



ELSEVIER

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Monitoring sources of event memories: A cross-linguistic investigation



Ercenur Ünal*, Adrienne Pinto, Ann Bunker¹, Anna Papafragou*

Department of Psychological and Brain Sciences, University of Delaware, 105 The Green, Room 108, Newark, DE 19716, USA

ARTICLE INFO

Article history:

Received 25 November 2014
revision received 25 October 2015
Available online 14 November 2015

Keywords:

Source monitoring
Inference
Perception
Event memory
Linguistic relativity
Eventuality

ABSTRACT

When monitoring the origins of their memories, people tend to mistakenly attribute memories generated from internal processes (e.g., imagination, visualization) to perception. Here, we ask whether speaking a language that obligatorily encodes the source of information might help prevent such errors. We compare speakers of English to speakers of Turkish, a language that obligatorily encodes information source (direct/perceptual vs. indirect/hearsay or inference) for past events. In our experiments, participants reported having seen events that they had only inferred from post-event visual evidence. In general, error rates were higher when visual evidence that gave rise to inferences was relatively close to direct visual evidence. Furthermore, errors persisted even when participants were asked to report the specific sources of their memories. Crucially, these error patterns were equivalent across language groups, suggesting that speaking a language that obligatorily encodes source of information does not increase sensitivity to the distinction between perception and inference in event memory.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Source monitoring

People gain knowledge about the world through various experiences. For instance, one might discover that it is raining by looking out of the window and seeing the rain, by hearing the sound of the rain, by hearing from someone that it is raining, or by drawing an inference based on evidence (e.g., a wet umbrella). Such experiences that characterize the conditions under which knowledge is gained are known as the *sources* of that knowledge (Johnson, Hashtroudi, & Lindsay, 1993). Sources of knowledge can be external (e.g., vision or audition) or internal (e.g.,

thoughts, imagination, dreams, or inferences; Johnson & Raye, 1981).

It has long been recognized that people do not readily bind their memories, knowledge or beliefs to the corresponding sources (Johnson, 1997, 2006; Johnson et al., 1993). Rather, they make attributions about the origins of these mental experiences based on subjective characteristics of these experiences through a process known as *source monitoring* (Johnson, 1988). Both the likelihood of attributing memories to particular sources and the accuracy of these attributions are affected by certain characteristics of memories, such as their vividness (Johnson et al., 1993), spatio-temporal contextual details (Johnson, Raye, Foley, & Kim, 1982), the similarity between memories whose origins need to be differentiated (Johnson, Bransford, & Solomon, 1973; Lindsay, Johnson, & Kwon, 1991; Mather, Johnson, & De Leonardis, 1999), and the subject's awareness of the cognitive operations that produced the memory (Durso & Johnson, 1980; Intraub & Hoffman, 1992; Johnson & Raye, 1981).

* Corresponding authors. Fax: +1 (302) 831 3645.

E-mail addresses: eunal@psych.udel.edu (E. Ünal), apapafragou@psych.udel.edu (A. Papafragou).

¹ Present address: Department of Linguistics, Indiana University, 322 Memorial Hall, 1021 E. Third St., Bloomington, IN 47405, USA.

Since people do not automatically bind source information with their memories, their source attributions are not always accurate. Several studies that investigated people's ability to discriminate between external/perceptual and internal/self-generated sources of memories have found that participants often tended to mistakenly attribute self-generated representations to perception (Durso & Johnson, 1980; Johnson, Kahan, & Raye, 1984; Johnson, Raye, Wang, & Taylor, 1979; Johnson, Taylor, & Raye, 1977). In one study, participants reported having performed an activity (e.g., tracing the outline of a line drawing) that they had previously simply imagined performing (Anderson, 1984). In another study, participants reported having seen photographs of scenes that they had previously only read descriptions of (Intraub & Hoffman, 1992). In other studies, people misremembered pragmatic implications of sentences as explicitly stated (Bransford & Franks, 1971; Brewer, 1977; Chan & McDermott, 2006; Fazio & Marsh, 2010; Harris, 1974; Harris & Monaco, 1978; McDermott & Chan, 2006).

False event memories also emerge when people are presented with post-event visual information that is consistent with the information delivered by an event. In one study (Hannigan & Reinitz, 2001), adults were shown multiple photographs depicting typical events within an episode (e.g., grocery shopping). One photograph per episode depicted the consequence of an action (e.g., oranges on the floor of a grocery store). In a later memory test, participants were more likely to treat photographs depicting causes of the original actions (e.g., someone removing an orange from the bottom of a stack) as 'old' compared to causes of novel actions (cf. Lyons, Ghetti, & Cornoldi, 2010 for developmental differences in these errors). A subsequent experiment in the same study showed that participants were much less likely to make the opposite error (i.e., to mistakenly treat consequences of causal actions as 'old' after perceiving only the actions themselves; cf. Durso & Johnson, 1980; Johnson et al., 1977, 1979, 1984; Lindsay & Johnson, 1989). In a related study by Strickland and Keil (2011), conceptually coherent post-event information led people to falsely believe that they had seen the moment that an event unfolded. These and other studies have suggested that the opposite error (i.e., attributing perceived events to internal sources such as imagination or inference) is much less likely (Anderson, 1984; Dallett & Wilcox, 1968).

Linguistic evidentiality and source monitoring

One question left open in the literature is whether source monitoring might be susceptible to linguistic influences. This is important because the expression of source of information—known as *evidentiality*—varies across languages (Aikhenvald, 2004, 2014; Aikhenvald & Dixon, 2001; Anderson, 1986; Chafe & Nichols, 1986; De Haan, 2001; DeLancey, 2001; Faller, 2001; Givón, 1982; Willet, 1988). For instance, in English, speakers do not need to encode the source of their information in their utterances, even though evidentiality can be optionally

expressed through lexical devices such as verbs (e.g., *see*, *hear*), or adverbials (e.g., *allegedly*, *reportedly*). Thus, sentence (1) can be used whether the speaker has directly experienced the event or has only indirect information about it.

(1) Ali played.

In contrast, Turkish obligatorily expresses evidentiality through verb morphology for past events. In Turkish, two past tense markers, *-dı* and *-miş* (realized as *-dı*, *-di*, *-du*, *-dü*, *-ti*, *-tı*, *-tu*, *-tü* and *-miş*, *-miş*, *-muş*, *-müŝ*, respectively, depending on phonological factors) differentiate direct experience from indirect experience (Aksu & Slobin, 1986; Aksu-Koç, 1988; Johanson, 2003; Kornfilt, 1997; Slobin & Aksu, 1982). Therefore, sentence (1) must be expressed as either sentence (2) or (3) in Turkish depending on the speaker's source of information. In sentence (2), *-dı* conveys that the speaker has directly experienced the event; in sentence (3), *-miş* conveys that the speaker has indirect information about the event through either the verbal report of another person or inference based on visual evidence.

(2) Ali oyna-dı.
Ali play-PAST. Direct.3sg
(I saw that) Ali played

(3) Ali oyna-mış.
Ali play-PAST. Indirect.3sg
(I heard/inferred that) Ali played

Could cross-linguistic differences in the encoding of evidentiality influence source monitoring? This question relates to a broader debate in the literature about the role of language in cognition (see Gentner & Goldin-Meadow, 2003; Gleitman & Papafragou, 2005, 2012; Gumperz & Levinson, 1996; Landau, Dessalegn, & Goldberg, 2010; Levinson, 2003; Wolff & Holmes, 2011 for reviews).² Several commentators have proposed that language has a deep and lasting effect on cognitive processes (Boroditsky, 2003; Imai & Gentner, 1997; Levinson, 2003; Lucy, 1992; see Whorf, 1956 for an early statement of this position). On this view, habitual differences in linguistic framing of events across language communities may lead to stable differences in how members of these communities remember and reason about events even when they are not explicitly using language (i.e., speaking or comprehending speech). Other researchers have pointed out that linguistic effects on cognition may be more ephemeral in nature (Gennari et al., 2002; Landau et al., 2010; Li, Dunham, & Carey, 2009; Papafragou,

² What is at stake in this debate is whether language can impact memory and other cognitive processes in situations in which people are not required to speak. It is widely acknowledged that when people have to speak, cognitive resources are mobilized to serve speech planning (see Papafragou, Hulbert, & Trueswell, 2008, for cross-linguistic evidence). Furthermore, it is understood that, if people perform cognitive tasks right after speaking, performance may be shaped by language-specific patterns (see, e.g., Gennari, Sloman, Malt, & Fitch, 2002).

