Perceiving and remembering events cross-linguistically:

Evidence from dual-task paradigms

John C. Trueswell

University of Pennsylvania

Anna Papafragou

University of Delaware

Address correspondence to: John C. Trueswell, Department of Psychology, University of Pennsylvania, 3401 Walnut Street C-Wing, Philadelphia PA 19104 (E-mail: trueswel@psych.upenn.edu)

ABSTRACT

What role does language play during attention allocation in perceiving and remembering events? We recorded adults' eye movements as they studied animated motion events for a later recognition task. We compared native speakers of two languages that use different means of expressing motion (Greek and English). In Experiment 1, eye movements revealed that, when event encoding was made difficult by requiring a concurrent task that did not involve language (tapping), participants spent extra time studying what their language treats as the details of the event. This 'linguistic encoding' effect was eliminated both when event encoding was made easier (no concurrent task) and when the concurrent task required the use of language (counting aloud). In Experiment 2, under conditions of a delayed concurrent task of counting aloud, participants used language covertly just prior to engaging in the additional task. Together, the results indicate that language can be optionally recruited for encoding events, especially under conditions of high cognitive load. Yet, these effects are malleable and flexible and do not appear to shape core biases in event perception and memory.

When we inspect a picture, or almost any visual scene, our eyes rapidly dart from person to person, place to place, object to object. Research into understanding the dynamics of scene perception indicates that these eye movements, although partially driven by bottom-up visual factors, reflect goal-directed categorization processes; the entities, events and states of affairs are placed into task-relevant categories, designed to achieve the person's immediate and longer-term goals (Yarbus, 1967; cf. Griffin & Bock, 2000; Henderson, 2003, 2007; Mennie, Hayhoe & Sullivan, 2007; Pelz & Canosa, 2001). The present paper follows this line of research, by asking: to what extent are task-relevant categorization processes that occur during scene perception necessarily or predominantly linguistic in nature? And if language is a typical currency for representing the world, does this fact influence what we attend to even when we are not using linguistic representations as a means of encoding?

Although there have been explorations into the role language plays in spatial reasoning and problem solving (e.g., Li & Gleitman, 2002; Levinson, 2003) and the memory and categorization of scenes/events (e.g., Gennari, Sloman, Malt & Fitch, 2002; Papafragou, Massey & Gleitman, 2002; Papafragou & Selimis, 2010), much less is known about whether and/or how linguistic knowledge plays a role in the dynamics of scene perception itself and the categorization processes that comprise it. For example, some potentially positive, albeit indirect, evidence that language does play a role in scene perception can be found in the early work of Loftus (1972), as well as from more recent work reported by Antes and Kristjanson (1993). These particular lines of research were not designed to test the relationship between language and scene perception (they instead explore the relationship between scene encoding and scene memory generally), yet both papers included experiments that could be interpreted as having some bearing on the issue of language. In these experiments, eye movements were recorded while participants studied static images of simple scenes, such as a farm, an urban neighborhood, etc. Participants were asked to view these images for a short time in preparation for a memory test, while either engaged in a concurrent task that involves language (e.g., counting backward by three) or not engaged in a concurrent task. Both Loftus (1972) and Antes and Kristjanson (1993) report that participants' eye movements and picture memory were negatively affected by this linguistic interference. Fixations became longer and, although insufficient details were provided about the specific gaze patterns of participants, the authors of both papers suggested that viewers spent less time inspecting important details of the images. Nevertheless, even when controlling for how often people looked at objects in these images, memory for the objects was impaired by linguistic interference, consistent with non-eyetracking studies of picture memory that also used interference tasks involving language during or after picture encoding (e.g., Allport, Antonis & Reynolds, 1972; Antes & Kristjanson, 1993; Rowe & Rogers, 1975; Wolfe, Horowitz & Michod, 2007). Such findings are only suggestive of a role for language however, largely because the studies were not designed to test the hypothesis that language is involved in perceiving and remembering scenes (e.g., direct comparisons with nonlinguistic interference tasks were not done).

The present study addresses the role of language in scene perception more directly, asking how and when language is recruited when people view depictions of simple events. These questions are particularly relevant in the context of a resurgence of interest in the relationship between language and perception/cognition (see Whorf, 1956; Gumperz & Levinson, 1996; Boroditsky, 2001; Levinson, 2003; cf. also papers in Bowerman & Levinson, 2001; Gentner & Goldin-Meadow, 2003). Until quite recently, most psycholinguists and cognitive psychologists believed, as part of their background assumptions for their theories of perception and language use, that linguistic encoding of the world was an optional process done either for purposes of communication (e.g., see Levelt, 1989) or as an aid for short- and long-term memory (e.g., see Miller, 1956; Paivio, 1971). In most other situations, lexical and syntactic characterizations of the world were not routinely deployed during perception, as these characterizations were believed to have little or no bearing on our physical interactions with the world.

Recently however, a number of researchers have questioned these background assumptions. For instance, it has been claimed that linguistic representations may actually be used as a matter of course, perhaps even automatically, in a range of nonlinguistic perceptual tasks, including online perceptual categorization of the features of objects and objects themselves (e.g., Gilbert, Regier, Kay & Ivry, 2008; Winawer, Witthoft, Frank, Wu, Wade & Boroditsky, 2007; cf. Roberson, Davidoff & Braisby, 1999). In particular, when discussing color categorization, Winawer and colleagues state that "language-specific categorical representations play an online role in simple perceptual tasks that one would tend to think of as being primarily sensory" and that these linguistic representations are "brought online spontaneously during even rather simple perceptual discriminations" (Winawer et al., 2007, p. 7784). Similar conclusions are drawn by Gilbert et al. (2008) who extend these conclusions beyond color categories to object categories such as dog and cat. They argue that the use of lexical information during object categorization is not a strategic process but rather "language affects discrimination on-line through the activation of lexical codes." Still others have argued that specific syntactic mechanisms are necessary for integrating feature/object information with location information (e.g., Hermer-Vazquez, Spelke, & Katsnelson, 1999). The processes of interest to these researchers (i.e., categorization of features/objects plus the integration with location information) are computations believed to be central to scene perception (e.g., Henderson, 2003). Thus these accounts predict that scene perception should routinely include the computation of linguistic information either because of automatic processes (such as spreading activation) or because the integrative encoding of the world requires the use of linguistic mechanisms.

Others have taken a more nuanced perspective on these issues. For instance, Slobin (2003, 2006) maintains the standard assumption about linguistic encoding being optional. However, he argues that it is a mistake to think that linguistic encoding of the world during perception is a rare or exceptional case for humans in their everyday lives. For instance, he notes that learning a first language most likely requires children to attend to conceptual contrasts more commonly encoded in their native tongue (e.g., language-specific contrasts relevant for encoding events) and not the contrasts less commonly encoded (Slobin, 2003). Slobin also notes that humans often anticipate a need to describe events later to other individuals, perhaps producing supplementary linguistic encoding of the events even when not currently talking about them. Slobin proposes that this pervasiveness of 'thinking for speaking' may in fact lead individuals to be especially attuned to conceptual contrasts made in their language, modulating attention to aspects of the world that are relevant to their particular language, even when linguistic encoding of the world is not occurring (Slobin, 2006). This latter conclusion suggests that language-specific effects on attention can have stable global consequences on our perceptual interrogation of the world.

These various positions have not always been clearly separated in the literature (nor are they easily separable in practice). Empirical studies have only just begun to investigate the role of language on event perception systematically (see the next section) but, so far, such studies have not attempted to distinguish between these theoretical options about the robustness of the role of language. In the present work, we address this gap by exploring the interface between language and perception/memory under different interference conditions and among speakers of different languages.

We take as our test bed the perception and memory of simple motion events (e.g., a man skating over to a hockey net) for which there are strong cross-linguistic differences in how they are routinely described (Talmy, 1975, and next section). Our own past work shows that differences also exist in how native speakers of different languages visually interrogate such scenes even when not asked to describe these events aloud (see below, and Papafragou, Hulbert & Trueswell, 2008). According to the traditional view of the relationship between language and attention outlined above, these cross-linguistic differences arise because subjects engaged in an optional strategy of linguistically encoding these perceived events in an attempt to aid memory. If so, such cross-linguistic differences ought to arise erratically when no severe processing load is placed on subjects at encoding (some subjects adopt the strategy, some do not), whereas these differences should be quite strong and consistent when subjects perceive the encoding task to be difficult and/or important (i.e., all should adopt a linguistic encoding strategy, resulting in strong cross-linguistic differences). Moreover, under such a view, cross-linguistic differences in attention should be eliminated under conditions of load when participants are prevented from doing linguistic encoding (e.g., a linguistic interference task). In contrast, for those theories that posit that linguistic encoding of events is so pervasive that it results in general attentional and/or conceptual biases, then the cross-linguistic patterns observed by Papafragou et al (2008) should be quite stable under a range of load conditions and should be immune to linguistic interference.

The cross-linguistic encoding of motion and its cognitive consequences

Consider the simple motion event depicted in Figure 1.

Insert Figure 1 about here

English speakers typically describe such a scene as:

(1) A boy is roller-skating into a hockey net.

Here, the manner of motion, roller-skating, appears in a syntactically prominent position, namely the matrix verb itself. The path (trajectory of motion), specifically the path endpoint (the goal), appears as an optional argument (*into the soccer net*); the sentence *A boy is roller-skating* is an acceptable, albeit less informative description of the same event. In contrast, speakers of Greek will tend to offer a different description, as in:

(2) Ena agori beni sta dihtia (me patinia).

'a boy enters in-the net (with skates)'

For Greek, the manner of motion is in an optional adjunct position (*me patinia*), whereas the path and path-endpoint are conveyed more centrally as the matrix verb and an argument PP *sta dihtia* (the spatial preposition in the PP *sta dihtia* is fairly underspecified, such that the main information about the spatial configuration of the path is carried by the verb). The Greek equivalent of *A boy is approaching a hockey net* is an acceptable but less informative rendering of the sentence in (2), and *A boy is approaching* is not acceptable.

These particular differences between Greek and English reflect a broader generalization regarding how events are encoded linguistically in the languages of the world (e.g., Talmy, 1975, 1985, 1991). All languages typically encode the path, or trajectory (e.g., *reaching/leaving* a point), and the manner of motion (e.g., *skating, flying*), 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*), adpositional phrases (*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 gerunds, adpositions, or adverbials (*trehontas* 'running', *me ta podia* 'on foot'; see Papafragou, Massey & Gleitman, 2002, for this terminology). The Manner/Path language distinction is not meant to imply that the relevant languages lack certain kinds of verbs altogether. But the most characteristic (i.e., colloquial, frequent, and pervasive) way of describing motion in these languages involves manner and path verbs, respectively.

These typological differences affect how speakers habitually talk about motion (e.g., Gennari et al., 2002; Özçalışkan & Slobin, 1999, 2003; Papafragou et al, 2002, 2006). These differences are already in place as early as 3 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 et al., 2002, 2006) and affect conjectures about the meanings of novel verbs in both children and adults (Naigles & Terrazas, 1998; Papafragou & Selimis, in press). Most relevant to the present research is a recent study from our own labs in which adult speakers of Greek and English described a set of short animated motion events very much like the one illustrated in Figure 1 (Papafragou et al., 2008). For English speakers, 78% of all sentences contained a manner verb as the matrix verb of the sentence, as compared to only 32% for Greek speakers. Unlike past cross-linguistic production studies, Papafragou et al. (2008) also recorded the eye movements of speakers as they watched the animations and as they proceeded to describe them. Quite interestingly, even early eye movements during the animations showed signs of this same cross-linguistic difference: Greek speakers were more likely than English speakers to fixate the path endpoint first rather than the manner of motion portion of the scene (projecting the path of motion to, for instance, the soccer net in Figure 1, rather than looking first at the roller-skates). This pattern is in accord with other eye movement production studies, which were done within a single language, where participants' preparation to describe aspects of a scene is preceded by fixations on these regions (e.g., Griffin & Bock, 2000).

Could such lexicalization and attention preferences 'percolate' into conceptual event encoding more broadly? Some have taken this stronger position (cf. Whorf, 1956, for an early version of the idea that habitual language use shapes cognition). For instance, Slobin has proposed that the manner of motion for 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 on-line 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." Advocates of these views then would expect similar differences in gaze patterns to be found cross-linguistically in motion event stimuli even when the task is not one of description. For instance, when simply studying these videos without describing them, speakers of Manner languages as compared to Path languages should look earlier and more often at manner of motion regions, whereas Path language speakers should prioritize the path endpoint more.

