 'same' bias (i.e., they incorrectly judged 60% or more of the changed items as unchanged from the original video). Results from the remaining participants are presented in Table 2. Decreases in accuracy indicate that participants had trouble remembering if a path-endpoint had been present or absent in the original video. As seen in the table, participants were generally quite accurate at judging path information, with the exception of the Greek speakers in the Nonlinguistic condition. Indeed, ANOVAs revealed that the Greek speakers' poorer performance resulted in a significant effect of Language Group overall ($F1(1,25)=3.76$, $p = 0.91$, $\eta_p^2 = 0.13$; $F2(1,11)=11.95$, $p = 0.001$, $\eta_p^2 = 0.52$) but that this effect interacted with task, i.e., Linguistic vs. Nonlinguistic ($F1(1, 25) =4.95$; $p = 0.01$, $\eta_p^2 = 0.17$; $F2(1,11)=10.69$, $p = 0.004$, $\eta_p^2 = 0.49$). Simple effects revealed that the effect of Language group was significant in the Nonlinguistic condition ($F1(1,10) =4.53$, $p = 0.03$, $\eta_p^2 = 0.31$; $F2(1,11)=12.90$, $p = 0.001$, $\eta_p^2 = 0.54$) but not significant in the Linguistic condition (Fs < 1).

This pattern of memory errors in the Nonlinguistic condition suggests that Greek speakers' efforts to attend to both the manner and the path during encoding came at a cost, resulting in decreased recognition performance for the path-endpoint. That is, studying both the manner

and the path area when the video froze in the Nonlinguistic condition (Fig. 2b) made it less likely for Greek speakers to later recall information about the path-endpoint object.

3. General Discussion

The experiment reported in this paper introduced and explored a novel tool for investigating the language-thought interface by cross-linguistically analyzing participants' eye movements as they inspected ongoing dynamic events. Our findings support the conclusion that preparing for language production has rapid differential effects on how people allocate visual attention to components of a scene: if people need to talk about what they see, they are likely to shift their focus of attention toward those aspects of the scene that are relevant for purposes of sentence planning. Such effects of sentence preparation appear on top of broader (language-independent) biases in attending to event components (cf. the overall tendency of participants in both language groups to inspect the path endpoint more than the manner of motion throughout the videos). These findings support and extend similar results from prior studies examining the rapid mobilization of linguistic resources for language production (Gleitman et al., 2007; Griffin & Bock, 2000; Levelt, 1989; Meyer, Sleiderink & Levelt, 1998), otherwise known as 'thinking for speaking' (Slobin, 1996). Additionally, for the first time, our study reveals how cross-linguistic differences in event encoding impact the process of segmenting and attending to different event components during the formulation of linguistic messages: where languages differ from each other in how they encode event structure, this difference shows up in how speakers interrogate scenes during speech planning. Our results demonstrate that 'thinking for speaking' effects can be brought about by consistent language-specific lexicalization patterns, specifically by distinct verb typologies.

Perhaps most importantly, our findings show that the linguistic encoding of events does not appear to affect the moment-by-moment processes underlying event perception in situations where there is no need for linguistic communication. As our nonlinguistic task shows, when inspecting the world freely, people are alike in how they perceive events, regardless of the language they speak. This result is hard to reconcile with recent claims according to which framing events in language affects the way events are perceptually experienced (e.g., Levinson, 1996), or proposals which have raised the possibility that semantic distinctions typically encoded in the verbal predicates of a language can become more salient/accessible in the mental life of the speakers of that language (Bowerman & Levinson, 2001; Gentner & Boroditsky, 2001; Slobin, 2003). Rather our non-linguistic findings are consistent with accounts emphasizing universal aspects of event perception and cognition.

Overall, our data point to the conclusion that the way events are inspected in the visual world depends on the perceivers' goals: if perceivers are preparing to speak, then their attention is directed to linguistically relevant event components; otherwise, language-specific preferences for event encoding do not intrude into the ordinary processes of event apprehension. When they do surface, language-driven effects on attention seem to be quite specific: at least in the domain of motion events, which is characterized by differences in (path/manner) verb typology, typological choices do not affect the global allocation of attention to event components, but simply the order in which these components are attended to by observers when they prepare to speak.

Nevertheless, our data also reveal clear linguistic effects in tasks that do not overly involve linguistic communication. Recall that in our nonlinguistic task, language differences emerged in the latter portions of the eye movement record and in performance on our memory task. Our interpretation of this result is that linguistic representations were used to support rapid encoding in declarative memory after a motion event unfolded; furthermore, this rapid translation had implications for memory accuracy later on in our task. Although this process points to a role

for language-specific categories in situations where people do not speak, its nature does not easily fit with relativistic views: in preparation for a memory task, speakers seem to shift their attention to categories typically *not* encoded in their language (at least in the main verb). Moreover, during the unfolding of a dynamic event, preparation to commit something to memory did not result in language differences, which contrasts sharply with the case in which a dynamic event is watched while preparing to talk about it i.e., our linguistic condition, in which ample effects of language were observed. The fact that people rely on language resources for rapid, post-perceptual processing and storage of event representations is consistent with previous reports of linguistic mediation in memory tasks (Baddeley, 2003; Conrad, 1964; Paivio, 1990). Even though the scope of such mediation is still a matter of active investigation, we expect that in situations where rapid translation of motion stimuli into linguistic utterances is blocked (e.g., through linguistic shadowing, or in speeded judgment tasks), both on-line attention allocation and memory for path/manner of motion would be comparable across speakers of different languages, regardless of verb encoding preferences.