Massey, & Gleitman, 2002). Several studies have found meaningful differences between speakers of different languages; however, these differences often diminish or disappear when people are prevented from explicitly using language (e.g., Frank, Fedorenko, Lai, Saxe, & Gibson, 2012; Trueswell & Papafragou, 2010; Winawer et al., 2007). Furthermore, preventing access to language during only encoding or both encoding and retrieval has been found to impact cognitive performance to a greater extent compared to preventing access to language during only retrieval (Frank et al., 2012). These findings suggest that implicit (subvocal) linguistic encoding can be recruited online to support cognitive processing, but that language might not permanently alter conceptual representations.

The two positions on the language–cognition interface make different predictions about whether and how language might become implicated in source monitoring. Recall that people commit source monitoring errors because they do not automatically tag their memories with source information but need to reconstruct source information on the basis of different types of cues (Johnson, 1988; Johnson et al., 1993). Source monitoring errors might originate either from failures in encoding the event characteristics relevant for source decisions or during the evaluative processes involved in retrieval (Johnson, 2006). If habitual use of language has lasting, long-term effects on people's cognitive processes, the systematic use of evidentiality markers in languages such as Turkish may facilitate source monitoring by creating general attentional biases that are active even when people do not speak. For instance, as a result of having to make distinctions between direct perception and inference from visual cues, Turkish speakers might be particularly good at attending to and encoding event characteristics that might later be useful in source monitoring decisions (see Johnson & Raye, 1981, for a description of such cues). Additionally, or alternatively, as a result of engaging in source decisions more frequently (in order to speak their language properly), Turkish speakers might use more sensitive criteria during retrieval while evaluating the origin of an event memory. As a result, Turkish speakers might be more accurate and/or faster to respond in source monitoring tasks compared to English speakers. But if language does not shape source monitoring, people from different language communities should converge in their ability to remember the sources of their event memories.

There have been only a few studies addressing the relation between evidentiality and source monitoring and these studies have not settled the issues arising from these competing theoretical possibilities. In one study reported in Aksu-Koç, Ögel-Balaban, and Alp (2009), three- to six-year-old Turkish children learned about events through perception, inference, or hearsay and had to describe the events. Children also completed two source monitoring tasks. In the first task (source report; adapted from Gopnik & Graf, 1988), children learned the contents of a container through visual access, inference, or the experimenter's verbal report, and later reported how they learned what was inside the container. There was no relationship between children's production of evidentials and their performance in this task. In the second task (speaker

choice; adapted from Drumme & Newcombe, 2002), children heard a series of statements uttered by one of two speakers. In a later memory test, children had to attribute a given statement to a speaker. Children's use of the indirect marker (*-miş*) in the production task significantly predicted their ability to remember the person from whom they acquired information initially. Even though these findings demonstrate some relationship between evidentiality and source monitoring, they are hard to interpret for several reasons. First, it is unclear why use of evidentiality markers would predict performance on the speaker-choice task since the indirect marker in its hearsay use does not actually encode *the person* from whom the information is acquired. The mapping between evidentiality markers and information sources is more straightforward in the source-report task, but in this task, there was no relation between linguistic evidentiality and source monitoring (see also Robinson, 2009 for discussion). Second, even though Aksu-Koç and colleagues suggested that the 4-year-old Turkish children in their study performed better than the English-speaking children of the same age in Drumme and Newcombe's work, they did not directly compare English- and Turkish-speaking populations. Finally, there is evidence that the acquisition of linguistic evidentiality in Turkish and other languages lags behind the conceptual understanding of sources, a fact that suggests that language reflects rather than drives cognitive development in this domain (Ozturk & Papafragou, *in press*; Ünal & Papafragou, 2013).

In a cross-linguistic study that is more relevant for present purposes, Tosun, Vaid, and Geraci (2013) compared Turkish- and English-speaking adults' memories for firsthand vs. non-firsthand information. Monolingual Turkish and English speakers and Turkish–English bilinguals read 30 sentences presented one at a time on a computer screen. The sentences in Turkish were marked with either the direct past tense (*-di*) or the indirect past tense (*-miş*) morpheme. The sentences in English included either only a past tense verb (e.g., *Mary missed her flight*), or an adverbial and a past tense verb to indicate non-firsthand information (e.g., *Mary allegedly missed her flight*). In a later memory test, participants were presented with the original 30 sentences mixed with 30 new sentences and had to indicate whether they had read each sentence before, and if so in which form (firsthand vs. non-firsthand). English speakers were equally likely to recognize the two types of sentences and were able to indicate the original form of the sentence regardless of whether that was firsthand or non-firsthand. Both Turkish monolinguals and Turkish–English bilinguals were less likely to recognize as familiar sentences presented in the non-firsthand form and misremembered the non-firsthand sentences as having been in the firsthand form. The authors take this finding as supporting an “evidentiality effect” in which sentences with non-firsthand information that is obligatorily encoded as such by the grammar are “discounted because one cannot be as certain of their truth value” (Tosun et al., 2013, p. 132).

While Tosun et al.'s results suggest that the linguistic framing of information might influence memory accuracy, two questions remain open at this point. First, Tosun et al.

explicitly manipulated linguistic form and tested the effect of form on subsequent memory for the information presented in an utterance and the form in which the utterance was presented. These findings are reminiscent of findings from English speakers showing that choices about the explicit linguistic framing of information affect eyewitness memory (e.g., Hart & Albarricín, 2011; Strack & Bless, 1994). This method differs from typical investigations of the language–cognition interface that use non-linguistic tasks to test whether different language groups conceptualize information in different ways even if language is not involved in the task (e.g., Dolscheid, Shayan, Majid, & Casasanto, 2013; for a review, see Gleitman & Papafragou, 2005). Currently, it remains an open question whether English and Turkish speakers would differ in source monitoring in the absence of explicit linguistic processing. Second, and relatedly, Tosun et al.'s findings concern memory for *someone else's* information sources, as indicated by their verbal reports. It is unknown whether there would be cross-linguistic differences in source memory when participants are tested on *their own* information sources.

In the current study, we take up both of these questions. Specifically, we ask whether English and Turkish speakers differ in their event source memory in a task that does not involve processing linguistic material, but rather simply learning about events directly or indirectly. Throughout, we compare participants' memories for events they had learned about *themselves* (either through visual perception or through inference from post-event visual evidence). The current studies have some methodological improvements over Tosun et al.'s study. First, the present studies include an independent measure of general memory performance in order to establish equivalence across English and Turkish studies. This is an important control that was lacking in the earlier study. Furthermore, notice that in Tosun et al.'s study the stimuli presented to English and Turkish groups were not equivalent: Turkish monolinguals and bilinguals were tested on their memory for morphemes, whereas English speakers were tested on their memory for lexical items (i.e., evidential adverbs). Thus, English speakers had to report merely the presence or absence of an evidential adverb, whereas Turkish speakers had to report which of two evidential morphemes marked the verb. We overcome this limitation by comparing the two language groups on a non-linguistic measure of cognition using *the exact same* stimuli. In this respect, the current study follows standard practice in experimental tests of the relation between language and thought in the literature (Gleitman & Papafragou, 2005).

Overview of current study

In the current study, we take a multi-step approach in exploring the way the grammatical encoding of evidentiality interfaces with source monitoring cross-linguistically. In Experiment 1, we sought to confirm and clarify the differences in the linguistic encoding of event sources between English- and Turkish-speaking adults. Obviously, in some contexts English speakers often exhibit very nuanced linguistic encoding of sources (legal and scientific

discourse is one such case). Our focus here was on ordinary, everyday situations in which English speakers *could* mark evidence linguistically but do not. We were especially interested in exploring the conditions under which the Turkish indirect evidential is used to mark inference from visual premises, and whether more careful inspection of its use might reveal sensitivity to different types of evidence that could have further cognitive implications. We were also interested in whether the evidential distinctions in the Turkish system might have a parallel in the intuitions of English speakers about what counts as a directly “seen” vs. an “inferred” event.

In Experiments 2 and 3, we asked whether English and Turkish speakers would differ in source monitoring—specifically, the ability to distinguish between direct visual evidence for an event vs. inferences based on post-event visual evidence in a memory task. Because prior work has revealed that source monitoring errors follow a certain direction (from internally generated representations to perception as opposed to the other way around; Durso & Johnson, 1980; Intraub & Hoffman, 1992; Johnson et al., 1977, 1979, 1984; Lindsay & Johnson, 1989), we focused on inference-to-perception errors, for which the potential benefits of speaking a language with obligatory evidentiality might be more pronounced.