Papafragou et al. (2008) found that this prediction was not supported. In fact, preferences emerged in the opposite direction from those expected from language typology. Specifically, when people were asked to just study the videos for a later memory test (and hence not describe them aloud), attention allocation during the animations themselves was, contrary to the description task, strikingly similar for both language groups: generally speaking, people prioritized looks to the path endpoint and inspected the manner of motion slightly later (notably in accord with the proposed universal tendency to prioritize Goal over Source information in motion events, see Lakusta & Landau, 2005; Lakusta, Wagner, O'Hearn & Landau, 2007).

Nevertheless, the same study revealed evidence of linguistic intrusions in event encoding. Late in each trial (three seconds into the trial, after the animation ceased but the image remained on the screen), a striking 'reverse-Whorfian' effect emerged: speakers of English spent more time inspecting the path endpoint (e.g., the soccer net) rather than the manner of Motion (e.g., the roller-skates) as compared to Greek speakers who tended toward the opposite pattern. Papafragou et al.'s (2008) interpretation of this reverse-Whorfian effect was that it reflected an optional linguistic strategy (cf. Miller, 1956), in which participants, in addition to perceptually and conceptually encoding the information, also developed a linguistic representation late in the trial, which was used to help remember the event and to guide attention toward further 'details' in the scene. What was considered a detail was linguistically determined. For Greek speakers, the manner of motion is a 'peripheral' aspect of the description, since it is usually encoded outside the main verb and it can therefore be omitted, but for English speakers, manner of motion is a core aspect of the event since manner is usually encoded in the main verb and thus has a central part within a linguistic representation (and the opposite is true of the path of motion). This interpretation is bolstered by observations from the scene perception literature; in picture-studying tasks, participants typically begin by studying the central aspects of the scene, and then increasingly fixate the ‘details’ (e.g., Antes, 1974; Loftus, 1972; Antes & Kristjanson, 1993) as defined, for instance, by other participants who rate scene regions for relevance (Antes, 1974).[1]

The reverse-Whorfian effect, however, is also compatible with a stronger view of the role of language in cognition. It may be that viewers do indeed adopt a strategy to study the extra details late during viewing, but not based on linguistic representations but rather on fundamental nonlinguistic conceptual representations that have been shaped over the years by routine language use. Suggestive evidence in favor of this alternative explanation can be seen in the Papafragou et al. (2008) eye movement data. Early in viewing videos in the non-linguistic condition (during which no descriptions were required), trends of the ‘normal’ Whorfian sort were observed: Greek speakers were slightly more likely to fixate path endpoint regions over manner regions as compared to English speakers, though these differences were not statistically significant. Perhaps conceptual differences only manifest themselves in attention allocation based on what is perceived as ‘effortful’ to encode, i.e., information that lies outside the core event concept.

---

[1] In a different context, Talmy (1985) has argued that linguistic material in the main verb is backgrounded while material in optional non-verbal modifiers such as gerunds and adverbials is foregrounded/more salient (e.g., *He went to Hawaii by plane* foregrounds manner compared to *He flew to Hawaii*). From this perspective, one might expect language to guide attention to what is encoded outside the main verb. However, Talmy’s generalization picks up on the fact that atypical ways of describing events (e.g., through the addition of optional modifiers when a verb would do) lead to the Gricean expectation that there is something atypical about the event itself. Talmy himself points out that such foregrounding cases are rare since optional modifiers are often omitted (*He went to Hawaii* would be more canonical than *He went to Hawaii by plane*). We therefore focus on Slobin’s definition of salience that is based on what is frequently and canonically encoded within a language (i.e., information carried by the main verb; cf. also Gentner & Boroditsky, 2001).

The methods of Papafragou et al. (2008) cannot adjudicate between these two possibilities, and thus leave open the question whether the observed effects are temporary intrusions of language into non-linguistic representation or more stable adjustments of cognition to linguistic encoding practices. Here, as in other areas in which the language and thought debate has unfolded, one would need systematic dual tasks (using linguistic vs. non-linguistic interference) to resolve the nature of attested linguistic effects. In this paper, we pursue several types of such dual tasks to gain a better understanding of the behavior and source of potential linguistic intrusions during event apprehension. This approach advances the state of knowledge in this area by asking not simply whether language has an effect on event apprehension but what the scope and potency of such an effect might be.

Experimental prospectus

The two experiments that follow build on our earlier work (especially Papafragou et al., 2008) and explore the potential role of native language in the way members of different linguistic communities attend to dynamic motion events. The logic of our studies is as follows: if linguistic intrusions during event apprehension (such as the reverse-Whorfian effect observed in Papafragou et al., 2008) reflect a transient linguistic encoding strategy, one should be able to block the strategy by asking participants to perform a concurrent linguistic task (linguistic interference) as they inspect events. However, the linguistic encoding strategy should remain when participants are asked to perform an equally distracting concurrent task that does not involve language (non-linguistic interference); if anything, efforts to fixate the linguistically-defined details under a concurrent non-linguistic task should be *exaggerated* compared to situations in which there is no concurrent task (no interference). If, on the other hand, linguistic intrusions reflect fundamental and stable conceptual differences across language communities (shaped by years of native language use), then linguistic and nonlinguistic interference should have similar effects on attention allocation. Under this strong relativistic view, increasing cognitive load during event apprehension through a dual task (either linguistic or nonlinguistic interference) should simply encourage further examination of the event details, with the assessment of what is a detail being derived from language-influenced conceptual event representations.

The present experiments also explore potential implications of linguistic-typological differences for spatial memory. The two theoretical positions presented above make different predictions about how certain event components might be extracted from an ongoing event, tracked, and stored in memory. Some proponents of linguistic relativity expect linguistic encoding patterns to permeate spatial memory and to surface in both recognition and recall in spatial tasks (see Slobin, 2006). If so, cross-linguistic differences in manner and path endpoint encoding should lead to systematic language-congruent patterns in memory. For instance, English speakers should be better at remembering manner of motion compared to Greek speakers under normal viewing conditions (e.g., without interference). Furthermore, linguistic and nonlinguistic interference at encoding (to the extent that they are of comparable difficulty) should not differ in terms of how they impact this basic difference in event memory. In contrast, if language does not create a stable bias for extracting and remembering certain event components over others but can simply be used (when possible) as a tool for memory, then language-congruent differences in memory, if at all present, should be restricted to situations in which the task of holding on to an event representation is difficult but language mechanisms are still free to be used implicitly for encoding the event (i.e., when the secondary task is nonlinguistic). Prior work comparing memory for motion events in members of different linguistic communities has shown no effects of native language on memory (Papafragou et al., 2002; Gennari et al., 2002) but none of the earlier studies have explored dual tasks.[2] Other studies with English speakers

---

[2] Our earlier eye-tracking study reported in Papafragou et al. (2008) also tested participants' memory for dynamic motion events. However, the memory task only included gross changes in the endpoint region (inserting a goal when none was present or removing a goal originally present in the event). In that task Greek speakers were found to be somewhat

have shown that memory for motion events can be biased if path vs. manner verbs accompany the events, regardless of whether the verbs are provided by the experimenter (Billman & Krych, 1998) or generated by participants (Billman, Swilley & Krych, 2000). Of interest now is whether similar effects obtain in the absence of overt naming for events, and whether speakers of languages with different rates of use of path and manner verbs would differ in memory for path or manner aspects of motion. In what follows, we compare English and Greek speakers' rate of detecting changes to either manner or path endpoint of motion after encoding conditions that involve single (no interference) or dual (linguistic vs. nonlinguistic interference) tasks.

Experiment 1

Method

Participants

Sixty adults participated. Half were native speakers of Greek and half were native speakers of English. Both groups were approximately 18 to 22 years of age and enrolled at a university. Greek speakers were students at the University of Athens; English speakers were students either at the University of Delaware or the University of Pennsylvania. Participants spoke their native language routinely since they lived in a culture whose dominant language was either Greek or English. The vast majority of participants were monolingual (a second language may have been known only via course instruction and not immersion). Participation was voluntary. Minor differences existed in compensation: Greek speakers received a gift card for a coffee shop valued at four Euros (at the time,

---

worse than English speakers but the difference was also observed in the filler items. The present study contains a more stringent memory task (see next section) that includes both manner and path endpoint changes.

approximately US$6). English speakers received extra course credit for their participation. All participants gave informed written consent in their native language.

Apparatus

The same Tobii 1750 remote eyetracking system was used with all participants. This system does not use an eyetracking visor and instead tracks both eyes and head position via optics embedded in a flat panel 17" computer display. Two laptop computers were used, both running the Windows XP Operating System. One laptop displayed the video stimuli on the eyetracker screen (via the Tobii, Inc., ClearView software). A second laptop was dedicated to acquiring the eyetracking data from the Tobii eyetracker (via the TET-server software, developed by Tobii, Inc.). Both laptops were disconnected from the Internet to increase timing accuracy. The data sampling rate was 50 Hz, and spatial resolution of the tracker was approximately 0.5 degrees of visual angle. Timing accuracy of this eyetracker in this setup was assessed separately using a high speed camera trained on the eyes of an experimenter being eyetracked while running the program. Onsets of eye movements were hand coded frame by frame in the video and were computed relative to the onset of the stimuli by including the Tobii screen display in the same high speed video image. The results showed that the Tobii 1750 has an average delay in detecting an eye movement of 30 ms plus or minus one 20 millisecond sample.

Stimuli

A set of 12 target video triples were prepared for this experiment. Each consisted of a silent clipart animation, in which movement occurred for exactly three seconds, at which time a beep was heard. After the beep, the final (still) frame of the video remained on the screen for exactly two seconds (thus the total duration of each video was five seconds). All clipart animations were first created in Microsoft PowerPoint and then converted to Audio Video Interleave (avi) files using a conversion program. Timing was verified in video editing software.

Each target video depicted a simple motion event like the example illustrated in Figure 1 above (see the Appendix for a brief description of each target item). Target videos always contained a moving agent (e.g., a boy) propelled with the assistance of some instrument (e.g., roller-skates) to a stationary path endpoint (e.g., a soccer goal). A simple, contextually appropriate background was created for each video (e.g., a sidewalk in a neighborhood). Clipart images were constructed such that the instrument was spatially separated from the torso and face of the agent. This was achieved by having instruments be vehicles of various sorts (boats, planes, cars, etc.) such that agents' torsos and faces were always higher up than the body of the vehicle (e.g., in a car scene, the car was a convertible) . The spatial separation was done so that the eye tracker could distinguish looks to the regions of Agent vs. Instrument (see Analysis section below).

A separate sentence production study using these same target stimuli (Papafragou, Mojaverian & Trueswell, in prep.) confirmed that these stimuli elicit the usual cross-linguistic differences in the production of manner and path verbs. Ten native Greek speakers and 10 native English speakers described each of the 12 target videos among 12 filler videos. For the target videos, English speakers' utterances contained a manner verb without a path verb 62% of the time and a path verb without a manner verb 23% of the time. Greek speakers showed the opposite pattern: Their productions contained 17% manner-verb-only utterances and 51% path-verb-only utterances. (The remaining utterances contained either a combination of manner and path verbs, or an irrelevant – i.e., non-motion – verb. Verb combinations accounted for 5% of the responses in each language; irrelevant verbs were 10% of the English and 25% of the Greek responses.)

For the present study, two modified versions of each target video were also created. In one, the manner of motion was changed (Manner of Motion Change, e.g., the roller-skates were changed to a skateboard). In another, the path endpoint was changed (Path Endpoint Change, the soccer goal was changed to a small gazebo on the lawn; see Figure 2). The Path Endpoint Change always affected the spatial relationship between the moving agent and the path landmark in a linguistically relevant way: rather than going *into* the net, the boy went *next to* the gazebo. These variants were used in the memory test described below.

Figure 2 about here

Twelve filler videos were also created. These videos were designed to look like target videos in that they involved motion for three seconds, a beep, and a final still frame displayed for two seconds. The filler motion events also involved at least one animate entity and at least one object (e.g., a boy raking some leaves in a yard; a man putting a lid on a box; a frog jumping up and down in place next to a bench).

Procedure and Experimental Design

Each participant was run individually at his/her university campus. For Greek speakers, only Greek was spoken during the experimental session (the experimenter was author A.P., a native speaker of Greek). For English speakers, only English was spoken. The participant was given instructions specific to his/her task (see below) and was allowed to ask questions of clarification, which were answered by the experimenter. The participant was then seated in front of the Tobii 1750 eyetracking monitor. The eyetracker was calibrated by having the participant view a moving blue ball on the computer screen (i.e., a standard five-point calibration procedure, devised by Tobii within their ClearView software). The calibration procedure was repeated until the fit of the samples met the default criteria of the ClearView software. (Typically only one calibration was necessary.) Viewing distance was approximately 24 to 36 inches.