Acknowledgments

Appendix. Stimuli list

Bounded versions of test items include the components in brackets

TEST ITEMS

1. A clown is driving a car (into a cave).

2. A wolf is roller-skating (into a cage).

3. A sea captain is sailing (to an island).

4. A dog is sledding down a hill (through a fence).

5. A man is skiing (to a rocket).

6. A military man is flying in an airplane (to a red flag).

7. A man is skating (to a snowman) on an ice rink.

8. A man with binoculars is riding a hot air balloon (over power lines).

9. A cowboy is riding a camel (to an oasis).

10. A little boy is riding a motorcycle (to a Ferris Wheel).

11. A rabbit is in sailing in a boat (under a bridge).

12. A man with a turban is riding a magic carpet from the top of one building (to another).

FILLER ITEMS

1. A main is painting with pink paint on a canvas.

2. A main is sliding down a railing.

3. A guy in a green trench coat is winking.

4. A maid is dusting a bureau.

5. A man is shaking his head "no".

6. A man in a suit is cutting a piece of paper with scissors.

7. A large woman sitting in a chair is knitting.

8. A young boy is waving.

9. A girl is throwing a yellow ball over her head.

10. A man is taking out the trash to a trash can.

11. A lumberjack is chopping a tree with an axe.

12. A man with one red sock is mopping the floor.

References

Allen S, Özyürek A, Kita S, Brown A, Furman R, Ishizuka T, Fujii M. Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. Cognition 2007;102:16–48. [PubMed: 16442518]

Altmann, G.; Kamide, Y. Now you see it, now you don't: Mediating the mapping between language and the visual world. In: Henderson, J.; Ferreira, F., editors. The interface between language, vision and action: Eye movements and the visual world. New York & Hove: Psychology Press; 2004. p. 347-385.

Aske, J. Path predicates in English and Spanish: A closer look; Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society; Berkeley, CA: BLS; 1989. p. 1-14.

Baddeley W. Working memory and language: an overview. Journal of Communication Disorders 2003;36:189–208. [PubMed: 12742667]

Beavers, J.; Levin, B.; Wei, T. A morphosyntactic basis for variation in the encoding of motion events; Paper presented at the Workshop on Diversity and Universals in Language; Stanford University; 2004 May. p. 21-23.

Bock, JK. Sentence production: From mind to mouth. In: Miller, J.; Eimas, P., editors. Handbook of perception and cognition, Vol. 11: Speech, language, and communication. Academic Press; 1995. p. 181-216.

Bock, K.; Irwin, D.; Davidson, D. Putting first things first. In: Henderson, J.; Ferreira, F., editors. The interface between language, vision and action: Eye movements and the visual world. New York & Hove: Psychology Press; 2004. p. 249-317.

Boroditsky L. Does language shape thought? Mandarin and English speakers' conception of time. Cognitive Psychology 2001;43:1–22. [PubMed: 11487292]

Bowerman, M.; Choi, S. Space under construction: Language-specific spatial categorization in first language acquisition. In: Gentner, D.; Goldin-Meadow, S., editors. Language in mind: Advances in the study of language and thought. Cambridge, MA: MIT Press; 2003. p. 387-427.

Bowerman, M.; Levinson, S., editors. Language acquisition and conceptual development. Cambridge: Cambridge University Press; 2001.

Choi S, Bowerman M. Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. Cognition 1991;41:83–122. [PubMed: 1790656]

Conrad R. Acoustic confusion in immediate memory. British Journal of Psychology 1964;55:75–84.

Cummins S. Le mouvement directionnel dans une perspective d'analyse monosemique. Langues et Linguistique 1998;24:47–66.

Dell G, Schwarz M, Martin N, Saffran E, Gagnon D. Lexical access on aphasics and nonaphasic speakers. Psychological Review 1997;104:801–838. [PubMed: 9337631]

Dobel C, Gumnior H, Bölte J, Zwitserlood P. Describing scenes hardly seen. Acta Psychologica. (in press).

Folli, R.; Ramchand, G. Getting results: Motion constructions in Italian and Scottish Gaelic; Proceedings of the West Coast Conference on Formal Linguistics; Somerville, MA: Cascadilla Press; 2001. p. 101-114.

Garrett, M. Processes in language production. In: Newmeyer, F., editor. Linguistics: The Cambridge Survey. Vol. 3. Cambridge: Cambridge University Press; 1988. p. 69-96.

Gennari S, Sloman S, Malt B, Fitch T. Motion events in language and cognition. Cognition 2002;83:49–79. [PubMed: 11814486]

Gentner, D.; Boroditksy, L. Individuation, relativity and early word learning. In: Bowerman, M.; Levinson, S., editors. Language acquisition and conceptual development. Cambridge: Cambridge University Press; 2001. p. 215-256.

Gentner, D.; Goldin-Meadow, S., editors. Language and mind. Cambridge, MA: MIT Press; 2003.

Gleitman, L.; Papafragou, A. Language and thought. In: Holyoak, K.; Morrison, B., editors. Cambridge handbook of thinking and reasoning. Cambridge: Cambridge University Press; 2005. p. 633-661.

Gleitman L, January D, Nappa R, Trueswell J. On the *give* and *take* between event apprehension and utterance formulation. Journal of Memory and Language 2007;57(4):544–569. [PubMed: 18978929]

Griffin Z, Bock K. What the eyes say about speaking. Psychological Science 2000;11:274–279. [PubMed: 11273384]

Griffin Z. Gaze durations during speech reflect word selection and phonological encoding. Cognition 2001;82:B1–B14. [PubMed: 11672707]

Gumperz, J.; Levinson, S., editors. Rethinking linguistic relativity. New York: Cambridge University Press; 1996.