Our source monitoring tasks departed from prior work on post-event inference (e.g., Hannigan & Reinitz, 2001; Strickland & Keil, 2011) in several ways. First, rather than looking at whether visuals (photo/video) of multiple time points in an event lead people to think they have perceived unseen parts of an event, we asked whether errors emerge if participants are given a single post-event view (thereby widening the gap between what was seen and what was inferred). Such situations more closely approximate real-life circumstances (e.g., eyewitness testimony) in which people often have to inferentially reconstruct an event on the basis of very limited visual evidence.

Moreover, unlike past studies that relied exclusively on forced-choice (old–new) recognition memory, we compared forced-choice recognition and multiple-choice source monitoring tasks that oriented people to the origins of their memories. There is considerable evidence that the two types of tasks are sometimes based on different information or processes, with detailed source attributions requiring more specific information compared to forced-choice recognition (Anderson, 1984; Hayes-Roth & Thorndyke, 1979; Johnson, Kounios, & Reeder, 1994; Johnson & Raye, 1981; Johnson et al., 1993; Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989). Most relevant for present purposes, in several studies, people who were engaged in a forced-choice task made more source misattributions compared to people who were asked more specific questions about the origin of their memories (e.g., whether an item was presented “only in picture”, “only in text”, “both in picture and in text” or “neither”, Lindsay & Johnson, 1989; cf. Hayes-Roth & Thorndyke, 1979). We reasoned that performance on a simple recognition task may mask people's ability to reconstruct the sources of their memories.

Finally, unlike most prior studies of source monitoring, in addition to obtaining accuracy data on people's source

discriminations, we also collected reaction time data to be able to examine the time course of such discriminations and compare them across people with different language backgrounds (see also Johnson et al., 1994).

Experiment 1

Experiment 1 contained two tasks. In a Linguistic task, we asked speakers of English and Turkish to describe photographs depicting the point after which an event took place (so that the event could be inferred) and photographs depicting the point at which an event was unfolding (so that the event could be seen). These stimuli would be used in subsequent cross-linguistic memory comparisons (Experiments 2 and 3). Our goal was to ensure that (a) both English and Turkish speakers would identify the events in similar ways but (b) only Turkish speakers would mark information about how they learned about each event in their descriptions. In a Source Identification task, we asked a new group of English speakers to judge whether each of these events was “seen” or “inferred”. We hypothesized that evidential distinctions encoded in Turkish might be reflected in the source judgments of this population that does not mark information source grammatically.

Method

Participants

In the Linguistic task, participants were 14 native speakers of English ($M_{\text{age}} = 19.1$, range = 18–22, 9 females) and 12 native speakers of Turkish ($M_{\text{age}} = 19.6$, range = 18–23, 8 females). In the Source Identification Task, participants were a new group of 24 English speakers ($M_{\text{age}} = 18.57$, range = 18–20, 17 females). All English speakers were recruited at the University of Delaware and all Turkish speakers were recruited at Koç University, Turkey. All participants received course credit for their participation.

Stimuli

Forty-four photographs of people engaged in various events served as the stimuli. Thirty-two photographs represented 16 target events organized in pairs in which the same event was presented through a different source: “inferred” events were depicted by the point after which the event was completed so that it could be inferred on the basis of visual evidence (for instance, a person was shown in a kitchen holding a paper towel); “seen” events were depicted by the point at which the event was unfolding when it could still be directly seen (in the same scene, a person was shown tearing a paper towel off a roll). The list of target events is presented in the Appendix A. Twelve additional photographs served as filler events.

Two lists of 28 stimuli were created, each containing 12 filler and 16 target events. Filler events were exactly the same across the lists. Within each list, half of target events were “inferred” events and the other half were “seen” events. Each list included only one version (either “inferred” or “seen”) of a given target event, and thus source for a given target event (“inferred” or “seen”) was

counterbalanced across lists. In each list, stimuli were arranged in a single fixed order.

Procedure

All participants were tested in their native language. In the Linguistic task, Turkish speakers were tested individually using a 13-in. MacBook Pro laptop. They were told that they would see photographs of past events and were asked to describe what happened in each photograph with one sentence by typing their responses in a box under the photograph. English speakers were tested in small groups. They were told that they would see photographs of past events and were asked to describe what happened in each photograph with one sentence by writing it on a response sheet. Photographs were projected on a screen one at a time. In the Source task, participants were tested in small groups. Participants were told that they would see photographs of “past” events one at a time projected on a screen. They were asked to write down what happened in each photograph using a single verb and indicate whether they had seen or inferred it by selecting “seen” or “inferred” on the response sheet. Each photograph remained visible until participants gave their answer.

Results

Linguistic task

For each language, descriptions were coded by a native speaker of the language under consideration. As a first step, we used these descriptions to ensure that participants in both language groups would interpret the stimuli in the same way. We coded how often descriptions included the intended verb for the target event (e.g., “The man pulled off a paper towel” for the picture of a man holding a paper towel), as opposed to other information (e.g., “The man is going to wipe something”, or “The man is in the kitchen” for the same event). Both English and Turkish speakers overwhelmingly identified the target events as expected ($M = 0.90$ and 0.89 , respectively, based on all descriptions).

Next, we sought to confirm the presence of cross-linguistic differences in how evidence for an event would be encoded. Both language groups sometimes used non-past utterances to describe the target events. For inferred events, these descriptions consisted of 23% of the data in both English and Turkish speakers. For seen events, non-past tense utterances consisted of 32% of the Turkish descriptions and 30% of the English descriptions. Because the frequency of non-past descriptions did not differ across language groups, we excluded these non-past utterances from our analyses.³

In the remaining descriptions, as expected, English speakers did not use any evidentiality devices, but Turkish speakers regularly did so. Specifically, Turkish speakers described seen events using the direct morpheme *-dı* ($M = 0.73$, range = 0.50–1.00), as opposed to the indirect morpheme *-miş* ($M = 0.25$), ($t(15) = 9.49$, $p < .001$), and inferred events using the indirect morpheme *-miş*

³ This step did not affect correct identification of the target event ($M = 0.92$ and 0.93 for English and Turkish speakers, respectively, in past-tense descriptions alone).

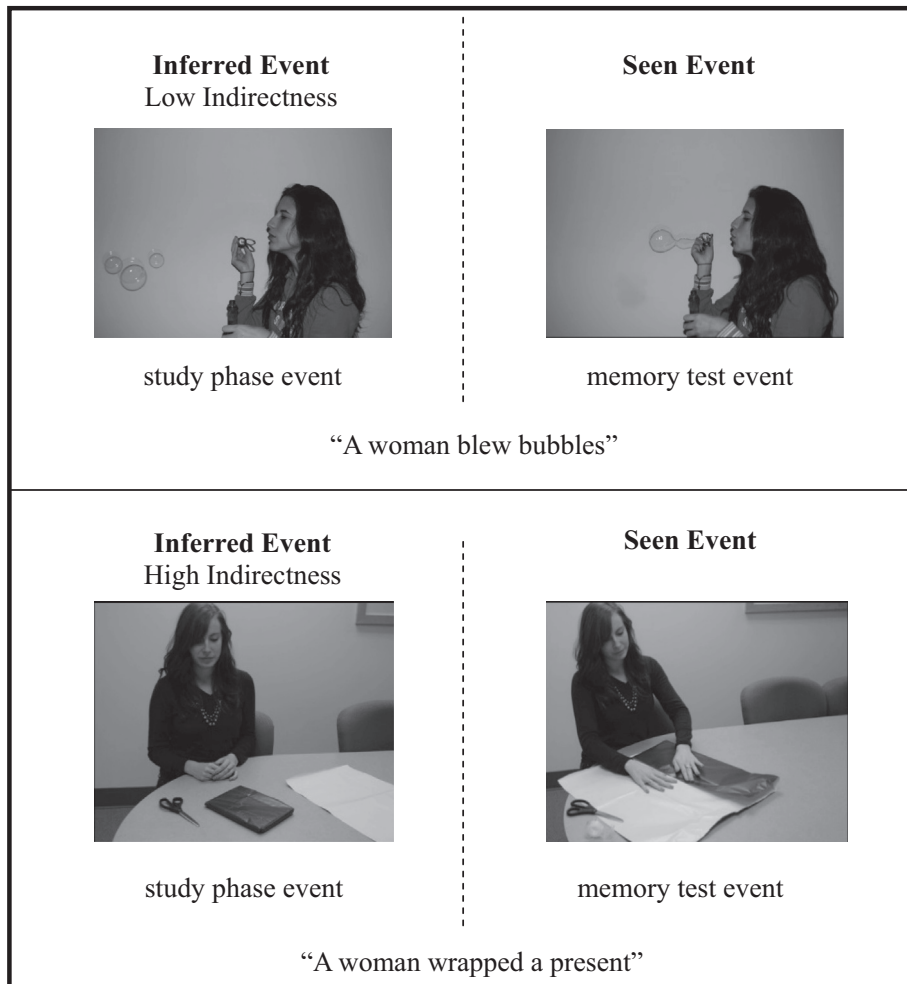


Fig. 1. Examples of target events undergoing Source Changes. The top panel depicts a Source Change pair with a Low Indirectness event and the bottom panel depicts a Source Change pair with a High Indirectness event. Inferred events were presented during the Study phase and Seen events were presented during the Memory phase.