Participants were informed that they would be viewing a series of clipart animations and that at the end of the experiment there would be a memory test, in which they would view additional videos and judge if they were the same or different from the original videos. During the initial encoding phase, participants viewed a sequence of 24 clipart computer animations (consisting of 12 Target videos and 12 Filler videos, each five seconds in duration, see Stimuli above). At the end of the encoding phase, participants proceeded directly into the memory test phase. Here they viewed an additional set of 24 clipart animations, each either identical to one viewed previously or altered in some way (12 of these animations were the same and 12 were different from the original set). The participants' task was to judge whether the video was "old" or "new" (same or different). Responses were made verbally. Sessions were audio recorded, and the experimenter also wrote down responses during the session. The presentation of videos during the Encoding and Memory phases was experimenter-paced, such that the experimenter pressed a button to proceed to the next video (in the Memory phase, this happened after a response had been made).

Two stimulus lists were created. In each, the 24 encoding videos were shown in a fixed random order, and the 24 memory videos were shown in the same fixed random order. All 12 target memory videos contained a change during the memory test phase. Six were a Manner Change and six were a Path Endpoint Change (see Stimuli) and were randomly intermixed. All 12 filler memory videos contained no change. The second stimulus list was identical to the first except that in the test phase, Manner Change videos became Path Endpoint Change videos and vice versa.

Participants were randomly assigned to one of three possible encoding conditions:

No Interference. In this condition, participants freely viewed the animations during encoding. They had no task other than to inspect and remember the videos for the upcoming memory test.

Nonlinguistic Interference (tapping). This condition was identical to the No Interference task, except that, concurrent with the video onset of each trial, participants heard a pre-recorded series of drum beats making up a simple one-measure rhythm. Participants then used a pencil to tap out the same series of beats on the table in front of them, repeating the rhythm throughout the animation, the beep, and the still frame. For each of the events, participants heard (and had to reproduce) a different rhythm. Tapping requires sequential processing but does not involve linguistic elements (such as words) nor does it involve vocal articulation.

Linguistic Interference (repeating numbers aloud). This condition was identical to the No Interference task, except that, concurrent with the video onset of each trial, participants heard a string of three two-digit numbers (e.g., 45, 93, 77) and were asked to repeat the string out loud and continue repeating it out loud until the video disappeared (that is, during the three-second animation and during the two-second still frame). For each of the events, participants heard (and had to reproduce) a different string of numbers. This condition served to disrupt linguistic encoding of the event, given that counting aloud engages linguistic representations and linguistic processes. The intention here was to provide an interference task of similar complexity to the Nonlinguistic Interference task (tapping). Pilot work was done in advance of this experiment to find two tasks that according to subjective experience were similar in difficulty. The memory results below suggest that the two tasks were successfully matched in difficulty: overall, decreases in memory performance were very similar for these two interference tasks.

Eyetracking Coding and Analysis

Only eyetracking data from the Encoding phase were analyzed. Eyetracking samples (50 per second) were time-locked to the onset of each video. Track loss was determined separately for each eye by the ClearView Software. If track loss occurred for only one eye, the gaze coordinates from the other eye were used in analyses; if neither eye had track loss, the gaze coordinates from each eye were averaged. A given trial was dropped from further analysis if it had more than 33% track loss. If this process resulted in the dropping of five or more trials for a given subject, the data from that entire subject were dropped. Two of the 30 Greek subjects (one in the Nonlinguistic Interference task and one in the Linguistic Interference task) had to be dropped from the eye tracking analysis; none of the English subjects had to be dropped. For all included subjects, 3% of trials on average were excluded.

For each target video, three spatial scoring regions were defined (see Figure 1 for an example). These regions were: (1) Manner of Motion (Instrument), which was defined as an invisible rectangular region surrounding the object used by the agent as the means of motion (e.g., the roller-skates); (2) Path Endpoint, which was defined as an invisible rectangular region surrounding the stationary path endpoint of the motion event (e.g., the hockey net); and (3) Agent, which was defined as the head and torso of the actor in the event (e.g., the boy).

Unlike Papafragou et al. (2008), in which eye tracking output videos were coded by hand, an automatic eyetracking data analysis procedure was developed here that used moving scoring regions. That is, the position of the Manner of Motion region was updated in the eyetracking data file so as to match the motion of the animated character on the screen. This was achieved by having an experimenter use a pre-processing program that allowed her to define scoring regions frame-by-frame, once for each target video. On the first frame of the video, the experimenter drew a rectangular scoring region once around the object of interest, and the coordinates were stored to a file. This rectangle was then moved by the experimenter for each subsequent video frame, allowing for the same size region, but now in an updated coordinate position. The resulting coordinate file was then merged with the eyetracking data file. An eye position sample was considered within a region of interest if its coordinates fell within the region as defined for that moment in time in the video.

Scoring regions, and their movement over time, were identical across the experimental tasks within each item. Scoring regions were typically on the order of two to three degrees of visual angle in width and/or height, well within the 0.5 degree visual angle accuracy of the eyetracker. The Agent and Manner of Motion regions did sometimes overlap with the Path Endpoint region, especially at the end of the animation as these two regions moved “into” or “next to” the Path Endpoint (see Figure 1). At these points of overlap, gazes were counted as looks to the occluding object (e.g., the boy, the rollerskates) and not the occluded object (e.g., the hockey net).

Looks to the Agent are not reported because viewers could be inspecting this region for a variety of reasons: viewers might be looking at the Agent simply to identify the character, but they also could be assessing the character's posture, which partially informs both the Manner of Motion (e.g., the characteristic arm swings of a skater) and the Path (e.g., which way the person is facing). In contrast, the viewers' reasons for inspecting the Manner of Motion and Path Endpoint regions are considerably less ambiguous and as such more informative for testing our hypotheses.

Predictions

If, as we have argued, linguistic intrusions during event apprehension (such as the reverse-Whorfian effect observed in Papafragou et al., 2008) reflect a transient strategy of encoding event details in language, then participants should be more likely to employ this strategy under Nonlinguistic Interference than No Interference, given the higher cognitive load of the interference task. Furthermore, this linguistic encoding strategy should be blocked in the Linguistic Interference task – that is, even though cognitive load is also increased with a Linguistic Interference task, this task necessarily blocks the ability to encode the event linguistically. Concretely, in English speakers, the difference between Path and Manner looks (as defined by a Path minus Manner difference score) should be higher in the Nonlinguistic Interference compared to the other two conditions – since Path is encoded outside the main verb in English and would be considered a 'detail.' In Greek speakers, the opposite pattern should occur, with the Path-Manner difference being lower in the Nonlinguistic Interference condition compared to the other two conditions. Crucially, English speakers should differ most strikingly from Greek speakers in terms of attention allocation in the Nonlinguistic Interference condition (where the linguistic encoding strategy should emerge), but less so in the No Interference condition and not at all in the Linguistic Condition.

An alternative possibility is that linguistic intrusions reflect fundamental and stable conceptual differences across language communities (in other words, what is a detail in an event is shaped by years of native language use). If so, then the Linguistic and Nonlinguistic Interference tasks should have similar effects on attention allocation: in both tasks, participants should inspect what they consider to be event details more than the core event as compared to the No Interference condition. This account does indeed expect Linguistic Interference will block the linguistic encoding of the event; however, the blocking of linguistic encoding should not mask stable conceptual differences that were shaped by years of past language use. This means that, for English speakers, the difference between Path and Manner looks should be similar for both interference tasks; similarly, for Greek speakers, there should be no difference between the two interference conditions in terms of attention to Manner and Path. But, because of conceptual/attentional differences, English and Greek speakers should differ from each other – with English speakers having a higher Path minus Manner score as compared to Greek speakers.

Results and Discussion

Eye movements

Figure 3 presents the eye movement results for English speakers (panels A & C) and Greek speakers (panels B & D). Plotted is the average accumulation of time spent looking at the Path Endpoint (panels A & B) and Manner of Motion (panels C & D) regions as a function of how long the video had been playing. These two measures are not simply the complement of each other, since participants also looked to the moving Agent and other elements of background in events. We chose to report cumulative looking times rather than proportion of looks to path or manner (cf. Papafragou et al., 2008) because cumulative time best illustrates the amount of time participants spent studying critical regions as a function of how long they had been viewing each video. In these figures, diverging slopes between conditions indicate different looking preferences across conditions during the corresponding period of time, whereas parallel slopes indicate similar looking preferences during that same time. For instance, consider Figure 3C, which illustrates English speakers' inspection of the Manner region. After 1000 ms of viewing the video, participants in each of the three conditions had looked at the Manner region for an average of only 50 ms. However, one second later, participants in the Linguistic and No Interference conditions had spent on average an additional 200 ms (250 minus 50 ms) whereas participants in the Nonlinguistic Interference condition had spent only 100 ms more time looking at the Manner (150 minus 50 ms). The slope for the Nonlinguistic condition continues to diverge from the other two conditions for an additional second indicating a continued divergence in looking preference. But thereafter the lines remain parallel indicating no additional changes after three seconds in looking preference across conditions.

Figure 3 about here

Informal inspection of Figure 3 reveals a pattern of effects consistent with the hypothesis that speakers adopt an optional linguistic encoding of events but inconsistent with the claim that there are fundamental conceptual biases in event representation based on native language. That is, a reverse-Whorfian effect was observed, but only in the Nonlinguistic Interference task: when cognitive load was high but did not disrupt the ability to use language (Nonlinguistic Interference), speakers of English (a Manner language), inspected the Path Endpoint more (panel A) and the Manner of Motion region less (panel C), as compared to when cognitive load was high but linguistic encoding was prevented (Linguistic Interference), or when cognitive load was low (No Interference). The opposite pattern was observed, albeit less strongly in the Greek speakers. With Nonlinguistic Interference, speakers of Greek (a Path language), inspected the Path Endpoint less (panel B) and the Manner of

Motion region (slightly) more (panel D), as compared to when linguistic encoding was prevented (Linguistic Interference), or when there was no secondary task (No Interference).

To better illustrate these patterns, and to test their reliability, we calculated the Path-over-Manner preference for each condition, which is simply the difference between Path Endpoint and Manner of Motion cumulative looking times (graphed in Figure 4). Positive values indicate a preference to spend more time looking at the Path Endpoint than the Manner of Motion; negative values indicate a preference to spend more time looking at the Manner of Motion than the Path Endpoint.

Figure 4 about here

As can be seen in Figure 4, the two Language groups diverge quite strikingly during the second of the one-second intervals (1.0 to 2.0 seconds). In the Nonlinguistic Interference condition as compared to the other two conditions, English speakers increasingly prefer to inspect the Path Endpoint rather than the Manner of Motion, whereas Greek speakers actually show a decrease in their preference to study the Path Endpoint in this same interval.

We tested the reliability of the patterns in Figure 4 by using multi-level mixed modeling over time, following Barr (2008). Our approach diverges from Barr's in that we used crossed random effects modeling of Subject and Items on non-averaged data rather than modeling Subject and Item means separately (see Locker, Hoffman & Bovaird, 2007, and Baayen, Davidson & Bates, 2008, for discussion). We modeled Path minus Manner gaze times using the following first-level predictors: Interference Task (Linguistic, Nonlinguistic or No Interference), Language (Greek, English), and Time. Data were sampled only from every 500 ms, resulting in 10 time periods (500 ms to 5000 ms). Because of the shapes of the curves in Figure 4, first and second order polynomial time (Linear, Quadratic) were used as predictors (see Mirman, Dixon & Magnuson, 2008). This allows us to capture the curvature of any effects over time. (Following standard procedures, predictor variables were centered before doing the analyses.) Finally, each Subject and Item was given separate random intercepts, and each Item was given a random slope for each factor (Task, Language, Linear Time and Quadratic Time). (Providing random slopes for each Subject gave consistently worse fits of the data and therefore such slopes were dropped from analyses.)

The results from the data modeling appear in Table 1. As can be seen in the table, reliable predictors of gaze preferences all involved interactions between components of Time, Language, and the contrast between the Nonlinguistic (Tapping) Interference task and the No Interference task. First, the contrast between Nonlinguistic and No Interference interacted with Linear time. This reflects the fact that the Nonlinguistic (Tapping) task resulted in a greater preference to look at Path Endpoints.[3] Second, the contrast between Nonlinguistic and No Interference also interacted with Quadratic time, reflecting the fact that looking preference over time in the Nonlinguistic condition has greater curvature (see Figure 4). Crucially however, both of these interactions with time interacted with Language, such that English and Greek speakers show opposite directions of these effects. None of the reliable interactions involved the Linguistic and No Interference contrast, suggesting little difference between these conditions.