Horrocks G, Stavrou M. Grammaticalized aspect and spatio-temporal culmination. Lingua 2007;117:605–644.

Jackendoff, R. The architecture of the linguistic-spatial interface. In: Bloom, P.; Peterson, M.; Nadel, L.; Garrett, M., editors. Language and Space. Cambridge, MA: MIT Press; 1996. p. 1-30.

Johansson G. Visual perception of biological motion and a model for its analysis. Perception and Psychophysics 1973;14:201–211.

Kopecka, A. The semantic structure of motion verbs in French: Typological perspectives. In: Hickmann, M.; Robert, S., editors. Space in languages: Linguistic systems and cognitive categories. Amsterdam: Benjamins; 2006. p. 83-101.

Lashley, KS. The problem of serial order in behavior. In: Jeffress, LA., editor. Cerebral mechanisms in behavior. New York: Wiley; 1951. p. 112-136.

Levelt, W. Speaking. Cambridge, MA: MIT Press; 1989.

Levelt W, Roelofs A, Meyer S. A theory of lexical access in speech production. Behavioral and Brain Sciences 1999;22:1–37. [PubMed: 11301520]

Levinson, S. Frames of reference and Molyneux's question: Crosslinguistic evidence. In: Bloom, P.; Peterson, M.; Nadel, L.; Garrett, M., editors. Language and space. Cambridge, MA: MIT Press; 1996. p. 109-170.

Lucy, J. Grammatical categories and cognition: A case study of the linguistic relativity hypothesis. Cambridge: Cambridge University Press; 1992.

Majid A, Bowerman M, Kita S, Haun D, Levinson S. Can language restructure cognition? The case for space. Trends in Cognitive Science 2004;8(3):108–114.

McDonough L, Choi S, Mandler JM. Understanding spatial relations: Flexible infants, lexical adults. Cognitive Psychology 2003;46:229–259. [PubMed: 12694694]

Meyer, A.; Lethaus, F. The use of eye tracking in studies of sentence generation. In: Henderson, J.; Ferreira, F., editors. The interface between language, vision and action: Eye movements and the visual world. New York & Hove: Psychology Press; 2004. p. 191-211.

Meyer A, Sleiderink A, Levelt W. Viewing and naming objects: Eye movements during noun phrase production. Cognition 1998;66:B25–B33. [PubMed: 9677766]

Miller, G.; Johnson-Laird, P. Language and perception. Cambridge, MA: Harvard University Press; 1976.

Munnich E, Landau B, Dosher B. Spatial language and spatial representation: A cross-linguistic comparison. Cognition 2001;81:171–208. [PubMed: 11483169]

Naigles L, Terrazas P. Motion-verb generalizations in English and Spanish: Influences of language and syntax. Psychological Science 1998;9:363–369.

Naigles L, Eisenberg A, Kako E, Highter M, McGraw N. Speaking of motion: Verb use in English and Spanish. Language and Cognitive Processes 1998;13:521–549.

Napoli DJ. Secondary resultative predicates in Italian. Journal of Linguistics 1992;28:53–90.

Noonan, M. Motion events in Chantyal. In: Shay, E.; Seibert, U., editors. Motion, direction and location in languages. Amsterdam: Benjamins; 2003. p. 211-234.

Paivio, A. Mental representations: A dual coding approach. New York: Oxford University Press; 1990.

Papafragou, A.; Selimis, S. Lexical and structural cues for acquiring motion verbs cross-linguistically; Proceedings from the 32nd Annual Boston University Conference on Language Development; Somerville, MA: Cascadilla Press; 2007.

Papafragou A, Massey C, Gleitman L. Shake, rattle, 'n' roll: The representation of motion in thought and language. Cognition 2002;84:189–219. [PubMed: 12175572]

Papafragou A, Massey C, Gleitman L. When English proposes what Greek presupposes: The linguistic encoding of motion events. Cognition 2006;98:B75–B87. [PubMed: 16043167]

Paul, H. The sentence as the expression of the combination of several ideas. In: Numenthal (trans.), AL., editor. Language and psychology: Historical aspects of psycholinguistics. New York: Wiley; 1970. p. 20-31.(Originally publ. in 1886).

Pederson E, Danziger E, Wilkins D, Levinson S, Kita S, Senft G. Semantic typology and spatial conceptualization. Language 1998;74:557–589.

Pinker, S. Learnability and cognition: The acquisition of argument structure. Cambridge, MA: MIT Press; 1989.

Slobin, D. From 'thought and language' to 'thinking for speaking'. In: Gumperz, J.; Levinson, S., editors. Rethinking linguistic relativity. New York: Cambridge University Press; 1996. p. 70-96.

Slobin, DI. Language and thought online: Cognitive consequences of linguistic relativity. In: Gentner, D.; Goldin-Meadow, S., editors. Language in mind: Advances in the study of language and thought. Cambridge, MA: MIT Press; 2003. p. 157-191.

Slobin, DI. The many ways to search for a frog: Linguistic typology and the expression of motion events. In: Strömqvist, S.; Verhoeven, L., editors. Relating events in narrative: Typological and contextual perspectives. Mahwah, NJ: Erlbaum; 2004. p. 219-257.

Slobin, D.; Hoiting, N. Reference to movement in spoken and signed languages: Typological considerations; Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society; Berkeley: BLS; 1994. p. 487-505.

Snyder W. On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. Language 2001;77:324–342.

Stringer, D. Acquisitional evidence for a universal syntax of directional PPs; Proceedings from ACL-SIGSEM Workshop: The linguistic dimensions of prepositions and their use in Computational Linguistics formalisms and applications; Toulouse: IRIT; 2003. p. 44-55.