($M = 0.64$, range = 0.29–1.00), as opposed to the direct morpheme *-dı* ($M = 0.36$), ($t(15) = 2.61$, $p = .020$). Participants' descriptions only included *-miş* (and not variants such as *-mişmiş* or *-mişdir*).

Closer examination of Turkish speakers' descriptions of the inferred events revealed that these events varied in the extent to which they prompted the use of the indirect marker *-miş*. Using a median split on *-miş* uses, the inferred events were placed into one of two Indirectness groups: High (items 1–8 in Appendix A) vs. Low (items 9–16 in Appendix A; see also Fig. 1). Mean use of *-miş* was 81% (range = 0.60–1.00) for the High Indirectness events but only 48% (range = 0.29–0.57) for the Low Indirectness events. These two means were significantly different from each other, $t(14) = 4.31$, $p < .001$.⁴ Inspection of the events

within each group suggests that Turkish adults' choice of an evidential form depends on the quality of post-event visual evidence. For some events (e.g., blowing bubbles), post-event information (e.g., someone next to a bubble traveling through the air) allowed observers to reconstruct the event with great clarity and precision, such that the boundary between what was seen and what was inferred was blurred. For such events, the speakers were less likely to encode the presence of an inference using the indirect evidential, even though the speakers had not actually seen the event. For other, perhaps more complex, events involving multiple, visually distinct steps (e.g., wrapping a present), post-event information (e.g., someone next to some wrapping paper, scissors and a wrapped present) was more distinct from direct visual information. For such events, the presence of an inference was more likely to be encoded with the indirect evidential. The next task allowed us to test this hypothesis more directly: if the distribution of the Turkish evidential markers depends on the quality of visual evidence, it might have a corollary in the way English speakers'

⁴ We focus on inferred events since these had to be remembered in the memory task in Experiments 2 and 3. There was no difference in terms of eliciting the direct past tense between the "seen" versions of the High ($M = 0.76$) and Low ($M = 0.71$) Indirectness events ($t(14) = 0.42$, *ns*).

judgments draw the subtle boundary between perception and inference.

Source identification task

Participants correctly identified the target event 78% of the time ($M_{\text{seen}} = 0.78$, $M_{\text{inferred}} = 0.77$) when asked to give a single verb for each scene (failures to identify the event were mostly due to the tendency to produce more general as opposed to more specific verbs). We looked at source assessments for correctly identified events. For seen events, participants reported having seen what happened in 109 of the 138 responses (or 79% of the time), whereas for inferred events, participants reported having inferred what happened in 101 of the 169 responses (or 60% of the time). The distribution of “seen” vs. “inferred” responses was significantly different for the seen and inferred events, $\chi^2(1, N = 307) = 46.72, p < .001$. The distribution of the “seen” and “inferred” responses for the two types of inferred events were examined separately using the median split of the Turkish descriptions in the Linguistic task. For High Indirectness inferred events, participants reported having inferred what happened in 64 of the 92 responses (or 70% of the time), whereas for Low Indirectness inferred events that were potentially closer to perception, participants reported having inferred what happened in only 37 of the 77 responses (or 48% of the time). These distributions were also significantly different from each other, $\chi^2(1, N = 169) = 8.07, p = .005$. Furthermore, the inferability judgments from this group of English speakers (as calculated by proportion of “inferred” responses out of total responses for each event) correlated significantly with the use of the indirect (*-miş*) morpheme in the Turkish descriptions obtained in the Linguistic task ($r(14) = .62, p = .010$).

Discussion

Experiment 1 demonstrated that both English and Turkish speakers draw inferences from post-event visual cues, but only Turkish speakers mark such inferences linguistically – through an indirect past-tense morpheme – and distinguish them from direct perception. Furthermore, Turkish speakers reserve the indirect evidential marker primarily for inferred events that are further removed from perception (High Indirectness events). Finally, Turkish speakers’ use of the indirect evidential closely maps onto English speakers’ judgments of whether an event was seen or inferred.

Our data confirm the presence of strong cross-linguistic differences in the way adults encode evidentiality in language (and complements developmental studies in Aksu & Slobin, 1986; Aksu-Koç, 1988, 2000; Aksu-Koç et al., 2009; Choi, 1995; Fitneva, 2009; Ozturk & Papafragou, in press; Papafragou, Li, Choi, & Han, 2007; Ünal & Papafragou, 2013, submitted for publication; de Villiers, Garfield, Gernet-Girard, Roeper, & Speas, 2009, see Matsui, 2014 for a review). The present data lead naturally to the question of whether the cross-linguistic differences

in source encoding for these events would be reflected in Turkish and English speakers’ event source memory. In Experiments 2 and 3, we put this question to test.

Nevertheless, the fact that English speakers’ reflective judgments about the source of a past event tracked the subtle use of evidential morphology in Turkish shows that the representations underlying grammatical evidential systems are not language-specific but are instead available to speakers of languages that lack those distinctions in grammar. More generally, these data suggest that inferences from post-event visual cues are not a homogeneous class of indirect evidence. Rather, inference from post-event visual cues and visual perception lie on a continuum indicating varying levels of directness (the higher the quality of visual evidence, the more direct the event is judged as). This continuum underlies both linguistic systems of evidentiality and judgments of how one found out about an event. We return to this conclusion in the experiments and discussion to follow.

Experiment 2

The goal of Experiment 2 was to test whether English and Turkish speakers would differ in terms of their memories for the sources of the events in Experiment 1. Specifically, we tested whether Turkish speakers who studied the point after which an event took place (the “inferred” events of Experiment 1) would be less likely compared to English speakers to misidentify the source of their memory by reporting that they had originally seen the point at which the event unfolded (cf. the “seen” events of Experiment 1). We were particularly interested in comparing source memory for inferred events of High vs. Low Indirectness (as defined by both Turkish speakers’ use of the indirect past tense and by English speakers’ inferability judgments in Experiment 1). Of interest was whether the speed or accuracy with which participants identified inference as opposed to perception as the source of the event memory would differ for the two levels of Indirectness, and whether this difference would be greater for Turkish speakers. To ensure that whatever differences we obtained would be independent of potential differences in general memory performance between English and Turkish speakers, we included as a control a memory task that did not involve source monitoring.

Because prior work has shown that asking participants to explicitly report the sources of their memories reduces source monitoring errors compared to simple forced-choice recognition tasks (see Hayes-Roth & Thorndyke, 1979; Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989, among others), half of the participants reported whether they had previously seen the target events or not and the other half reported whether they had seen, inferred or not seen or inferred the target events. We were interested in seeing whether the more specific, three-choice response option would lead to greater accuracy with source memory compared to the coarser,

two-choice option, and whether response options would interact with participants' language background.

Method

Participants

Participants were a new group of 48 English speakers ($M_{\text{age}} = 19.29$, range = 18–29, 27 females) and 48 Turkish speakers ($M_{\text{age}} = 21.71$, range = 18–33, 33 females). English speakers were recruited at the University of Delaware and Turkish speakers were recruited at Koç University, Turkey. Participants received course credit or \$10 for their participation. Data from 14 additional participants were discarded due to failure to use one of the keys to respond in the memory task ($n = 12$), or experimenter error ($n = 2$).

Critical task

The critical task served as a measure of source memory.

Stimuli. Ninety-six photographs of people engaged in various events served as the stimuli. There were three types of stimuli: Source Changes (16 pairs of photographs), Event Changes (32 photographs) and No Changes (16 photographs). Source Changes consisted of the same pairs of photographs for “inferred” and “seen” events as in Experiment 1. Inferred events were further classified in terms of Indirectness (High, Low) based on the linguistic distinction observed in Turkish speakers in Experiment 1. Event Changes were photographs of different events (e.g., a woman bit an apple, a woman peeled a banana): half of these events were presented in the study phase, and the other half in the memory (test) phase. No Changes consisted of the 12 filler events in Experiments 1 and 4 additional events that were included to have an equal number of items per type of stimuli (e.g., a man raked leaves). These events did not change across the study phase and the memory (test) phase. To ensure similarity to the Source Change stimuli (and thus discourage simple response strategies in the memory task), within each of the sets of Event Changes and No Changes, half of the events were depicted at the point after they were completed (similar to the “inferred” events in Source Changes) and the other half at the point at which they unfolded (similar to the “seen” events in the Source Changes). Additionally, we counterbalanced the assignment of Event Changes and No Changes stimuli to the memory and test phase such that each phase would have an equal number of ongoing vs. completed events from each stimulus type.