Insert Table 1 about here.

Traditional statistical treatments of the data yielded similar results. For instance, following our earlier paper (Papafragou et al., 2008), we also computed non-cumulative Path minus Manner looking times for five one-second intervals, and entered the resulting Subject and Item means into separate

---

[3] See Figure 6 of Mirman et al. (2008) for a discussion of how best to interpret linear and quadratic components of time.

Analyses of Variance (ANOVAs) for each one second interval. Specifically, for each one second interval of a trial, the proportion of time spent looking at the Path and the proportion of Time spent looking at the Manner was calculated. Each proportion was then arcsin transformed and the difference was taken. Subject and Item means of this difference score were entered into separate ANOVAs with Language and Interference task as factors. Consistent with the changes in slopes seen in Figure 4 during the 2nd one-second interval, traditional ANOVAs on non-cumulative looking time during the 2nd one-second interval showed a reliable interaction between Language and Interference ($F1(2,52)=3.97$, $p<0.05$; $F2(2,22)=10.30$, $p<0.001$). No other reliable effects or interactions obtained.

This pattern supports the hypothesis that fundamental conceptual biases between the language groups, if they exist at all, do not explain the reverse-Whorfian effects. Rather, speakers have the option to encode the events linguistically and do so when the task is difficult and the conditions allow it (Nonlinguistic Interference). But are there any conceptual differences of the normal Whorfian sort? To answer this, we re-plotted the data from Figure 4 in Figure 5, except now separately for each task, comparing directly English and Greek speakers within each task.

Figure 5 about here

Figure 5 shows a potentially interesting pattern. In the No Interference condition (panel A) and the Linguistic Interference condition (panel B), we see a trend for Greek speakers to prefer the Path Endpoint more than the English speakers – a (potentially) true Whorfian effect. However, separate multilevel mixed models within each of these tasks reveal that this effect is quite small and tenuous. As shown in Table 2, in the best fitting model within the No Interference condition, Language did not interact with Linear Time. However, Language did interact with Quadratic time ($p<0.05$), reflecting the fact that in Figure 5a, the plot for Greeks has greater curvature than the plot for English. Thus, there appears to be a very small ‘true’ Whorfian effect, which is much less striking than effects of Interference. We note also that traditional Subject and Item ANOVAs on non-cumulative time did not yield a corresponding effect of Language in any of the one second time windows in the No Interference Condition nor the Linguistic Interference Condition, suggesting this effect of Language may be spurious.

Two aspects of our results merit some further discussion. First, it might appear puzzling that no canonical (non-reverse) Whorfian encoding precedes the reverse-Whorfian effect in the eye movement data. If participants, guided by linguistic encoding, fixate non-essential aspects of the scene at some later stage of event encoding, it seems reasonable that they should have first encoded the core aspects of the scene in some way congruent with their native language. If so, one might expect their eye movements early on to reflect language-specific linguistic planning (of the sort seen, e.g., when participants are asked to overtly describe motion events; see Papafragou et al., 2008). Despite the initial plausibility of this idea, we do not believe that linguistic encoding for purposes of memory is necessarily equivalent to a full-blown, structured linguistic string (i.e., an utterance). For instance, order of mention and other considerations that accompany actual speech planning might well be irrelevant for linguistic encoding purposes. Indeed, based on a large amount of experimental evidence, language production researchers have long assumed that producing an actual spoken description of the world involves preparation of three levels of representation: a message level, a functional level and a positional level (see, e.g., Garrett, 1988; Bock & Levelt, 1994; Bock, Irwin, Davidson & Levelt, 2003). It is only at the positional level of representation that things get ‘spelled out’ in an order required for a sentence. The message and function planning does not contain order information, and the mapping from the full conceptual representation of the world onto the message happens quickly and in parallel, with components of the message forming over time in a partially overlapping fashion. We suspect that individuals are planning a message to aid memory up to a functional level, but not a spelled-out sentence, to aid memory. Consequently, one should not expect eye movements to reflect speech planning per se but to correspond partially to the kinds of message and functional information that can be encoded within a language. Indeed, the fact that the canonical Whorfian effects are small and statistically questionable in our data can be seen as consistent with the view that message and functional encoding is dispersed over time and highly variable between subjects.

Second, the present study largely replicates our prior findings in Papafragou et al. (2008). Both studies show that, when asked to inspect and remember events, people's attention does not follow language-specific lines. Moreover, both studies reveal a role for language in memory encoding under certain circumstances (what we have called the 'reverse-Whorfian' effect). There appears to be one difference between the two data sets: in Papafragou et al. (2008), this reverse-Whorfian effect was observed late in the nonlinguistic (memory) manipulation, after participants had the opportunity to watch the event unfold; however, in the present No Interference condition, no reverse-Whorfian strategy was observed at any point. We believe that this difference can be explained by a difference in procedure between our previous work and the present study (the two studies used similar, even though not identical, stimuli). In Papafragou et al. (2008), participants in the beginning of the non-linguistic (memory) session were given examples of possible event changes as practice (e.g., a woman eating an apple turned into a woman eating a pie); also participants were told that the memory phase would involve a single static image taken from the original motion event, which they would have to identify as same or different from the original event. On the basis of the examples, participants knew that changes in the memory task would be quite subtle; furthermore, they knew that the format of the items in the memory phase (static clips) would be different from the format of events in the inspection phase (dynamic events). Both of these task features might have made them treat the task as difficult and spontaneously adopt a linguistic encoding strategy after they first inspected the events. By contrast, our present No Interference task did not include example items, and the format of items in the encoding and memory phase was the same (dynamic events). Thus participants may have been less worried about performance and therefore less likely to adopt additional language-based strategies. When the current study was made more demanding (as in the Nonlinguistic Interference task), participants fell back on the linguistic encoding strategy and, in fact, used it quite early during event apprehension (e.g., for English speakers, it occurred about 800 or 900 ms from video onset – or after about 3 fixations of typical scene viewing). Indeed, the presence of the interference task would serve as a constant reminder that a memory test was coming up at the end of the experiment, further heightening participants' concerns to study the details.

Memory responses

The results of the video recognition test at the end of the experiment also support the conclusion that no conceptual differences exist across the language groups. The proportion of correct responses for Changed Manner, Changed Path Endpoint, and No Change (filler) memory trials appears in Figure 6. As can be seen in the figure, memory performance in the No Interference condition is overall better than either Interference condition (Nonlinguistic or Linguistic). Both Interference tasks decreased accuracy similarly and in ways that did not seem to interact with whether one was a Greek or an English speaker. Overall, Greek speakers did somewhat worse on the memory task, but this did not appear to interact with the type of Change (Manner or Path). And quite remarkably, it appears that probable linguistic encoding strategies adopted during the Nonlinguistic Interference condition had little effect on participants' ability to recognize changes.

Figure 6 about here

To test the reliability of these patterns, binary accuracy data from each test trial were submitted to a mixed logit model with the following fixed factors: Language (Greek, English); Interference (No

Interference, Linguistic Interference, Nonlinguistic Interference) and Type of Change (Path, Manner). In level two, random intercepts were provided for each Subject and each Item (i.e., crossed random Subject and Item intercepts). Consistent with the description above, this model revealed a significant effect of Interference (Wald-Z=2.11; $p<0.05$) and Type of Change (Wald-Z=-2.36; $p<0.05$). No other effects or interactions were significant (all p's>0.4).

Experiment 2

Experiment 2 further explored the conditions under which the linguistic encoding strategy is used for memory purposes. This experiment was a modified version of the Linguistic Interference task of Experiment 1, in which participants only heard and were required to repeat a sequence of numbers *after* they had freely inspected the event for 3s (i.e., after the beep was heard and the last frame of the event froze on the screen, at which time the numbers were heard and then repeated back). Thus participants had the opportunity to encode the event non-linguistically and/or linguistically in the first (interference-free) phase but were later prevented from using a linguistic encoding strategy to maintain the event details in memory. This experiment leaves open three potential outcomes: (a) participants could inspect events without encoding them linguistically at any point (as in the No Interference condition of Experiment 1); (b) participants could initially encode events freely (without resource to language), and attempt to use a reverse-Whorfian strategy only towards the end of the three-second-long inspection phase in anticipation of the interference task; (c) finally, participants could employ the reverse-Whorfian strategy early on in the Encoding Phase, while the linguistic code was still available (just as they did in the Nonlinguistic Interference task of Experiment 1). Which of these outcomes should obtain depends on how difficult the Delayed Interference task would be considered by participants, with (a) entailing a judgment of least difficulty and (c) of highest difficulty. Based on our earlier results, we expected the new Delayed Interference task to be (judged as) easier compared to the earlier Linguistic Interference task but harder than a simple No Interference task – hence we expected (b) to be the most likely outcome (or, alternatively, a version of (c) in which linguistic encoding happens throughout the trial but the effect is weaker than the effect observed in the earlier, more demanding Linguistic Interference condition). Such an outcome would suggest that deploying the use of linguistic representations is under the control of participants (i.e., it is optional) and is based on perception of task difficulty.

A secondary goal of Experiment 2 was to address a concern about the memory task of Experiment 1. As Figure 6 shows, in the Interference conditions (both Linguistic and Nonlinguistic) of the previous experiment, performance in the recognition task hovered only slightly above 50%. One possible concern is that performance might have been so poor in these conditions that no cross-linguistic differences in memory could have been observed. The Delayed Linguistic Interference task included in Experiment 2 was expected to be easier than the Linguistic Interference task of Experiment 1 (since it introduced interference only after the inspection phase) and was therefore expected to lead to higher success rates on the memory task. Of interest was whether such higher success rates would reveal cross-linguistic differences in Manner vs. Path Endpoint representation.

Method

Participants

Ten Greek speakers and 10 English speakers were recruited from the same populations as in Experiment 1. None of them had completed the previous experiment.

Stimuli and Procedure

Stimuli and Procedure were as in the Linguistic Interference task of Experiment 1 with the following modification: only after motion event (at the beep) were participants given the set of three numbers and hence cued to start counting (i.e., linguistic interference occurred only during the two-second still frame). As in Experiment 1, all participants were warned about this secondary task in the beginning of the experiment. This Delayed Linguistic Interference task allowed participants to freely encode the motion events during the first 3s of the display.

Results and Discussion

Eye movements

Figure 7 presents the eye movement results for English and Greek speakers. Plotted is the average accumulation of time spent looking at the Path Endpoint region (panel A), the average accumulation of time spent looking at the Manner of Motion region (panels B) and the difference between these two measures (panel C). As can be seen in the figure, differences between Greek and English speakers emerge just before second 3, when they are about to hear the digits and repeat them aloud. The effect is the reverse-Whorfian effect: Greek speakers start studying Manner of Motion more than English speakers whereas English speakers start studying Path Endpoints more than Greek speakers. This change in looking preferences (as illustrated by changes in slopes of the two lines in each panel) is localized to 2.5 – 4.0 seconds, just before and during the hearing of the digits. Such a pattern, although surprising, suggests that in preparation for the linguistic interference task, participants have encoded the events linguistically and made a decision about what details to focus on during the interference task.

Figure 7 about here

The results from the data modeling appear in Table 3, and confirm our description of the changes in means. As can be seen in the table, the only reliable predictor of gaze preferences was the interaction between the Language of the participant and Linear Time. This interaction reflects the fact that, as can be seen in Figure 7, preference to look at the Manner of Motion rather than the Path Endpoint becomes substantially greater at 3 to 4 seconds for Greek as compared to English speakers.

Insert Table 3 about here.

Traditional statistical treatments of the data yielded similar results. As in Experiment 1, non-cumulative Path minus Manner looking times were entered into separate Subject-mean and Item-mean ANOVAs for each of five one-second intervals. Specifically, for each one second interval of a trial, the proportion of time spent looking at the Path and the proportion of Time spent looking at the Manner was calculated. Each proportion was then arcsin transformed and the difference was taken. Subject and Item means of this difference score were entered into separate ANOVAs with Language and Interference task as factors. Consistent with the changes in slopes seen in Figure 7c during the 4th one-second interval, traditional ANOVAs on non-cumulative looking time during the 4th one-second interval showed a reliable effect of Language and Interference ($F1(1,18)=4.81$, $p<0.05$; $F2(1,11)=5.78$, $p<0.05$) such that Greek speakers showed a stronger preference to look at the Manner of Motion region than did English speakers.