Talmy, L. Lexicalization patterns: Semantic structure in lexical forms. In: Shopen, T., editor. Grammatical Categories and the Lexicon. Vol. Vol. III. Cambridge: Cambridge University Press; 1985. p. 57-149.

Talmy, L. Path to realization: A typology of event conflation; Proceedings of the Berkeley Linguistics Society; 1991. p. 480-519.

Triesh J, Ballard DH, Hayhoe MM, Sullivan BT. What you see is what you need. Journal of Vision 2003;3:86–94. [PubMed: 12678628]

Trueswell J, Sekerina I, Hill N, Logrip M. The kindergarten-path effect. Cognition 1999;73:89–134. [PubMed: 10580160]

Washio R. Resultatives, compositionality, and language variation. Journal of East Asian Linguistics 1997;6:1–49.

Whorf, BL. Language, Thought and Reality. Carroll, J., editor. Cambridge, MA: MIT Press; 1956.

Wienold, G. Lexical and conceptual structure in expressions for movement and space. In: Egli, U.; Pause, P.; Schwarze, C.; von Stechow, A.; Wienold, G., editors. Lexical knowledge in the organization of language. Amsterdam/Philadelphia: Benjamins; 1995. p. 301-340.

Wundt, W. The psychology of the sentence. In: Blumenthal (trans.), AL., editor. Language and psychology: Historical aspects of psycholinguistics. 1970. p. 20-31.Originally publ. in 1900.

Zlatev, J.; Yangklang, P. A third way to travel: The place of Thai in motion-event typology. In: Strömqvist, S.; Verhoeven, L., editors. Relating events in narrative: Typological and contextual perspectives. Mahwah, NJ: Erlbaum; 2004. p. 159-190.

Animation unfolds (3 s)   BEEP   *Animation freezes (2 s)*
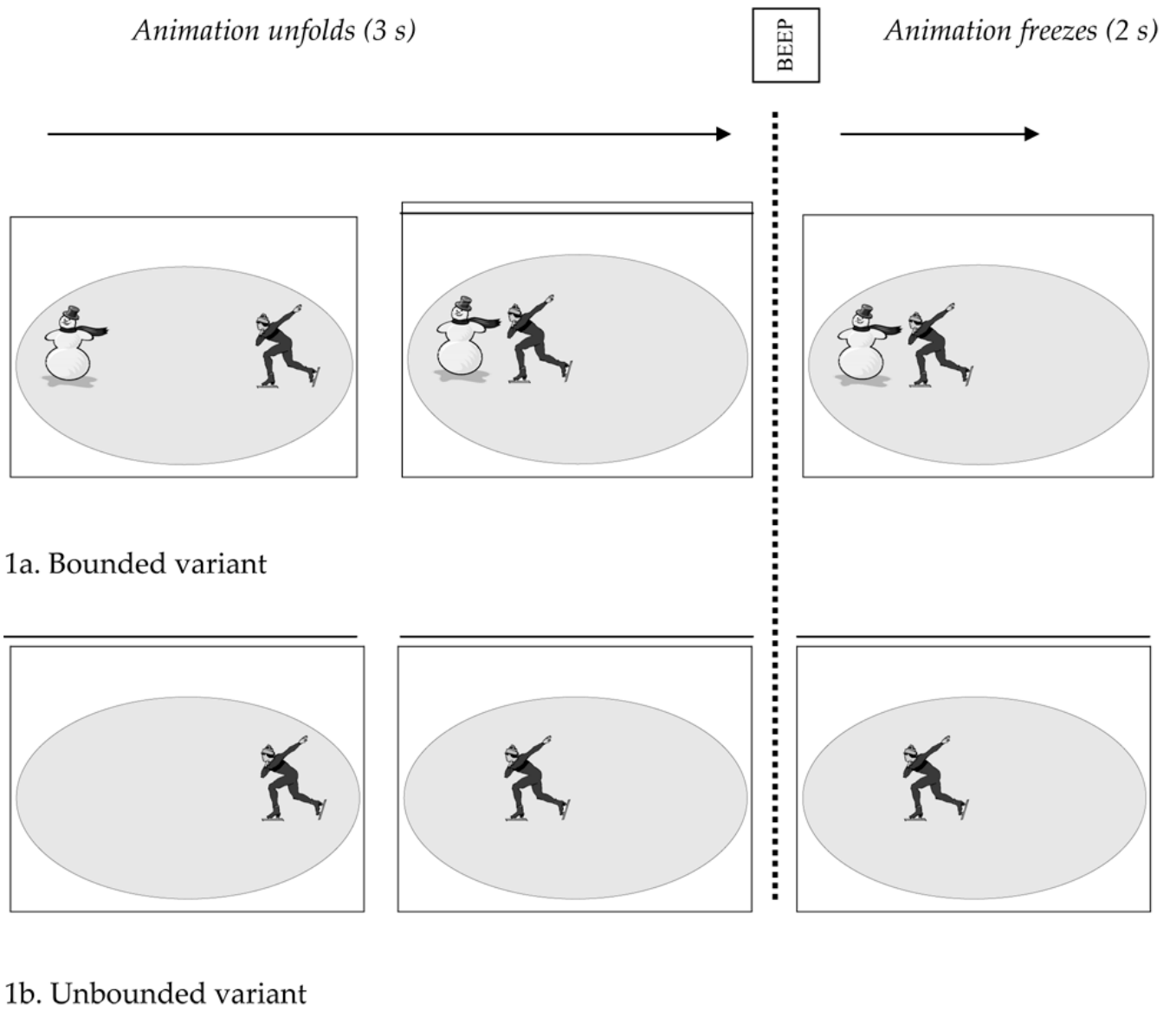