These stimuli were used to create separate lists (each consisting of 48 photographs) for the study and memory phase of the Critical task. The list for the study phase consisted of the inferred event from each of the 16 Source Change pairs (8 High Indirectness events, 8 Low Indirectness events), 16 events chosen from the Event Changes (8 unfolding and 8 completed) and the 16 No Changes (also 8 unfolding and 8 completed events). Items were arranged in a single fixed order in the study phase list. The list for the memory phase consisted of the seen event from each of the 16 Source Change pairs, the 16 Event Changes that were not used in the study phase (8 unfolding events and 8 completed events) and the 16 No Changes (also shown

during the study phase). Items were arranged in a different random order for each participant.

Procedure. Participants were tested using E-Prime (Schneider, Eschman, & Zuccolotto, 2002a, 2002b) run on a Dell Latitude E6520 computer. All participants were tested in their native language.

In the study phase, participants studied a set of photographs for 3 s each in preparation for a memory test. They were told that the photographs were of events in the past. The memory phase was presented immediately after the study phase. In the memory phase, participants were presented with another set of photographs as described above. Participants were assigned to one of two conditions. Participants in the Coarse Source Reporting condition were asked to indicate whether they had “seen” or “not seen” the event by pressing a computer key as quickly as possible. The correct response for No Changes was “seen” and the correct response for Source Changes and Event Changes was “not seen.” Participants in the Specific Source Reporting condition were asked to indicate whether they had “seen”, “inferred”, or “not seen or inferred” the event by pressing a computer key as quickly as possible. The correct response for No Changes was “seen”, the correct response for Source Changes was “inferred”, and the correct response for Event Changes was “not seen or inferred”. Participants responded using their dominant hand.

At the beginning of the Critical Task, all participants received three pairs of training items, one example of a No Change, one example of a Source Change, and one example of an Event Change. Later, participants in the Coarse Source Reporting condition were asked to indicate whether they had “seen” or “not seen” these events before by pressing a computer key. Participants in the Specific Source Reporting condition were asked to indicate whether they had “seen”, “inferred” or “not seen or inferred” these events by pressing a computer key. Participants did not receive any feedback regarding the accuracy of their responses.

Control task

The Control task served as a general measure of memory equivalence between the two language groups.

Stimuli. Ninety-six photographs of people engaged in various events served as the stimuli. There were two kinds of stimuli: Object Changes (24 pairs of pictures) and No Changes (24 pictures). Object Change stimulus pairs depicted the same event with two different objects. For instance, the study phase item would be a picture of a man reaching for apples and the memory phase item would be a picture of a man reaching for pears. No Changes were events that did not change across the study phase and the memory (test) phase.

Two separate lists, each containing 48 photographs, were created using these stimuli. The list for the study phase consisted of the 24 events from Object Change pairs and 24 No Changes. The list for the study phase arranged items in a single fixed order. The list for the memory phase consisted of 24 complementary events from Object Change

pairs and the 24 No Changes (also shown during the study phase). Items were arranged in a different random order for each participant in the memory phase list.

Procedure. Participants were tested using E-Prime (Schneider et al., 2002a, 2002b) run on a Dell Latitude E6520 computer. Half of the participants received the Control Task first, and the other half received it after the Critical Task. All participants were tested in their native language.

The Control task consisted of a study phase and a memory phase that was presented immediately after the study phase. In the study phase, participants studied a set of photographs for 3 s each in preparation for a memory test. In the memory phase, participants were presented with the set of photographs from the memory phase list and were asked to indicate whether they had “seen” or “not seen” each event before by pressing a computer key as quickly as possible. The correct response for No Changes was “seen” and the correct response for Object Changes was “not seen”. Participants responded using their dominant hand.

At the beginning of the task, participants received two training items, one example of an Object Change and one example of a No Change. They were asked to study the events in preparation for a memory task. Later, participants had to indicate whether they had “seen” or “not seen” these events before by pressing a computer key. Participants did not receive feedback on their accuracy in the training trials.

Results and discussion

Critical task

Accuracy data. The descriptive statistics for accuracy on the three types of change by Turkish and English speakers across source reporting conditions are presented in Table 1. Participants performed above chance level for No Changes, Event Changes, and Source Changes in both Specific and Coarse Source Reporting (single-sample *t*-tests, all $p < .001$, two-tailed).

We assessed each type of stimulus separately since correct responses differed both across stimulus types and between source reporting conditions. The mean accuracy for No Changes was submitted to a 2×2 between-subjects ANOVA with Language (English, Turkish) and Condition (Coarse Source Reporting, Specific Source Reporting) as factors. The ANOVA revealed only a significant main effect of Condition ($F(1,92) = 4.44$, $p = .038$, $\eta^2 = .046$): participants were more accurate in the Coarse Reporting condition ($M = 0.88$) than in the Specific Source Reporting condition ($M = 0.79$). The same analysis for Event Change items also returned a significant main effect of Condition ($F(1,92) = 19.10$, $p < .001$, $\eta^2 = .172$), such that participants in the Coarse Source Reporting condition ($M = 0.95$) were more accurate at identifying events that they had not seen than participants in the Specific Source Reporting condition ($M = 0.76$).

A similar ANOVA for Source Changes using Indirectness (High, Low) as an additional (within-subjects) factor returned a significant main effect of Indirectness ($F(1,92) =$

6.27 , $p = .014$, $\eta^2 = .064$), but this effect was qualified by an interaction between Indirectness and Condition ($F(1,91)$, 4.49 , $p = .038$, $\eta^2 = .046$). When Indirectness was High, accuracy did not change across Coarse ($M = 0.73$) and Specific ($M = 0.69$) Source Reporting conditions ($t(94) = 0.91$, *ns*), but when Indirectness was Low, accuracy was lower in the Specific ($M = 0.59$) than in the Coarse Source Reporting condition ($M = 0.72$) ($t(94) = 2.47$, $p = .015$). Thus, English and Turkish speakers did not differ in terms of their memories for sources; furthermore, both language groups had a harder time when making explicit source judgments compared to merely identifying that they had not seen these events before—but only for inferred events that were somewhat closer to perceived events.

Overall, then, the expectation that explicit source judgments would improve source monitoring is not borne out in our data. In fact, for inferences that are based on strong perceptual cues, we observe a drop in performance in the Specific Source Reporting condition compared to the Coarse Source Reporting condition.

Error analysis. To further explore performance within the Specific Source Reporting group, we performed an error analysis by comparing the frequency of choosing one vs. the other of the two incorrect responses. The results are presented in Table 2.

For No Changes, when participants responded incorrectly, they were equally likely to respond “inferred” ($M = 0.59$) vs. “not seen or inferred” ($M = 0.41$) ($t(36) = 1.43$, *ns*). For Event Changes, when participants responded incorrectly, they were more likely to say that they had “inferred” ($M = 0.82$) than “seen” ($M = 0.18$) the event ($t(38) = 6.31$, $p < .001$). Including Language as a factor in the analyses did not reveal any significant effects or interactions, and so it was omitted.

For Source Changes, we examined the frequency with which participants chose one of the two incorrect response options with a $2 \times 2 \times 2$ mixed ANOVA on the proportion of responses with Language (English, Turkish) as the between-subjects factor and Error Type (“seen”, “not seen or inferred”) and Indirectness (High, Low) as the within-subjects factors. The analysis revealed only a three-way interaction between Error Type, Language and Indirectness ($F(1,37)$, 5.94 , $p = .023$, $\eta^2 = .131$). Follow-up tests on this interaction revealed that, for Source Changes that characterized Low Indirectness events, when participants made an error, both language groups were more likely to pick “seen” as opposed to “not seen or inferred” ($t(42) = 2.15$, $p = .037$). However, for Source Changes that characterized High Indirectness events, when participants responded incorrectly, Turkish speakers were equally likely to respond with “seen” vs. “not seen or inferred” ($t(19) = -1.07$, *ns*), and in English speakers there was a trend indicating that they were more likely to respond with “seen” than “not seen or inferred” ($t(20) = 1.91$, $p = .070$). Furthermore, for Low Indirectness events, both groups were equally likely to make each of the two possible types of error but, for High Indirectness events, the frequency of “seen” errors was higher in English speakers ($t(39) = 2.08$, $p = .045$) and the frequency of “not seen or inferred” errors was higher in Turkish speakers ($t(39) = -2.08$, $p = .045$). An

Table 1

Means (SD) of proportion of accurate responses across types of change and language groups (Experiment 2).