Memory responses

As shown in Figure 8, recognition memory did improve in this Delayed Linguistic Interference task compared to the Interference conditions of Experiment 1 (an average around 70 percent correct, as compared to less than 60 percent in the previous experiment). Even with this improved performance, effects or interactions with native language were not observed. Specifically, based on a mixed logit model of accuracy in this condition, only a main effect of Type of Change was observed (Wald-Z=-2.20; $p<0.05$), with no effect of Language ($p=0.13$) and no interaction between Type of Change and Language ($p=0.64$). Thus, it seems unlikely that low performance in Experiment 1 masked potential language-driven memory differences.

Figure 8 about here

General Discussion

Implications and conclusions

Our findings lead to several major conclusions about the relationship between event perception, memory and language. First, basic processes underlying event perception and memory are independent from one's native language: speakers of different languages do not differ in terms of attention allocation when they inspect unfolding motion under viewing conditions that do not involve cognitive load (No Interference; first phase of Delayed Linguistic Interference) and under conditions that block linguistic encoding (Linguistic Interference). Furthermore, they perform comparably in tasks that test their memory of event components that are encoded differently in the languages they speak. These results extend and support prior evidence for the independence of motion cognition from linguistic-encoding patterns (Papafragou et al., 2008; Papafragou et al., 2002, 2006; Gennari et al., 2002; cf. also Malt, Gennari, Imai, Ameel, Tsuda & Majid, 2008).

Second, our findings point to a role for language as a tool selectively used to support memory under conditions in which (a) committing something to memory is difficult (e.g., because of high cognitive load), and (b) language is accessible as an encoding medium (as in our Nonlinguistic

Interference and our Delayed Linguistic Interference tasks).[4] This (implicit) use of the verbal code gives rise to cross-linguistic differences in attention allocation as events are apprehended and committed to memory. In our data, the most pronounced effect of this linguistic encoding (as measured by eye movements) is to direct people's attention to the 'details' (rather than the 'core') of a visual event soon after the event has begun - the point presumably being that event details might be forgotten and thus need to be bolstered through the linguistic code. Since languages differ in what they consider 'details' of the event (event components not encoded in the main verb and therefore omissible), linguistic encoding leads to what we have called a 'reverse-Whorfian' effect: partway through the inspection of motion, participants' attention strongly diverges in directions which are *incongruent* with the information preferentially encoded by the main verb in each language (manner in English, path in Greek). This finding offers a novel perspective on previous observations according to which scene details can become the focus of attention (Antes, 1974; Loftus, 1972; Antes & Kristjanson, 1993).

Third, the effects of language as an encoding tool are neither permanent nor pervasive. Support for this conclusion comes from the fact that, as pointed out already, effects of language encoding on attention do not emerge under many viewing conditions (i.e., in our No Interference condition of Experiment 1, or the early phases of the Delayed Interference task of Experiment 2) and appear to be related to how difficult the task is perceived to be; furthermore, language-specific attention effects driven by encoding disappear under Linguistic Interference (Experiment 1). This conclusion is compatible with independent evidence that language is not immediately and automatically implicated in memory retention. For instance, studies have shown that overt naming of pictures during study does not improve recognition memory (Bahrick & Boucher, 1968), and that the naming latency of pictures does not have an effect on either picture recognition or recall (Intraub, 1979; *contra* Paivio, 1971).

---

[4] This type of effect may also emerge in cases in which participants implicitly use words as a strategy for interpreting ambiguous experimental instructions (Li, Abarbanell & Papafragou, 2005; Pinker, 1994). In that case, too, effects of language are task-specific and can be disrupted.

Rather than language encouraging (or forcing) a specialization of the mind by permanently altering the relative salience of manners and paths of motion over time in the minds of English and Greek speakers, language can be seen as an additional route for encoding motion information in case linguistic encoding is appropriate and available.

Finally, the linguistic encoding strategy has limited efficacy: in our data, it does not predict how accurate speakers of different languages will be in remembering events. The most likely explanation for this is that implicit linguistic encoding (and the accompanying attention effects) cannot override the perceptual encoding of the events which is shared by both English and Greek participants. Prior studies have failed to find an effect of language on memory for path and manner of motion even in circumstances in which people were asked to describe each of the target events prior to the memory task – and thus performed linguistic encoding overtly (Gennari et al., 2002; Papafragou et al., 2002). It remains an open possibility, of course, that in cases in which language offers a more efficient way of packaging a complex event (e.g., through a single predicate), linguistic encoding strategies might be more likely to affect memory performance (see Papafragou et al., in prep., for suggestive data).

Comparison to other research findings

It is important to note that our findings show striking convergence with two other recent studies that ask whether native language influences human performance on perceptual-cognitive tasks in the presence of a secondary interference task (Winawer et al., 2007; Gilbert et al., 2008). In both of these studies, effects of language on a perceptual-cognitive task were eliminated in the presence of a linguistic interference task but not a nonlinguistic interference task. For instance, Winawer et al. (2007) found that Russian speakers but not English speakers were speeded on a color-matching task when the color comparisons were made across two Russian color categories (light blue, *goluboy,* and dark blue, *siniy*) as compared to color comparisons made within either of these Russian color categories. These cross-linguistic differences went away when a secondary linguistic task was performed by participants (both groups failed to show the Russian-defined language category effect). The cross-linguistic differences returned, however, under a nonlinguistic interference task.

These researchers (Gilbert et al., 2008; Winawer et al., 2007) conclude, like us, that language is one of several ways that a perceptual stimulus can be encoded. However, both groups of researchers make what we believe is a logical error by additionally concluding from these results that use of linguistic codes is pervasive in perceptual and cognitive tasks and shapes nonlinguistic perceptual categories. For instance, Winawer et al. (2007) conclude: "The critical difference in this case is not that English speakers cannot distinguish between light and dark blues, but rather that Russian speakers cannot avoid distinguishing them: they must do so to speak Russian in a conventional manner. This communicative requirement appears to cause Russian speakers to habitually make use of this distinction even when performing a perceptual task that does not require language."

Such a conclusion does not follow, however. Even if it were the case that Russian speakers automatically compute linguistic labels in the presence of colors, it does not mean that they "habitually make use" of them when performing perceptual tasks. The use of different sources of information in categorization tasks is known to be notoriously context dependent (e.g., Armstrong, Gleitman & Gleitman, 1983; Labov, 1973). For instance, consider the categorization task known as lexical decision: the decision to use orthographic, phonological and/or semantic information to decide if a string of letters is a legal word of a language depends on the properties of those trials for which people respond 'no': If all nonwords are illegal letter strings (XGFZ), participants need only use (and do indeed only use) orthographic cues to correctly accept real words. If nonwords frequently include pseudo-homophones, which are nonwords that are pronounceable and sound like real words (BRANE), then participants must rely on word meaning to make the decision to correctly accept real words (e.g., Antos, 1979; Kellas, Ferraro, & Simpson, 1988; Stone & Van Orden, 1993). These studies suggest that there is a real difference between what is automatically computed during the perception of a word and what is used in making a categorization. In relation to this, it is important to note that the stimuli used in the Winawer et al. (2007) study were all shades of blue; no other colors were used. Thus, it is likely that such a task would make Russian speakers (but not English speakers) become aware of a lexical contrast in the stimuli, which could be used to aid some decisions. (Likewise, only cats and dogs were used as stimuli for categorization in the Gilbert et al., 2008, study.) The implication here is that the inclusion of a wider range of stimuli (more than just blues) would make such an effect go away because the use of linguistic terms would no longer be perceived as providing a possible performance advantage.

Moreover, the proposal that the mere perception of color automatically triggers linguistic representations becomes quite dubious in light of other classic findings such as the Stroop effect. When naming the color that a word is printed in, color terms interfere with the task (Stroop, 1935), suggesting that strings of letters automatically trigger the computation of linguistic information. However, the opposite is not true; when reading color words aloud, text printed in incongruent colors has no effect on word naming times, suggesting that colors do not automatically trigger the computation of linguistic information (Stroop, 1935; MacLeod, 1991). If computing linguistic labels had become a habitual part of perceiving color, as Winawer et al. (2007) conclude, we should see interference effects happening in both directions. The asymmetric nature of the Stroop effect instead suggests that computing the linguistic labels from perceived colors is not automatic and is highly task specific.[5] Indeed, a 'reverse Stroop effect' only appears under very specific conditions consistent with this account, such as when participants have just recently spent an inordinate amount of time practicing

---

[5] The asymmetry in the Stroop effect might reflect a weaker automatic connection between color and linguistic representations. However, if this were so, it becomes difficult to explain why such a weak connection has any effect in Winawer et al. (2007). Reverse Stroop is so conspicuously absent in the 75-year literature that it indicates that the mere perceiving of colors has no automatic impact on accessing linguistic labels. There are of course 'connections' between color sensation and the representation of color categories in the brain, but these connections need to be engaged by controlled processes (executive attention), not the mere presence of color.

the verbal labeling of color patches (an effect that goes away quickly, e.g., Stroop, 1935) or when the response to a word is a visual search task that encourages the use of the color of the print (e.g., Durgin, 2000; see also Blais & Besner, 2007, for other specific task conditions related to providing participants with limited response sets).

Relation between language and thought

Taken together, our data suggest a nuanced interplay of linguistic and cognitive factors in perceptual processing. Specifically, they point to a mechanism through which language is involved in 'nonlinguistic' tasks - but in a way far less radical than proposed by strong proposals about the relationship between linguistic encoding and perception (e.g., Winawer et al., 2007; Gilbert et al., 2008). Unlike what such accounts propose, languages do not alter the underlying perceptual or conceptual event representations of speakers but offer an alternative route for the efficient and quick packaging of event information. As we show here, this mechanism has implications for the on-line allocation of attention to event components and can lead to cross-linguistic differences in event inspection in certain (but not all) contexts. Our data are compatible with research showing that verbal labels can be used to enhance working memory capacity in adults (Baddeley, 1976, 1986, 2003; Crowder, 1976), even though not in young children (Hitch, Halliday, Schaafstal & Heffernan, 1991; Palmer, 2000). The circumstances under which language is spontaneously recruited in nonlinguistic tasks remain to be clarified by further research. Language is useful for some tasks but not others (Potter, 1979; Papafragou & Selimis, 2010); furthermore, this account predicts that ways of increasing task difficulty other than through cognitive load would also encourage the use of linguistic symbolic representations, provided that the linguistic system is not otherwise preoccupied. In this way, this account is quite compatible with Slobin's (2003, 2006) 'thinking for speaking' account, in which individuals sometimes anticipate the need for linguistic encoding of perceived events. However, given the instability of the cross-linguistic differences when there was no interference task, the findings do not suggest, as Slobin (2006) claims, that linguistic encoding is so pervasive that it affects attention allocation generally even when linguistic encoding is not occurring.

This position also coheres with several recent findings in the domains of space and number that suggest a role for language in fast and efficient information processing. In one such demonstration, Dessalegn and Landau (2008) have shown that language can be useful as a temporary medium for visual feature binding in the representation and retention in memory of static spatial stimuli. Other work has shown that number words in natural language function as a cognitive technology that allows representations of exact cardinalities of sets to be remembered and communicated accurately across contexts (Frank, Everett, Fedorenko & Gibson, 2008). Speakers of a language that lacks number words can still perform numerical computations involving the equivalence of large sets but they have trouble retaining this information in memory (Frank et al., 2008). Finally, in the domain of spatial orientation, one's native language can be used in efficiently compressing and retaining navigational information; such effects, however, disappear under linguistic interference (Hermer-Vasquez, Spelke & Katsnelson, 1999; cf. Ratcliff & Newcombe, 2008). In all these domains, linguistic effects are evanescent and temporary and do not percolate to the underlying cognitive or perceptual mechanisms. Together with our data, these findings support the view that, rather than permanently reshaping the processes supporting event perception and processing, language offers an alternative, optionally recruited system of encoding, organizing and tracking experience.

References

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T. & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese, and Turkish. *Cognition*, *102*, 16-48.

Allport, D. A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single-channel hypothesis. *Quarterly Journal of Experimental Psychology, 24,* 225–235.

Antes, J.R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103(1),* 62-70.

Antes, J.R. & Kristjanson, A. (1993). Effects of capacity demands on picture viewing. *Perception and Psychophysics, 54,* 808-813.

Antes, J.R. (1975). The time course of picture viewing. *Journal of Experimental Psychology, 103(1),* 62-70.

Antos, S.J. (1979). Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception & Performance. 5(3),* 527-545.

Armstrong, S.L., Gleitman, L.R. & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13(3),* 263-308.