1a. Bounded variant

1b. Unbounded variant

Figure 1. Schematic representation of a stimulus event
(a) bounded variant and (b) unbounded variant. (The animation is presented here as a series of static clips.)
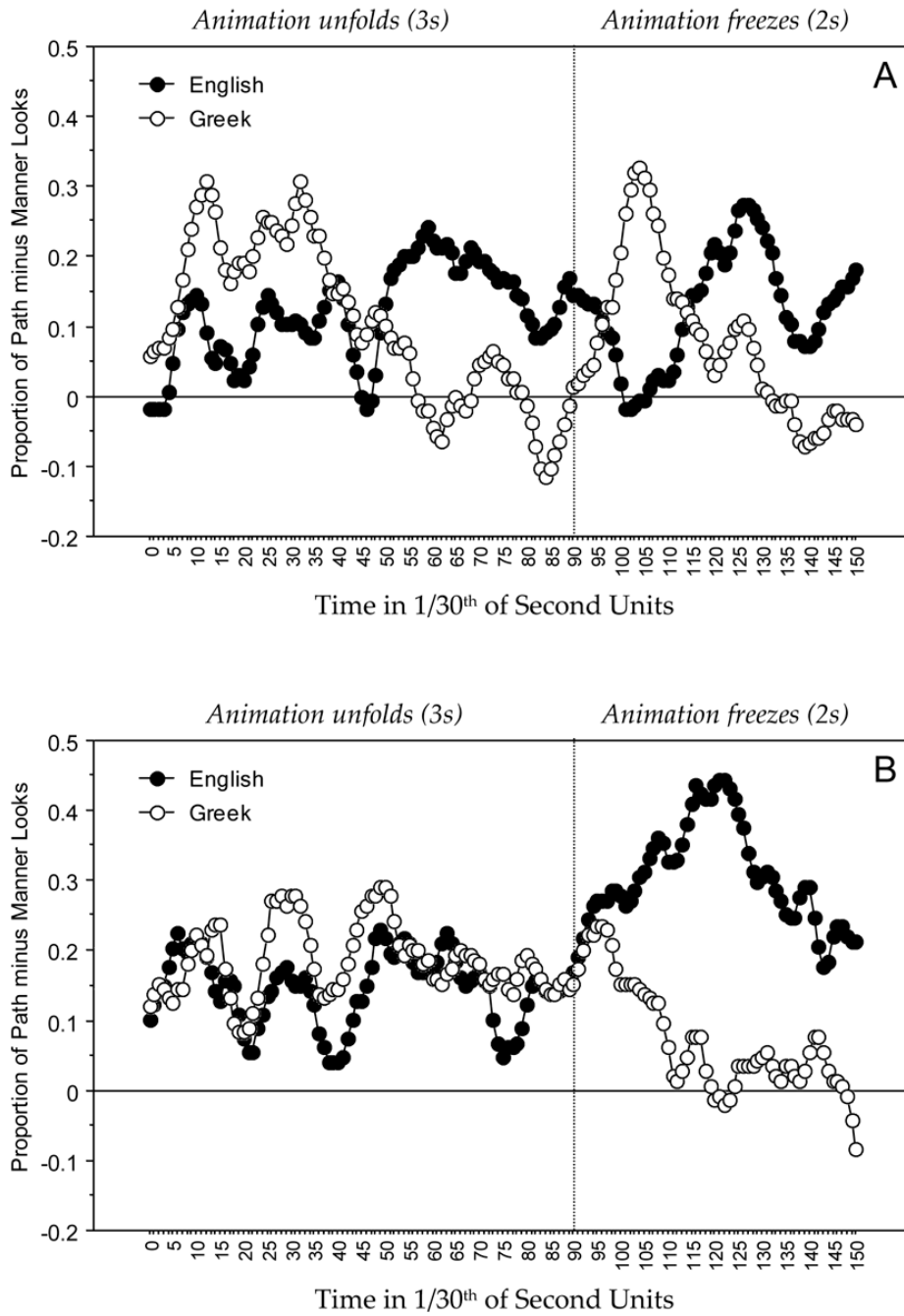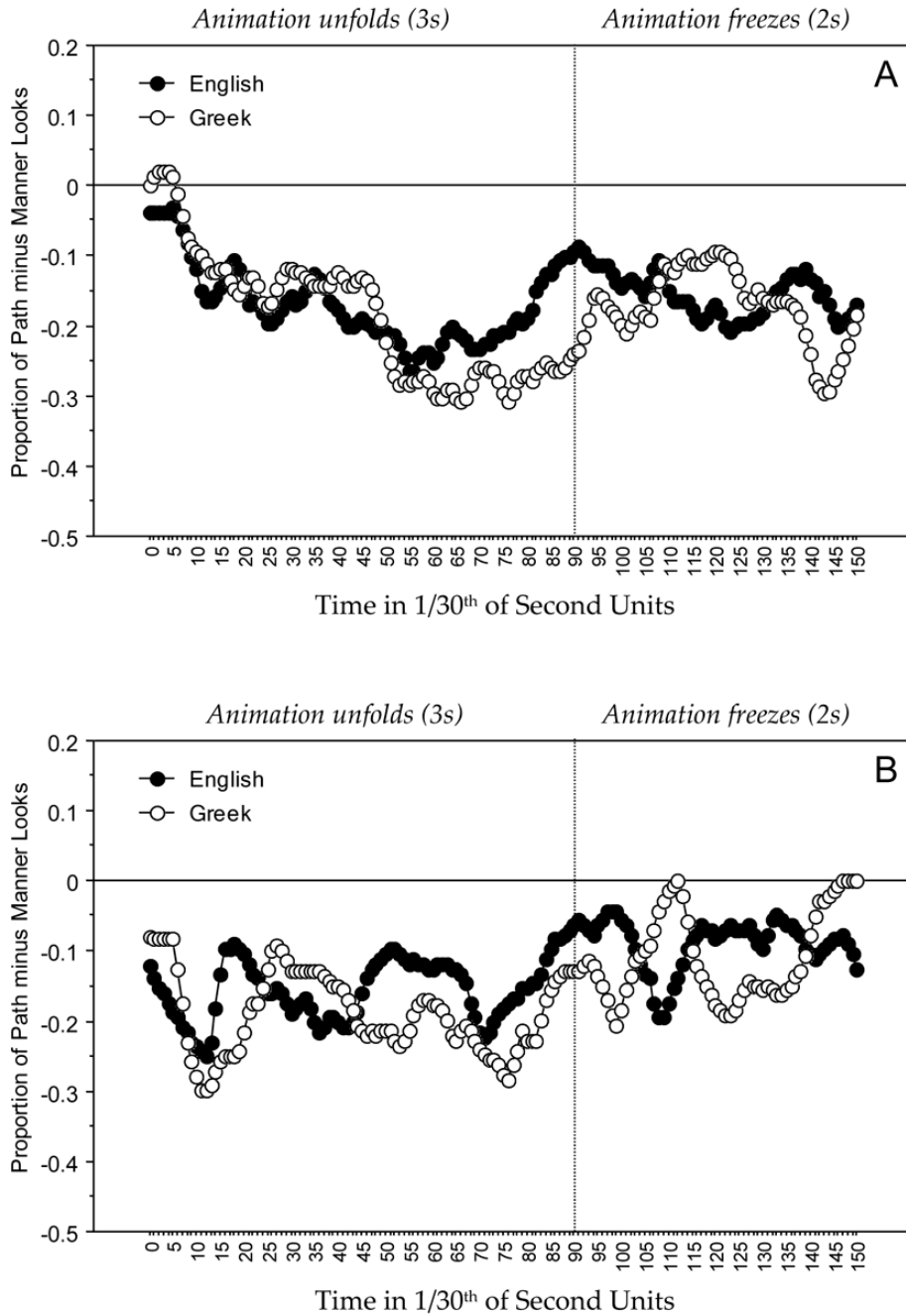
Figure 2. Eye movement results for bounded events
(A) Linguistic task, (B) Nonlinguistic task. Plotted over time from video onset, on a frame-by-frame basis, is the proportion of trials for which participants fixated the path-endpoint component of the event (e.g., the snowman) minus the proportion of trials participants fixated the manner component (e.g., the skates). Positive values reflect a preference to look at the path-endpoint, negative values reflect a preference to look at the manner.