		No Change	Event Change	Source Change	
				High Indirectness	Low Indirectness
English	Coarse Source Reporting	0.87 (0.13)	0.94 (0.12)	0.72 (0.27)	0.74 (0.29)
	Specific Source Reporting	0.85 (0.12)	0.77 (0.32)	0.63 (0.30)	0.59 (0.30)
Turkish	Coarse Source Reporting	0.88 (0.14)	0.97 (0.06)	0.73 (0.22)	0.70 (0.16)
	Specific Source Reporting	0.73 (0.30)	0.75 (0.27)	0.73 (0.26)	0.58 (0.29)

Table 2

Mean (SD) of errors to responses on the critical task in the specific source reporting condition (calculated over the total number of errors) (Experiment 2).

		No Change		Event Change		Source Change			
						High Indirectness		Low Indirectness	
		"Inferred"	"Not seen or inferred"	"Seen"	"Inferred"	"Seen"	"Not seen or inferred"	"Seen"	"Not seen or inferred"
English		0.57 (0.35)	0.42 (0.35)	0.12 (0.26)	0.88 (0.26)	0.65 (0.37)	0.35 (0.37)	0.63 (0.39)	0.37 (0.39)
		0.61 (0.42)	0.39 (0.42)	0.24 (0.35)	0.76 (0.35)	0.41 (0.39)	0.59 (0.39)	0.63 (0.39)	0.37 (0.39)

additional way of exploring the interaction revealed that the frequency of “seen” errors was lower for High Indirectness events compared to Low indirectness events in Turkish speakers ($t(18) = -2.86, p = .010$), but the frequency of “seen” errors did not differ for High vs. Low Indirectness events in English speakers ($t(19) = 0.51, ns$). Furthermore, the frequency of “not seen or inferred” errors was higher in High Indirectness events than it was for Low Indirectness events in Turkish speakers ($t(18) = 2.89, p = .010$), whereas the frequency of “not seen or inferred” errors was similar for High Indirectness and Low Indirectness events in English speakers ($t(19) = -0.51, ns$).

The pattern of the errors for No Changes suggests that, when participants failed to recognize an already-seen event, they were randomly choosing one of the two other options that indicated a difference in the event. For Event Changes, participants avoided choosing “seen” and instead chose “inferred”, the option that indicated some kind of change to the event that they were initially presented with. Finally, the error pattern for Source Changes suggests that participants across languages were likely to attribute Low Indirectness events that were inferred from visual evidence to perception. For High Indirectness events undergoing a Source Change, error patterns were not directional in Turkish and were marginally directional (toward “seen” responses) in English. For this last class of events, “seen” errors were higher in English compared to Turkish speakers and “not seen or inferred” errors were more likely in Turkish compared to English speakers. We return to the significance of these patterns in the Discussion below.

Reaction time data. Mean reaction time data for the critical task are presented in Table 3. First, the reaction times for accurate responses for No Changes were subjected to a 2×2 ANOVA with Language (English, Turkish) and Condi-

tion (Coarse Source Reporting, Specific Source Reporting) as between-subjects factors. The analysis revealed a main effect of Condition ($F(1,90) = 11.62, p = .001, \eta^2 = .114$), but this effect was qualified by an interaction between Language and Condition ($F(1,90) = 6.46, p = .013, \eta^2 = .067$): in the Coarse Source Reporting condition Turkish and English speakers were equally fast ($t(46) = 0.62, ns$), whereas in the Specific Source Reporting condition, Turkish speakers were slower than English speakers ($t(44) = 2.62, p = .012$). The same analysis for Event Changes showed that the main effect of Condition was significant ($F(1,91) = 50.03, p < .001, \eta^2 = .355$): across the two language groups, participants were slower in the Specific Source Reporting condition ($M = 1829.40$) than in the Coarse Source Reporting condition ($M = 1139.31$).

Finally, a similar ANOVA on the reaction times for accurate responses for Source Changes with Indirectness (High, Low) as an additional factor returned a significant main effect of Condition ($F(1,87) = 11.06, p = .001, \eta^2 = .113$), as well as an interaction between Language and Condition ($F(1,87) = 4.27, p = .042, \eta^2 = .047$): in the Coarse Source Reporting condition, Turkish and English speakers were equally fast to respond ($t(45) = 0.72, ns$); in the Specific Source Reporting condition Turkish speakers were slower than English speakers ($t(42) = 2.14, p = .038$). Notice that Turkish speakers seem to take longer to respond compared to English speakers, not only for Source Changes but also for No Changes in the Specific Source Reporting condition of the Critical Task (cf. also Object Changes in the Control task below). We consider the interpretation of this finding in the Discussion below.

Control task

Accuracy data. Table 4 presents the results for the Control Task. Overall, participants were highly accurate in the

Table 3

Means (SD) of Reaction times for accurate responses across types of change and language groups (in ms) (Experiment 2).

		No Change	Event Change	Source Change	
				High Indirectness	Low Indirectness
English	Coarse Source Reporting	1252.77 (368.81)	1143.59 (373.16)	1566.30 (951.58)	1578.23 (660.71)
	Specific Source Reporting	1328.12 (352.00)	1650.69 (566.94)	1721.43 (557.49)	1745.41 (528.16)
Turkish	Coarse Source Reporting	1191.88 (306.26)	1135.03 (199.89)	1490.26 (476.11)	1405.92 (340.96)
	Specific Source Reporting	1708.25 (608.33)	2008.11 (641.40)	2182.91 (949.43)	2091.60 (817.26)

Table 4

Means (SD) of proportion of accurate responses and means (SD) of reaction times for accurate responses (in ms) across types of change and language groups in the control task (Experiment 2).

	Accuracy		Reaction time	
	No Change	Object Change	No Change	Object Change
English	0.87 (0.10)	0.78 (0.15)	1283.33 (420.11)	1400.18 (368.09)
Turkish	0.85 (0.14)	0.73 (0.12)	1261.74 (257.49)	1609.78 (442.15)

Control task and performed above chance level for both No Changes and Object Changes (single-sample *t*-tests, all $p < .001$, two-tailed).

First, we tested whether English and Turkish participants had equivalent memory performance with a one-way between-subjects ANOVA on the accuracy for No Changes and Object Changes separately with Language (English, Turkish) as the factor. The analyses did not return any language effects on either No Changes or Object Changes.

Next, we calculated *d*-prime and criterion (*c*) scores for each participant and assessed whether sensitivity to changes and the criterion adopted when detecting changes differs across language groups. Sensitivity and criterion scores did not differ across speakers of English ($M_{d'} = 1.88$, $M_c = 0.17$) and Turkish ($M_{d'} = 1.69$, $M_c = 0.19$). The *d'* analysis demonstrates equivalence among speakers of English and Turkish in terms of sensitivity to detect changes. Overall, these findings show that English and Turkish speakers had comparable general memory performance in a task that did not involve memory for sources.

Reaction time data. We assessed whether English and Turkish speakers differ in how fast they respond to No Change and Object Change items in separate one-way between-subjects ANOVAs with Language (English, Turkish) as a factor. The analysis for No Changes did not reveal any significant differences across language groups. The same analysis for Object Changes revealed a Language effect ($F(1,94) = 6.37$, $p = .013$, $\eta^2 = .063$), such that English speakers were faster than Turkish speakers. These data suggest that English and Turkish speakers were equivalent

in general memory accuracy, but Turkish speakers might be slower to respond than English speakers.

Discussion

The findings of Experiment 2 demonstrate that participants were highly accurate in a memory task that did not involve memory for sources (Control task). However, their memory for sources was in error 36% of the time (Critical task). Moreover, participants committed source monitoring errors to a greater extent for inferred events that were relatively closer to perceived events. This suggests that the directness of inference from visual evidence plays a role in the degree to which inferential reasoning leads to source monitoring errors. Importantly, both high- and low-indirectness inferences led to source monitoring errors regardless of the participants' native language, suggesting that speaking a language with obligatory/grammaticalized evidentiality such as Turkish may not prevent source monitoring errors compared to speaking a language lacking grammaticalized evidentials such as English. Furthermore, Turkish speakers were not faster than English speakers in accurately identifying Source Changes – in fact, they were slower in the Specific Source Reporting condition. It is obvious that this asymmetry does not reflect a speed-accuracy trade-off. One might hypothesize that the slower responses of Turkish speakers might indicate an “evidentiality” effect, such that Turkish speakers were being more deliberate in their source decisions as a consequence of their experience with the evidential categories in their language. However, the fact that Turkish speakers were slower in the Specific Source Reporting condition both for Source Changes and also for No Changes (and were also slower for Object Changes in the Control Task) suggests that this hypothesis is unlikely to be true.