Baayen, R.H. Davidson, D.J. & Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59(4),* 390-412.

Baddeley, A. (1976). *The Psychology of Human Memory*. New York: Basic Books.

Baddeley, A. (1986). *Working Memory.* Oxford: Clarendon Press.

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders, 36,* 189-208.

Bahrick, H.P., & Boucher, B. (1968). Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology, 78,* 417-422.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59,* 457-474.

Billman, D., & Krych, M. (1998). Path and manner verbs in action: effects of "Skipping" or "Exiting" on event memory. *Proceedings of the 20th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Billman, D., Swilley, A., & Krych, M. (2000). Path and manner priming: verb production and event recognition. *Proceedings from the 22nd Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Blais, C. & Besner, D. (2007). A reverse Stroop effect without translation or reading difficulty. *Psychonomic Bulletin & Review, 14 (3),* 466-469.

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. & Levelt, W.J.M. (1994). Language Production: Grammatical Encoding. In M.A. Gernsbacher (Ed.) *Handbook of Psycholinguistics* (pp. 945-984). Sand Diego: Academic Press.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conception of time. *Cognitive Psychology, 43,* 1-22.

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

Bowerman, M., & Levinson, S. eds. (2001). *Language Acquisition and Conceptual Development.* Cambridge: Cambridge University Press.

Crowder, R. G. (1976). The locus of the lexicality effect in short-term memory for phonologically identical lists. *Bulletin of the Psychonomic Society, 7(4),* 361-363.

Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science, 19 (2),* 189-195.

Durgin, F. H. (2000). The reverse Stroop effect. Psychonomic Bulletin & Review, 7, 121-125.

Frank, M.C., Everett, D.L., Fedorenko, E., and Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition, 108,* 819-824.

Garret, M.F. (1988). Processes in language production. Language: Psychological and biological aspects. In F..J. Newmeyer, (Ed) *Linguistics: The Cambridge Survey, Vol. 3. (*pp. 69-96). New York: Cambridge University Press.

Gennari, S., Sloman, S., Malt, B., & Fitch, T. (2002). Motion events in language and cognition. *Cognition, 83,* 49-79.

Gentner, D. & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge: Cambridge University Press.

Gentner, D., & Goldin-Meadow, S. eds. (2003). *Language and mind.* Cambridge, MA: MIT Press.

Gilbert, A.L., Regier T., Kay P., Ivry, R.B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language, 105(2),* 91-98.

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

Gumperz, J. & Levinson, S., eds. (1996). *Rethinking linguistic relativity*. New York: Cambridge University Press.

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

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16,* 219-222.

Hermer-Vasquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology, 39*, 3-36.

Hitch, G. J., Halliday, M. S., Schaafstal, A.M., & Heffernan, T. M. (1991). Speech, 'inner speech', and the development of short-term memory: Effects of picture-labeling on recall. *Journal of Experimental Child Psychology, 51,* 220-234.

Intraub, H. (1979). The role of implicit naming in pictorial encoding. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 78-87.

Kellas, G., Ferraro, E R., & Simpson, G. B. (1988). Lexical ambiguity and the time-course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 14,* 601-609.

Labov, W. (1973). The boundaries of words and their meanings. In C.J.N. Bailey & R.W. Shuy (Eds.), *New ways of analyzing variations in English.* Washington, DC: Georgetown University Press.

Lakusta, L. & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition, 96,* 1-33.

Lakusta, L., Wagner, L., O'Hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language Learning and Development, 3(3),* 179-197.

Levelt, W. (1989). *Speaking.* Cambridge, MA: MIT Press.

Levinson, S. C. (2003). *Space in language and cognition*. Cambridge: Cambridge University Press.

Li, P., Abarbanell, L. & Papafragou, A. (2005). Spatial reasoning skills in Tenejapan Mayans. *Proceedings from the Annual Meeting of the Cognitive Science Society* 27. Hillsdale, NJ: Erlbaum.

Li, P. & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83,* 265-294.

Locker L Jr, Hoffman L, Bovaird JA. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavioral Research Methods, 39(4),* 723-730.

Loftus, G. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology, 3,* 525-551.

Malt, B.C., Gennari, S., Imai, M., Ameel, E., Tsuda, N., & Majid, A. (2008). Talking About Walking: Biomechanics and the Language of Locomotion. *Psychological Science, 19(3),* 232-240.

MacLeod, C. M. (1991). John Ridley Stroop: Creator of a landmark cognitive task. *Canadian Psychology, 32,* 521-524.

Mennie, N., Hayhoe, M., Sullivan, B. (2007) Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research, 179,* 427-442.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mirman, D., Dixon, J.A., and Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59(4),* 475-494.

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

Naigles, L. R. & Terrazas, P. (1998). Motion-verb generalizations in English and Spanish: Influences of language and syntax. *Psychological Science*, *9,* 363-369.

Özçalışkan, Ş., & Slobin, D.I. (1999). Learning how to search for the frog: Expression of Manner of motion in English, Spanish, and Turkish. In A. Greenhill, H. Littlefield & C. Tano (eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 541-552). Somerville, MA: Cascadilla Press.

Özçalışkan, Ş., & Slobin, D.I. (2003). Codability effects on the expression of Manner of motion in Turkish and English. In A. S. Özsoy, D. Akar, M. Nakipoğlu-Demiralp, E. Erguvanlı-Taylan &

A. Aksu-Koç (eds.) *Studies in Turkish linguistics* (pp. 259-270). Istanbul: Boğaziçi University Press.

Paivio, A. (1971). *Imagery and Verbal Processes*. Holt, Rinehart & Winston.

Palmer, S. (2000). Working memory: A developmental study of phonological encoding. *Memory*, *8,* 179-193.

Papafragou, A., Hulbert, J. & Trueswell, J.C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108(1),* 155-184.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in thought and language. *Cognition, 84,* 189-219.

Papafragou, A., Massey, C., & Gleitman, L.R. (2006). When English proposes what Greek presupposes: The linguistic encoding of motion events. *Cognition, 98,* B75-87.

Papafragou, A., Mojaverian, T., & Trueswell, J.C. (in prep.). Cross-linguistic effects on language production.

Papafragou, A., & Selimis, S. (2010). Event categorization and language: A cross-linguistic study of motion. *Language and Cognitive Processes, 25,* 224-260.

Papafragou, A., & Selimis, S. (in press). Lexical and structural biases in the acquisition of motion verbs cross-linguistically. *Language Learning and Development.*

Pelz, J.B. & Canosa, R. (2001). Oculomotor Behavior and Perceptual Strategies in Complex Tasks. *Vision Research. 41,* 3587-3596. Penney, C.G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin, 82,* 68-84.

Pinker, S. (1994). *The language instinct.* New York: Morrow.

Potter, M. C. (1979). Mundane symbolism: The relations among objects, names, and ideas. In N.R. Smith and M.B. Franklin (eds.), *Symbolic functioning in childhood* (pp.41-65). Hillsdale, NJ: Erlbaum.

Ratcliff, K.R., & Newcombe, N.S. (2008). Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cognitive Psychology, 56,* 142-163.

Roberson, D., Davidoff, J. & Braisby, N. (1999) Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71,* 1-42.

Rowe, E.J. & Rogers, T.B. (1975). Effects of concurrent auditory shadowing on free recall and recognition of pictures and word. *Journal of Experimental Psychology: Human Learning and Memory*, *1,* 415-422.

Slobin, D. (1996). From 'thought and language' to 'thinking for speaking'. In J. Gumperz and S. Levinson (eds.), *Rethinking linguistic relativity* (pp. 70-96). New York: Cambridge University Press.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (eds.), *Language in mind: Advances in the study of language and thought* (pp. 157-191). Cambridge, MA: MIT Press.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 219-257). Mahwah, NJ: Erlbaum.

Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (pp. 59-81). Amsterdam/Philadelphia: John Benjamins.

Steiger, J.H., Shapiro, A., & Brown, M.W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50(3),* 253-264.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744-774.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology 18,* 643-662.

Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and semantics*, Vol. IV (pp. 181-238). New York: Academic Press.

Talmy L., (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (ed.), *Grammatical Categories and the Lexicon,* Vol. III (pp. 57-149). Cambridge: Cambridge University Press.

Talmy, L. (1991). Path to realization: A typology of event conflation. *Proceedings of the Berkeley Linguistics Society, 17,* 480-519.

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf.* Edited by J. B. Carroll. Cambridge, MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Science, 104 (19),* 7780-7785.

Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research, 47,* 955-964.

Yarbus, A.L. (1967). Eye movements during perception of complex objects. In L.A. Riggs (Ed.), *Eye Movements and Vision.* New York: Plenum Press.

Author Note

This work was partially funded by grant BCS#0641105 from the National Science Foundation to A.P. and J.T. We want to thank Anna Piata, Anthi Haida, Kate McEldoon, U'nek Brown, Ann Bunger and especially Taraneh Mojaverian for their help in conducting the studies and coding the data, and Alon Hafri for commenting on manuscript drafts. We also gratefully acknowledge the hospitality of the faculty of the Department of Philology (Linguistics sector) at the University of Athens, especially Amalia Mozer, Antonis Botinis, Dimitra Hila-Markopoulou, Keti Bakakou-Orfanou and Christoforos Charalambakis, during the time the Greek portion of the data was collected.

Tables

Table 1. Fixed effects from best fitting multi-level linear model of cumulative looking time preference, Path minus Manner (Experiment 1).

Effect	Estimate	S.E.	t-value
Intercept	7.66	144.40	0.05
Task (Linguistic Interference vs. No)	13.60	114.31	0.12
Task (Nonlinguistic Interference vs. No)	166.48	104.45	1.59
Language (Greek vs. English)	8.78	108.50	0.08
Linear Time	-247.96	219.25	-1.13
Quadratic Time	-40.20	56.43	-0.71
Task(Ling. Int. vs. No) x Language	17.23	135.31	0.13
Task(Nonling. Int. vs. No) x Language	-198.87	135.21	-1.47
Task(Ling. Int. vs. No) x Linear Time	67.33	72.02	0.94
Task(Nonling. Int. vs. No) x Linear Time	***229.02***	***71.40***	***3.21****
Language x Linear Time	-74.75	72.26	-1.03
Task(Ling. Int. vs. No) x Quadratic Time	-34.77	72.01	-0.48
Task(Nonling. Int. vs. No) x Quadratic Time	***-150.38***	***71.39***	***-2.11****
Language x Quadratic Time	-125.04	72.15	-1.73
Task(Ling. Int. vs. No) x Language x Linear Time	7.69	104.23	0.07
Task(Nonling. Int. vs. No) x Language x Linear Time	***-223.38***	***103.11***	***-2.17****
Task(Ling. Int. vs. No) x Language x Quadratic Time	64.38	104.14	0.62
Task(Nonling. Int. vs. No) x Language x Quadratic Time	***209.95***	***103.03***	***2.04****

* $p<0.05$ (on normal distribution)

Note: Linear Time = first order polynomial time; Quadratic Time = second order polynomial time. Models presented here and in Tables 2 and 3 all have significantly better fits than empty models with no fixed effects, based on a chi-square test of the change in -2 restricted log likelihood (Steiger, Shapiro & Browne, 1985). All models reporting interactions were significantly better fits than the corresponding models that did not have these interaction terms. When effects or interactions do not appear, it is because adding them to the models did not reliably improve the fit. Formula in R: DepVar ~ Condition * Language * LinTime + Condition * Language *QuadTime + (1 | Subject) + (1 + Condition + Language + LinTime + QuadTime | Item)

Table 2

Table 2. Fixed effects from best fitting multi-level linear models of cumulative looking time preference, Path minus Manner (separate models for each type of interference, Experiment 1).