Figure 3. Eye movement results for unbounded events
(A) Linguistic task, (B) Nonlinguistic task. Plotted over time from video onset, on a frame-by-frame basis, is the proportion of trials for which participants fixated the path-endpoint component of the event minus the proportion of trials participants fixated the manner component. Results are largely negative (biased toward Manner looks) because looks to the Path-endpoint constituted looks to the empty region where the endpoint object would have been if it had been present (e.g., where the snowman had been in the bounded event condition).
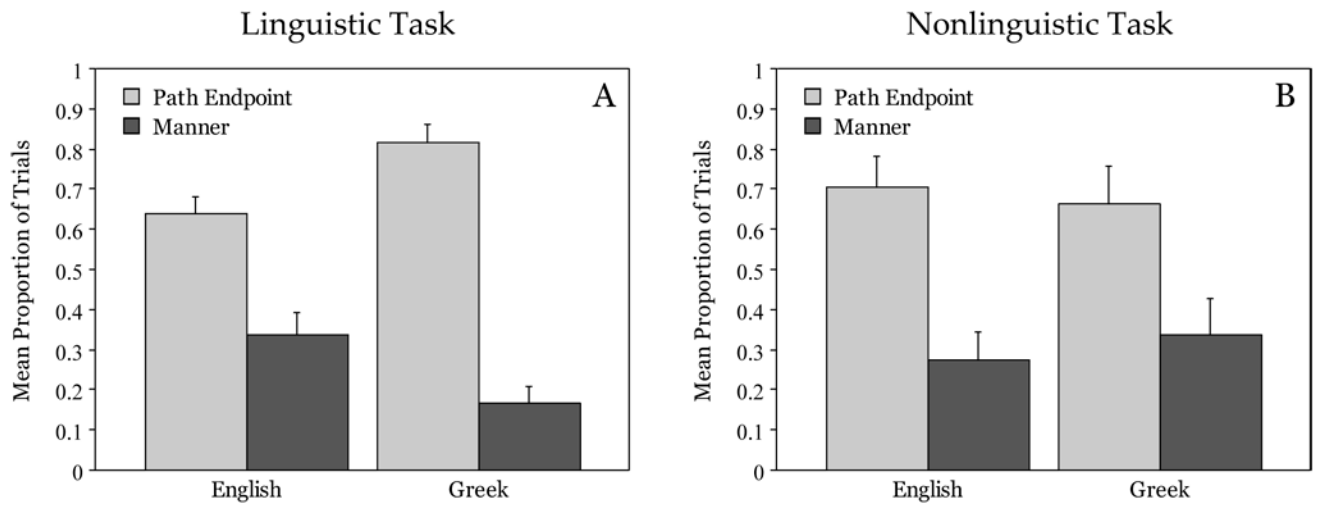
Figure 4. First Look analysis for bounded events
(A) Linguistic task, (B) Nonlinguistic task. Average proportion of trials that participants looked first to the Path Endpoint Region (e.g., the snowman) or the Manner Region (e.g., the skates) at video onset. (Error bars indicate standard errors.)