There is a further aspect of our findings in the Specific Source Reporting condition that is worth discussing. In that condition, when participants made source monitoring errors with Low Indirectness events, they typically characterized the events as “seen” (and did so to the same extent across languages). This is reminiscent of the finding that people misremember events that they have acquired from indirect sources such as reading or imagination as “seen” (see Durso & Johnson, 1980; Johnson et al., 1977, 1979, 1984; Lindsay & Johnson, 1989). For High Indirectness events, when participants responded incorrectly, English

speakers showed a trend to report having seen the events that they had only inferred and Turkish speakers had no bias to choose the “seen” over the “not seen or inferred” option. Overall, for this class of events, there were more “seen” errors in English compared to Turkish speakers, and more “not seen or inferred” errors in Turkish compared to English speakers.

One possible hypothesis is that this language-specific pattern is an effect of grammaticalized evidentiality on source monitoring. We do not think that this is likely. Recall that the proposal that language affects cognition predicts that Turkish speakers should commit source monitoring errors to a lesser extent compared to English speakers. However, this proposal makes no predictions regarding the kinds of errors expected of each group. In fact, the starting point of this proposal is that both language groups should identify the base events in the same way but would differ in the way they would encode access to these events (direct vs. indirect). The fact that Turkish speakers are more likely than English speakers to treat Source Changes for highly indirect events as entirely novel events (“not seen or inferred”) is thus unexpected.

An alternative hypothesis is that the error patterns of Turkish speakers are the result of a slightly weaker understanding of the task (and of overall uncertainty), especially in the more demanding Specific Source Reporting condition of the Critical task, compared to the English group. This hypothesis meshes well with the fact that the Turkish group was slower to respond in the No Change and Source Change items in the Specific Source Reporting condition (and is consistent with the slower performance of the Turkish group in the Object condition of the Control task). In Experiment 3, we revisit these two hypotheses so as to clarify the origin and robustness of the differences in error patterns between the two language groups in Experiment 2.

A final observation about the results from the present experiment is that, across language groups, and contrary to what we expected based on prior work (e.g., Hayes-Roth & Thorndyke, 1979; Lindsay & Johnson, 1989), sensitivity to the distinction between inferred and seen events did not increase when we asked participants to report the specific sources of their memories. In fact, in terms of both accuracy and reaction times, the Specific Source Reporting condition was found to be more challenging compared to the Coarse Source Reporting condition for Low-Indirectness events (i.e., events that were close to perception). We return to the significance of this finding in the General Discussion.

Experiment 3

A possible concern with the design of Experiment 2 is that participants in the Specific Source Reporting condition were not given sufficient information on how to draw the subtle distinction between having seen an event vs. having inferred it on the basis of (visual) evidence. Thus, performance on the Specific Source Reporting condition may have underestimated people’s ability to distinguish sources of event memories. A first aim of Experiment 3

was to investigate whether providing extra training on what the “inferred” response option meant would increase the participants’ ability to distinguish inferred events from already seen or completely new events. A second aim of Experiment 3 was to test whether improved performance on the Specific Source Reporting condition might more reliably reveal cross-linguistic differences in source monitoring compared to our earlier, less transparent task. A third, related aim was to clarify the nature of the difference in error patterns between English and Turkish speakers in the High Indirectness events of the Specific Source Reporting condition of Experiment 2. If the difference was driven by effects of the linguistic marking of evidentiality, the clarifications and training added in Experiment 3 should not matter, and the difference should reemerge. But if the difference was due to a weaker understanding of the task in the Turkish group, clarifying the nature of the task (especially, the meaning of the various response options and the way they mapped onto response keys) in Experiment 3 might make the difference disappear.

Method

Participants

Data were collected from a new group of 50 native speakers of English ($M_{\text{age}} = 18.60$, range = 18–21, 31 females) and 50 native speakers of Turkish ($M_{\text{age}} = 19.79$, range = 18–23, 31 females). English speakers were recruited at the University of Delaware, and Turkish speakers were recruited at Koç University, Turkey. All participants received course credit for their participation. The data from seven additional subjects were discarded due to following reasons: experimenter error ($n = 4$), background noise during the session ($n = 2$), or not being a native speaker of English ($n = 1$).

Stimuli

The stimuli were exactly the same as the stimuli in the Critical task of Experiment 2.

Procedure

Participants received only the Critical task. As in Experiment 2, half of the participants were assigned to the Coarse Source Reporting condition and the other half were assigned to the Specific Source Reporting condition. With a few exceptions (detailed below), the procedure was exactly the same as in Experiment 2.

A first set of changes involved the practice trials and was aimed at making the goal of the task (especially, the distinction between “seen” and “inferred” events) more transparent to participants. Participants were presented with two photographs side by side, the first corresponding to what they might see in the Study phase and the second one corresponding what they might be tested with in the Memory phase. There was one example of each type of change (No Change, Event Change, and Source Change).

Moreover, two new sets of practice items were created and replaced those for Experiment 2. In the first set, participants studied and were tested with three pairs of photographs: an example of a No Change, an example of an Event Change and an example of a Source Change. After

Table 5

Means (SD) of proportion of accurate responses across types of change and language groups (Experiment 3).

		No Change	Event Change	Source Change	
				High Indirectness	Low Indirectness
English	Coarse Source Reporting	0.91 (0.11)	0.95 (0.08)	0.72 (0.22)	0.63 (0.24)
	Specific Source Reporting	0.90 (0.10)	0.84 (0.21)	0.68 (0.23)	0.65 (0.24)
Turkish	Coarse Source Reporting	0.91 (0.08)	0.97 (0.06)	0.65 (0.18)	0.65 (0.25)
	Specific Source Reporting	0.82 (0.17)	0.82 (0.19)	0.70 (0.17)	0.60 (0.23)

participants responded to each item in the first practice memory phase, they were presented with a screen that indicated the correct response for that item. All participants received feedback regardless of the accuracy of their responses. In the second set of practice items, participants studied and were tested with six pairs of photographs: two No Changes, two Event Changes, and two Source Changes. In the memory phase for the second set of practice items, participants were not allowed to move to the next item unless they responded correctly. After this elaborated practice phase, participants proceeded to the main task.

A second set of changes involved the way responses were made. Participants in the Specific Source Reporting condition were asked to respond “neither” instead of “not seen or inferred” for events that they did not recognize. Furthermore, the response keys were labeled to help participants remember the response options. In English, for instance, in the Coarse Source Reporting condition, the A key was labeled as “S” (for “seen”) and the L key was labeled as “NS” (for “not seen”). In the Specific Source Reporting condition, the A key was labeled as “S” (for “seen”), the L key was labeled as “I” (for “inferred”), and the Spacebar key was labeled as “NEITHER.”

Results

Accuracy data

Table 5 presents the descriptive statistics for the accuracy data. As in Experiment 2, both language groups performed above chance level on No Changes, Event Changes, and Source Changes in both Coarse and Specific Source Reporting conditions (single-sample *t*-tests, all $p < .05$, two-tailed).

First, the mean accuracy for No Changes was submitted to a 2×2 ANOVA with Language (English, Turkish) and Condition (Coarse Source Reporting, Specific Source Reporting) as between-subjects variables. The ANOVA revealed only a significant main effect of Condition ($F(1,96) = 5.03$, $p = .027$, $\eta^2 = .050$): both language groups were more accurate in the Coarse Source Reporting condition ($M = 0.91$) than in the Specific Source Reporting Condition ($M = 0.86$). The same analysis for Event Changes also returned a main effect of Condition, ($F(1,96) = 19.26$, $p < .001$, $\eta^2 = .167$), with a similar advantage for the Coarse ($M = 0.96$) compared to the Specific Source Reporting condition ($M = 0.83$).

A similar ANOVA for Source Changes that additionally included Indirectness (High, Low) as a within-subjects factor returned a significant main effect of Indirectness ($F(1,96) = 5.44$, $p = .021$, $\eta^2 = .055$): for both language groups, accuracy was higher when the Indirectness was High ($M = 0.69$) compared to when it was Low ($M = 0.64$). No other effects or interactions were significant. Thus, even with additional training, participants did not become better at identifying the sources of their memories in the Specific Source Reporting compared to the Coarse Source Reporting condition. However, unlike in Experiment 2, Specific Source Reporting did not decrease accuracy for either types of Source Changes. Furthermore, even after this additional training, Turkish speakers did not perform any differently from English speakers in distinguishing inference from perception. The only factor that affected performance was the degree of indirectness of the event.