Interference	Effect	Estimate	S.E.	t-value
No Interference	Intercept	7.87	134.74	0.06
	Language (Greek vs. English)	11.44	101.02	0.11
	Linear Time	-337.42	220.60	-1.53
	Quadratic Time	-40.65	53.22	-0.76
	Language x Linear Time	-96.79	88.95	-1.09
	Language x Quadratic Time	***-122.94***	***61.57***	***-2.00****
	Linear Time x Quadratic Time	317.51	161.75	1.96
	Language x Linear Time x Quadratic Time	80.84	231.08	0.35
Ling. Interference	Intercept	21.44	151.15	0.14
	Language (Greek vs. English)	24.72	141.86	0.17
	Linear Time	-208.78	236.04	-0.88
	Quadratic Time	***-104.17***	***46.68***	***-2.23****
Nonling. Interference	Intercept	173.16	141.04	1.23
	Language (Greek vs. English)	-190.67	122.20	-1.56
	Linear Time	-19.99	220.67	-0.09
	Language x Linear Time	**-299.20**	**79.40**	**-3.77***
	Quadratic Time	**-150.33**	**43.53**	**-3.45***

* $p<0.05$ (on normal distribution)

Note: Linear Time = first order polynomial time; Quadratic Time = second order polynomial time. (See additional information in the note for Table 1.) Formulas in R were: No Interf: DepVar ~ 1 + Language * LinTime *QuadTime + (1 | Subject) + (1 + LinTime + Language + QuadTime | Item); Ling. Interf: DepVar ~ 1 + LinTime + Language + QuadTim + (1 | Subject) + (1 + LinTime + Language + QuadTime | Item); NonLing. Interf: DepVar ~ 1 + LinTime * Language + QuadTime+ (1 | Subject) + (1 + LinTime + Language + QuadTime | Item)

Table 3. Fixed effects from best fitting multi-level linear models of cumulative looking time preference, Path minus Manner (Experiment 2, Delayed Linguistic Interference).

Interference	Effect	Estimate	S.E.	t-value
No Interference	Intercept	-9.68	148.22	-0.07
	Linear Time	-216.74	225.69	-0.96
	Quadratic Time	-52.95	55.30	-0.96
	Language (Greek vs. English)	-65.91	109.12	-0.60
	Language x Linear Time	***-163.55***	***61.30***	***-2.67****

* $p<0.05$ (on normal distribution)

Note: Linear Time = first order polynomial time; Quadratic Time = second order polynomial time. (See additional information in the note for Table 1.) Formula in R: DepVar ~ 1 + LinTime * Language + QuadTime + (1 | Subject) + (1 + LinTime + Language + QuadTime | Item)

Figures

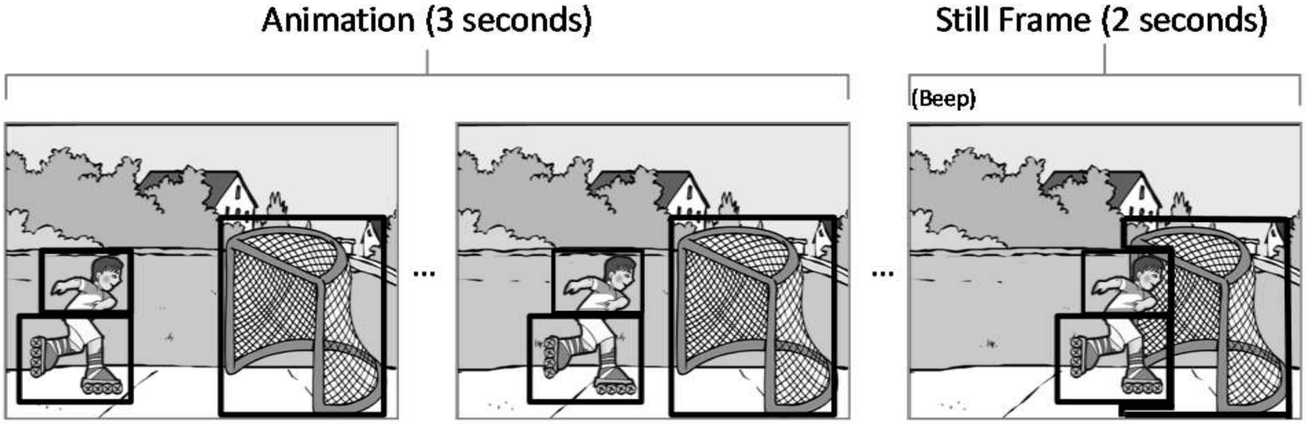


Fig.1. Sample test event. The rectangles illustrate the approximate size and position of eye tracking scoring regions, which moved as the video unfolded. The rectangles were invisible to participant. The three scoring regions correspond to the Agent (head and torso of the boy), the Manner of Motion (the roller skates) and the Path Endpoint (the hockey net). When scoring regions overlapped (as illustrated in the rightmost image), the occluded scoring region (i.e., the Path Endpoint) was modified to obey the rules of occlusion.

Figure 2

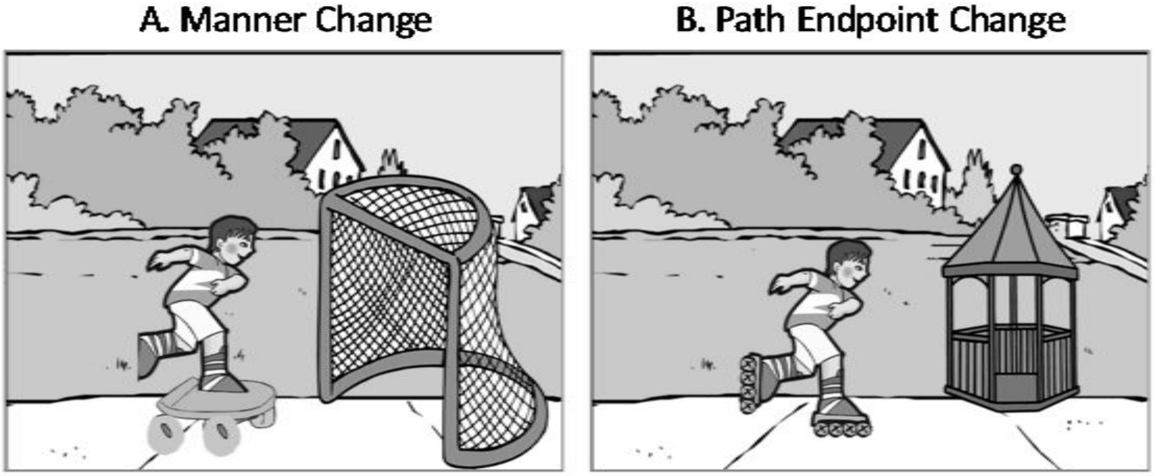


Fig.2. Sample variants of the test event in Fig.1 used in the memory task. Panel A presents a Manner Change and panel B a Path Endpoint Change.

Figure 3

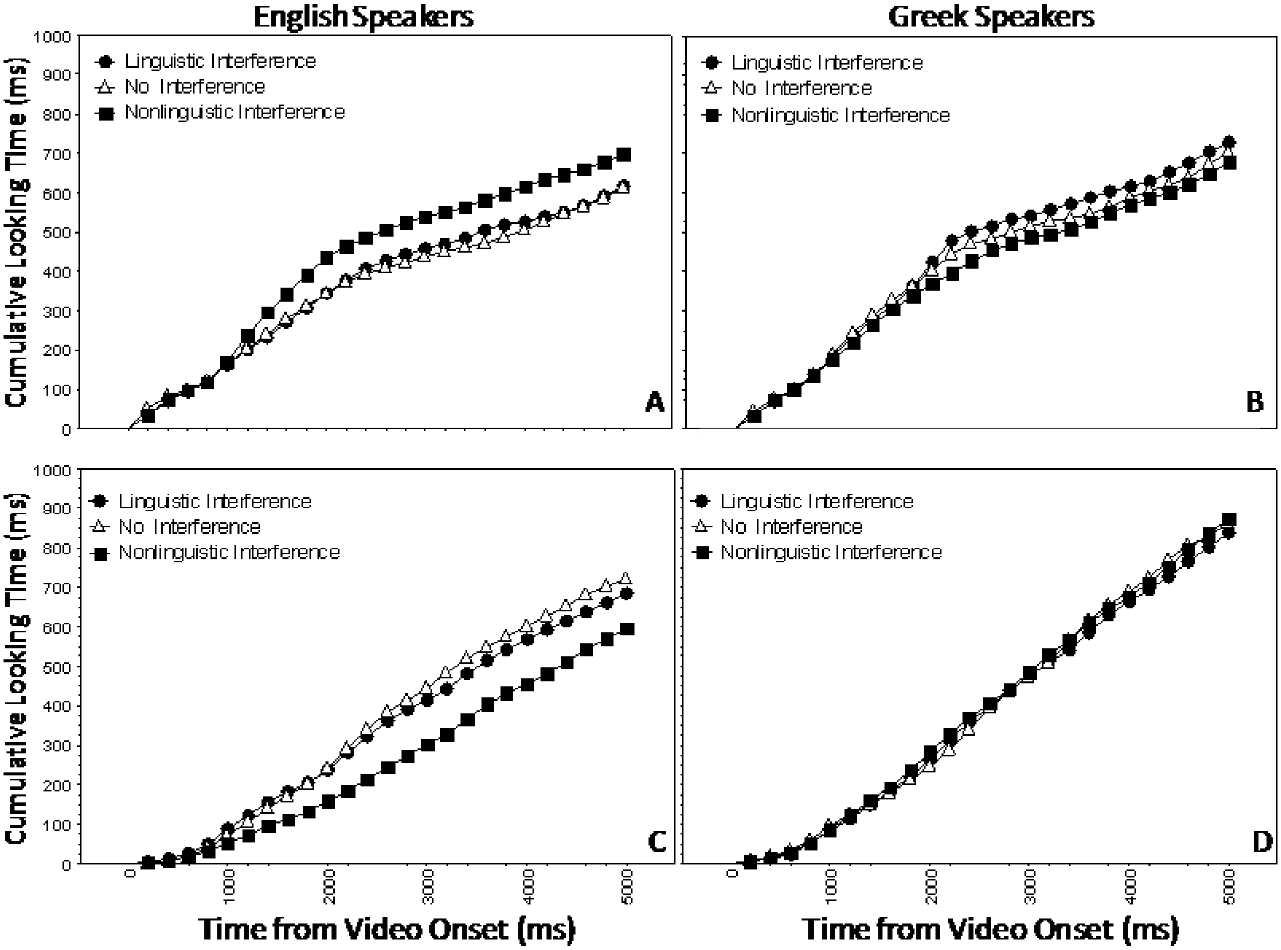


Fig.3. Average accumulation of time that English and Greek speakers spent looking at the Path Endpoint (panels A & B) and Manner of Motion (panels C & D) regions as a function of how long the video had been playing in Experiment 1. (Average of subject means.)

Figure 4

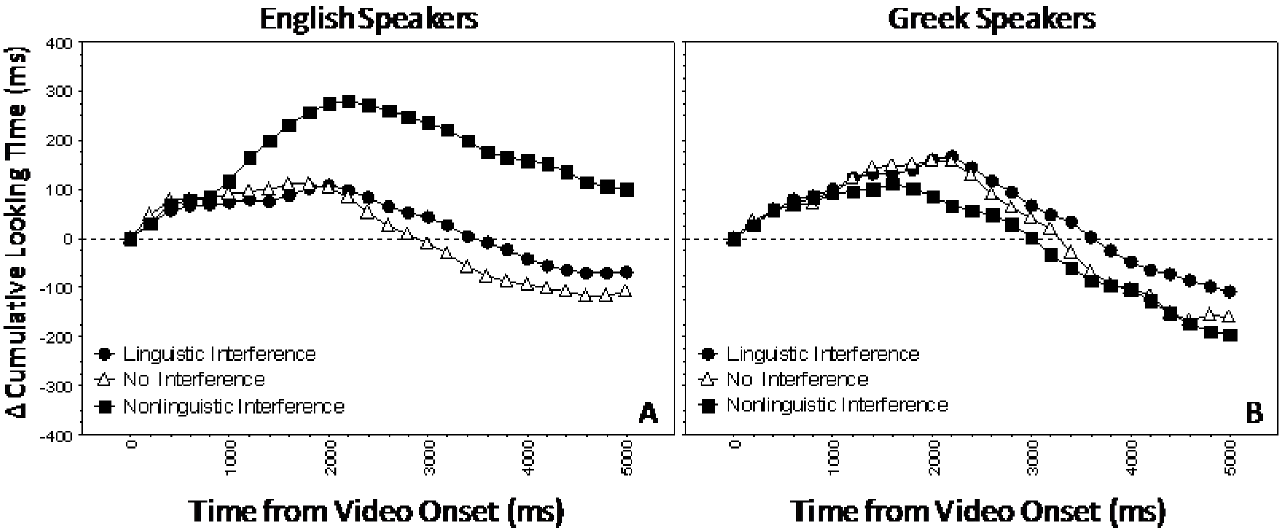


Fig.4. Path-over-Manner preference defined as the difference between Path Endpoint and Manner of Motion cumulative looking times (Experiment 1). Positive values indicate a preference to spend more time looking at the Path Endpoint than the Manner; negative values indicate a preference to spend more time looking at the Manner than the Path Endpoint. (Average of subject means.)