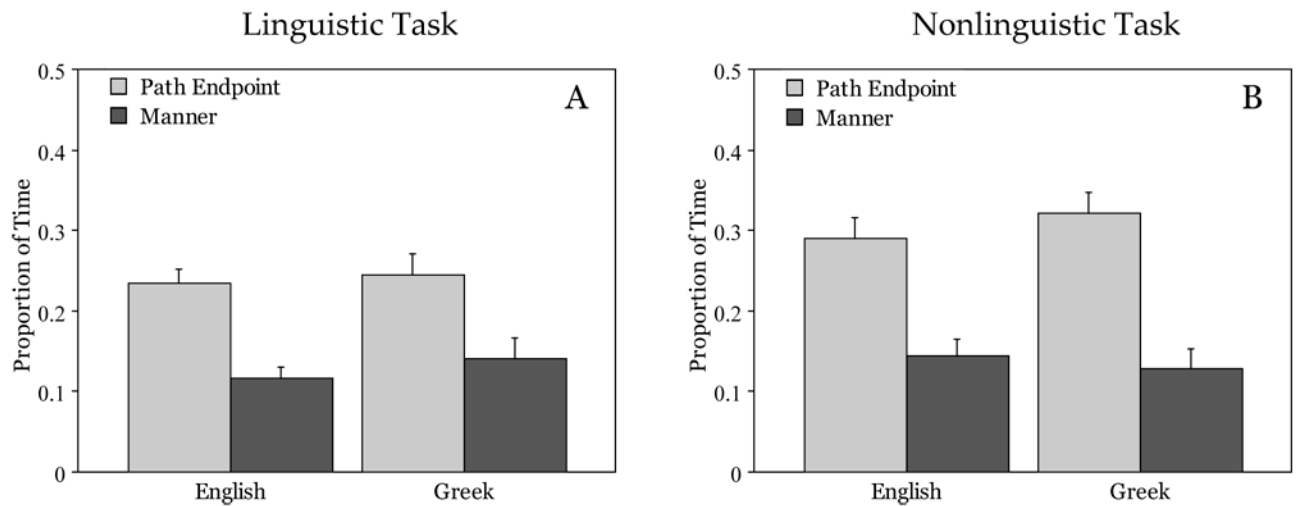
Figure 5. Looking Time for bounded events
(A) Linguistic task, (B) Nonlinguistic task. Average proportion of time that participants spent looking at the Path Endpoint Region (e.g., the snowman) or the Manner Region (e.g., the skates) during the 3 seconds of animation. (Error bars indicate standard errors.)
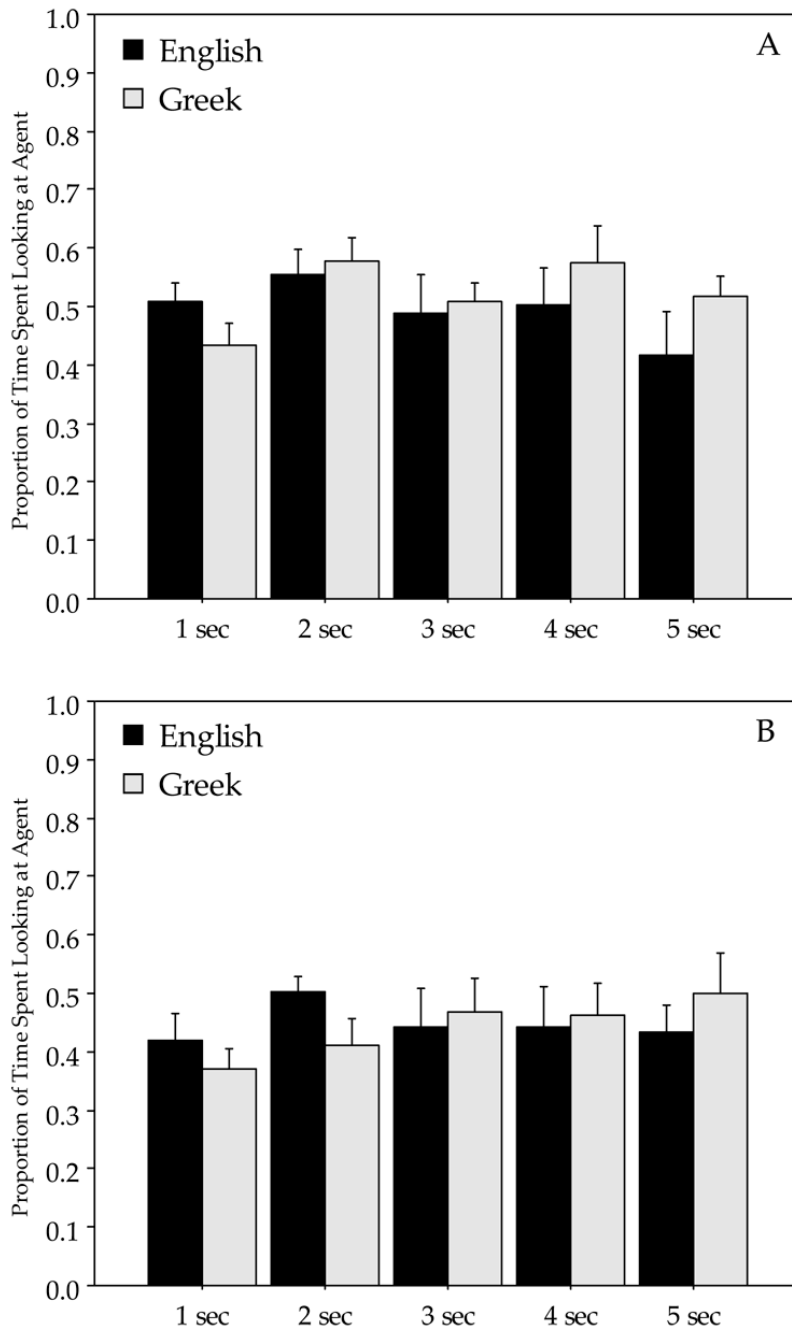
Figure 6. Proportion of time spent looking at agent for bounded events
(A) Linguistic task, (B) Nonlinguistic task. (Error bars indicate standard errors.)
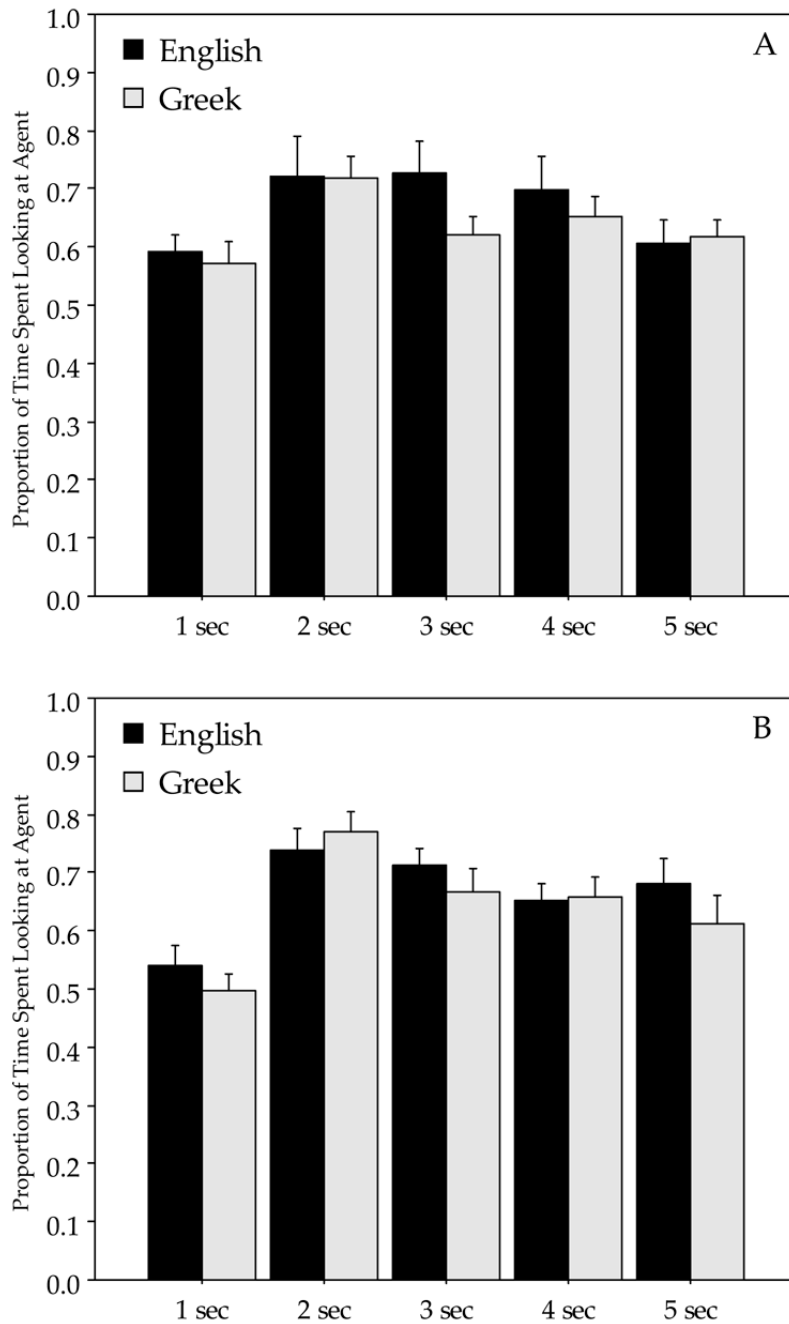
Figure 7. Proportion of time spent looking at agent for unbounded events
(A) Linguistic task, (B) Nonlinguistic task. (Error bars indicate standard errors.)

Table 1

Linguistic production data (percentage of responses)

Sentence structure	English		Greek	
	Bounded	**Unbounded**	**Bounded**	**Unbounded**
PATH Vs				
Path V only	0	0	5	0
Path V + Path modifier	2	0	6	1
Path V + Manner modifier	2	0	6	10
Path V + Manner modifier + Path modifier	8	6	22	0
Path V + Path modifier + Manner modifier	4	6	0	0
Path Vs total	*16*	*12*	*39*	*11*
MANNER Vs				
Manner V only	6	18	7	22
Manner V + Manner modifier	8	18	3	16
Manner V + Path modifier	38	22	15	11
Manner V + Manner modifier + Path modifier	12	6	11	5
Manner V + Path modifier + Manner modifier	8	6	0	0
Manner Vs total	*72*	*76*	*36*	*54*
MIXED Vs (Path V+ Manner V)	4	2	3	0
OTHER Vs	4	14	13	31
Total	***100***	***100***	***100***	***100***

Table 2

Percentage of correct responses in Memory task

	Type of Event	
	Bounded	**Unbounded**
Linguistic Condition		
Native English Speakers	92	91
Native Greek Speakers	92	92
Nonlinguistic Condition		
Native English Speakers	90	88
Native Greek Speakers	73	71