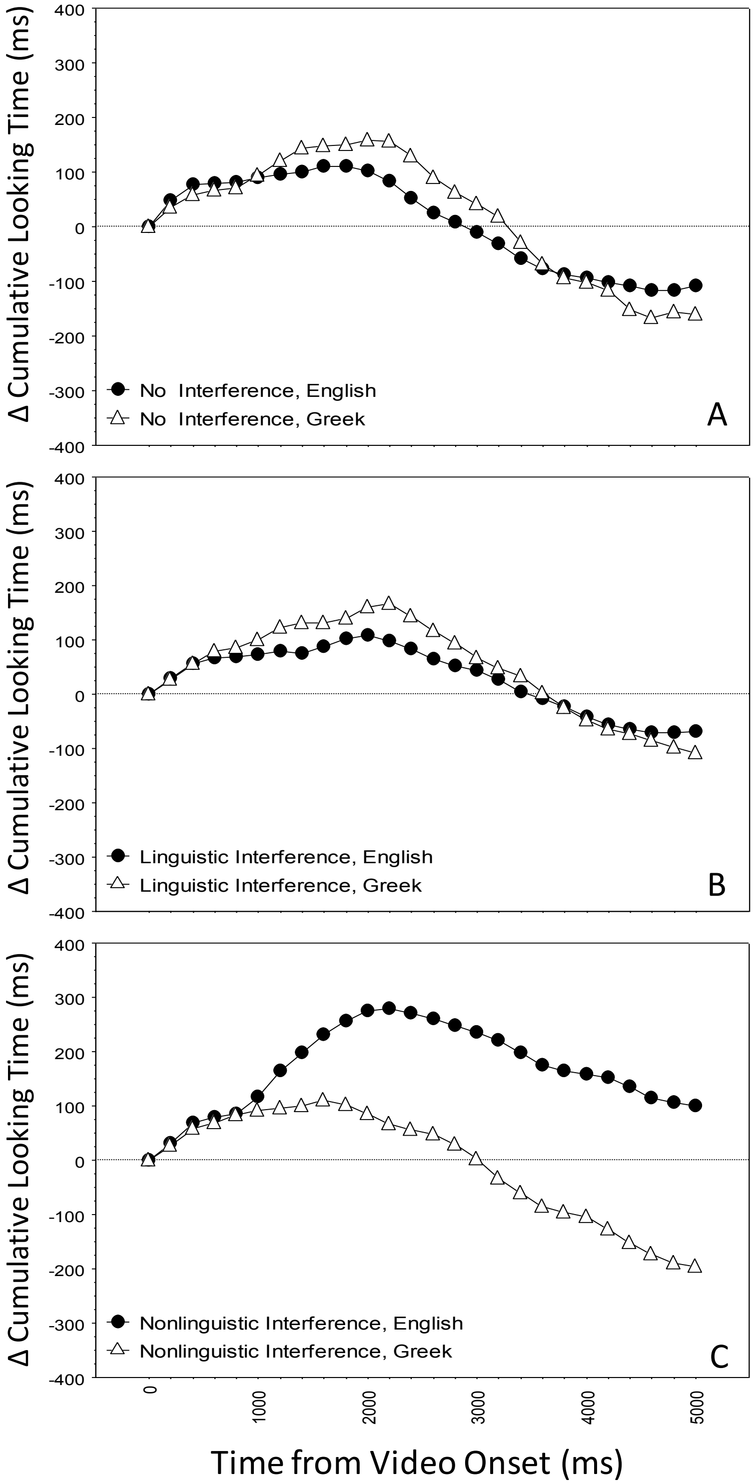
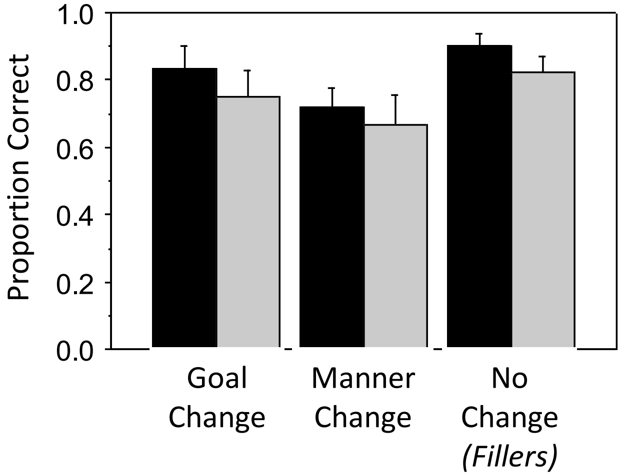


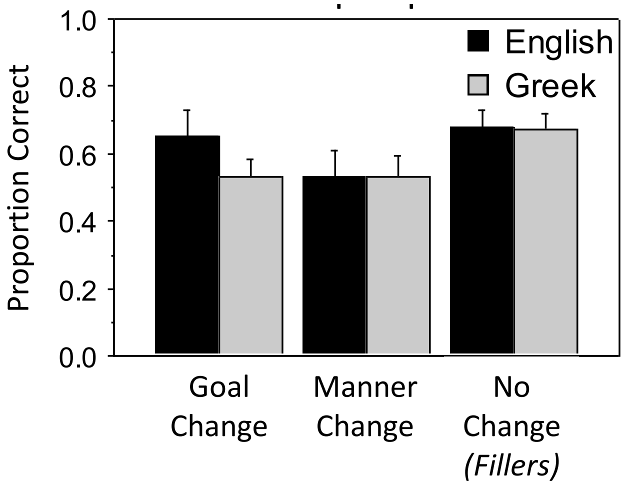
Fig.5. Replotted data from Fig. 4: the difference between Path Endpoint and Manner of Motion cumulative looking times is presented separately for each condition of Experiment 1. (Average of subject means.)

Figure 6

A. No Interference Task



B. Nonlinguistic Interference Task



C. Linguistic Interference Task

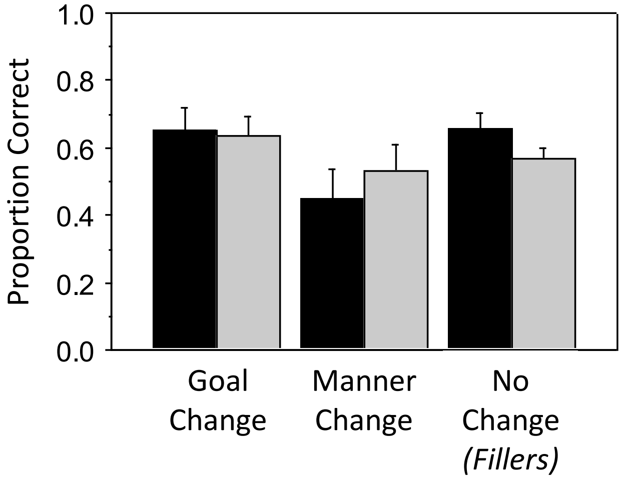


Fig.6. Proportion of correct responses from the recognition memory test in Experiment 1. (Average of subject means. Error bars indicate ±1 Standard Error.)

Figure 7

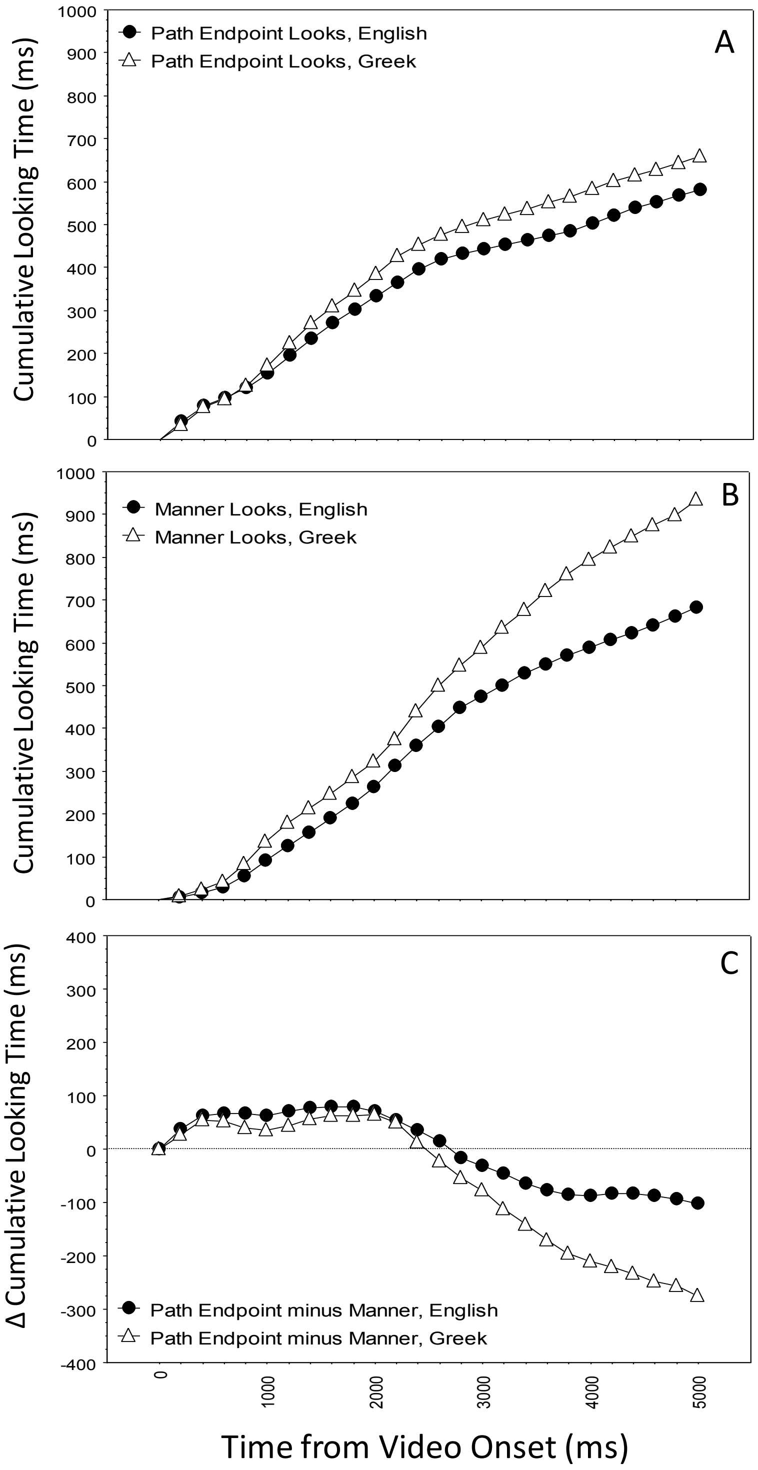


Fig 7. Eye movement results from Delayed Linguistic Interference (Experiment 2). A) Cumulative looking time for Path Endpoint. B) Cumulative Looking time for Manner of Motion. C) Difference between cumulative looking times. (Average of subject means.)

Figure 8

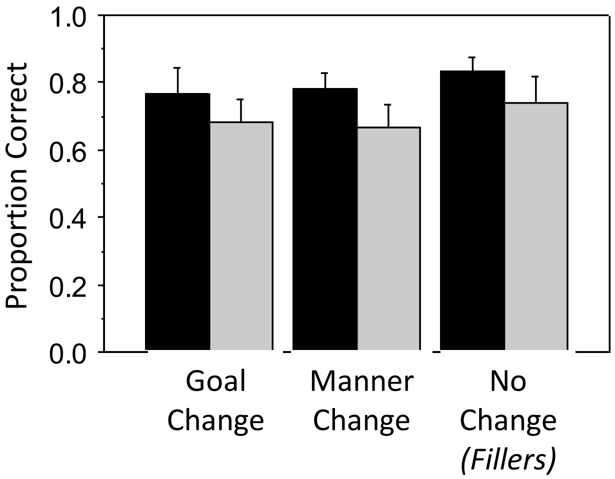


Fig 8. Proportion of correct responses from the recognition memory test in the Delayed Linguistic Interference task of Experiment 2. (Average of subject means. Error bars indicate ±1 Standard Error.)

Appendix. Stimuli list (test events shown during encoding phase).

1. A boy is rollerblading into a hockey net.
2. A man is canoeing to a dock.
3. A man is driving a motorcycle into a garage.
4. A man is landing a plane on a platform.
5. A man is parachuting onto a tree.
6. A man is sailing a boat to a shore.
7. A duck is ice-skating into a house.
8. A girl is riding a scooter into a cave.
9. A man is riding a hot air balloon onto the top of a building.
10. A man is skiing to the finish line.
11. A woman is flying to the moon on a magic carpet.
12. An alien is driving a car into a cave.