In order to test whether the training helped improve performance in Experiment 3, the overall accuracy in Experiment 3 was compared to the overall accuracy for those participants who received the Critical task first in Experiment 2 with a 2×2 between-subjects ANOVA with Condition (Coarse Source Reporting, Specific Source Reporting) and Experiment (2,3) as the factors. The analysis revealed an interaction between Source Reporting Condition and Experiment, $F(1,144) = 4.93$, $p = .028$, $\eta^2 = .033$. For the Specific Source Reporting condition, accuracy in Experiment 3 ($M = 0.78$) was higher than it was in Experiment 2 ($M = 0.70$, $t(72) = 2.31$, $p = .024$); however, for the Coarse Source Reporting condition, accuracy did not differ between Experiment 2 ($M = 0.86$) and 3 ($M = 0.85$, $t(72) = 0.47$, *ns*). Thus training helped improve performance but only in the Specific Source Reporting condition.

Error analysis

Performance in the Specific Source Monitoring condition was further assessed by comparing the frequency of choosing one vs. the other of the two incorrect response options. Table 6 presents the findings of the error analysis.

For No Changes, when the participants responded incorrectly, they were more likely to respond with “inferred” ($M = 0.70$) as opposed to “neither” ($M = 0.30$), $t(37) = 3.75$, $p = .001$. For Event Changes, when participants made an error, they were more likely to respond with “inferred” ($M = 0.82$) than “seen” ($M = 0.18$), $t(42) = 7.18$, $p < .001$. Including Language as a factor in these analyses

Table 6

Mean (SD) of errors to responses on the critical task in the specific source reporting condition (calculated over the total number of errors) (Experiment 3).

	No Change		Event Change		Source Change			
					High Indirectness		Low Indirectness	
	“Inferred”	“Not seen or inferred”	“Seen”	“Inferred”	“Seen”	“Not seen or inferred”	“Seen”	“Not seen or inferred”
English	0.69 (0.38)	0.31 (0.38)	0.21 (0.32)	0.78 (0.32)	0.71 (0.33)	0.28 (0.37)	0.78 (0.28)	0.22 (0.28)
Turkish	0.72 (0.30)	0.28 (0.30)	0.15 (0.26)	0.85 (0.26)	0.60 (0.33)	0.40 (0.33)	0.71 (0.28)	0.29 (0.28)

Table 7

Means (SD) of reaction times for accurate responses across types of change and language groups (in ms) (Experiment 3).

		No Change	Event Change	Source Change	
				High Indirectness	Low Indirectness
English	Coarse Source Reporting	1317.59 (368.76)	1210.43 (384.91)	1467.87 (516.08)	1513.3 (489.23)
	Specific Source Reporting	1414.60 (328.79)	1520.61 (290.08)	1772.52 (422.64)	1833.57 (427.50)
Turkish	Coarse Source Reporting	1291.28 (423.98)	1262.54 (355.25)	1637.95 (481.69)	1555.09 (551.16)
	Specific Source Reporting	1559.38 (335.98)	1664.28 (343.64)	2097.82 (729.07)	2055.76 (701.90)

did not reveal any significant effects or interactions, so it was omitted.

Crucially, for Source Changes we analyzed error patterns with a $2 \times 2 \times 2$ mixed ANOVA on the proportion of erroneous responses with Language (English, Turkish) as the between-subjects factor, and Error Type (“seen”, “neither”) and Indirectness (High, Low) as the within-subjects factors. The analysis returned only a main effect of Error Type ($F(1,39) = 29.50$, $p < .001$, $\eta^2 = .431$): participants were more likely to erroneously consider an inferred event as “seen” ($M = 0.70$) than as “neither seen not inferred” (i.e., a new event, $M = 0.31$). The pattern of errors for Source Changes, together with our previous results (and prior studies, such as Durso & Johnson, 1980; Johnson et al., 1977, 1979, 1984; Lindsay & Johnson, 1989), shows that people typically misattribute information derived from indirect sources to perception. Unlike Experiment 2, the bias to select “seen” responses was strongly present across both language groups and events of different Indirectness levels; there were no cross-linguistic differences in error patterns.

Reaction time data

Mean Reaction times for Experiment 3 are presented in Table 7. First, we analyzed the reaction times for accurate responses for No Changes with a 2×2 between-subjects ANOVA with Language (English, Turkish) and Condition (Coarse Source Reporting, Specific Source Reporting) as factors. The analysis revealed a main effect of Condition ($F(1,96) = 6.21$, $p = .014$, $\eta^2 = .061$): both language groups were slower in the Specific Source Reporting condition ($M = 1486.99$) as compared to the Coarse Source Reporting condition ($M = 1304.44$), perhaps because of the need to

select among three responses. The equivalent ANOVA for Event Changes also returned a main effect of Condition ($F(1,95) = 26.23$, $p < .001$, $\eta^2 = .216$), such that participants were slower in the Specific Source Reporting condition ($M = 1592.44$) than in the Coarse Source Reporting condition ($M = 1236.49$).

Finally, a similar analysis for Source Changes that also included Indirectness (High, Low) as a within-subjects factor revealed only a significant main effect of Condition ($F(1,93) = 16.34$, $p < .001$, $\eta^2 = .149$): both language groups were slower in the Specific Source Reporting condition ($M = 1939.92$) compared to the Coarse Source Reporting Condition ($M = 1543.56$). As in Experiment 2, even though the Indirectness of the events undergoing Source Changes had an effect on accuracy, it did not affect how fast participants responded in the source memory task.

Discussion

The findings of Experiment 3 support the finding from Experiment 2 that inferential reasoning leads to source monitoring errors. Overall, for 34% of the Source Changes, speakers of English and Turkish reported having seen an event that they had only inferred from post-event visual evidence. Moreover, we replicated the finding that the degree to which post-event visual information is indirect affects the likelihood of source monitoring errors: such errors were highest when the participants were initially presented with post-event inferential evidence that was somewhat closer to direct visual evidence. Crucially, as in Experiment 2, the likelihood of source monitoring errors remained the same regardless of the participants’ native language: neither accuracy nor reaction time data provided

any support for increased sensitivity to sources in Turkish speakers compared to English speakers. Importantly, unlike Experiment 2, there were no language-specific patterns in error distributions. If such patterns were the product of language-specific source monitoring processes in the earlier experiment, they would be unlikely to disappear in the present version of the same task that simply had clearer instructions and training. We therefore conclude that the observed differences in Experiment 2 were probably due to some other factor related to task interpretation in the Turkish group: once clarification was provided about the nature of the task and the response options, the difference between language groups disappeared. In support of this possibility, we note that in Experiment 3, unlike the earlier study, there were no differences in reaction times between the two language groups, presumably as a result of clearer task demands.

Across language groups in Experiment 3, orienting participants to the sources of their memories by asking them to specifically report whether they had inferred or seen an event did not increase their sensitivity to sources. As in Experiment 2, making specific source attributions was more challenging for participants for all types of changes, as indicated by longer reaction times for responses. However, unlike in Experiment 2, these explicit source judgments did not increase source monitoring errors with inference. Furthermore, performance on Specific Source Reporting improved between Experiments 2 and 3. This improvement in performance might have washed away the negative effect of specific source judgments on accuracy, especially for less indirect events undergoing Source Changes.

Before we conclude, we need to address a potential objection to the above interpretation of our findings. One might hypothesize that the high rate of errors for source changes is not specific to source attributions but is rather due to the fact that the test items for source changes (the “seen” events) were not visually very distinct from the original stimuli (the “inferred” events). To address this possibility, we tested a new group of English speakers ($n = 20$) using the exact procedure of the Coarse Source Reporting in Experiment 3 with a single change to the Source Change items: participants originally studied “seen” versions of Source Changes, and were tested on the “inferred” versions in the test phase. If the high rate of errors for Source Changes is due to limited visual discriminability between seen and inferred versions of an event, then error rates for source changes should stay the same. However, if the observed error rates are specific to memory for sources (specifically, to overattribution of memories to perception), then error rates should diminish. The results from this control task supported the second possibility. Accuracy was much higher in participants who were tested with seen-to-inferred Source Changes ($M = 0.82$) compared to those who were tested with inferred-to-seen Source Changes ($M = 0.68$), $F(1,43) = 7.14$, $p < .001$, $\eta^2 = .142$. These results demonstrate that low accuracy for Source Changes is not simply due to the visual similarity between study and test items. Together with the error data, these findings show that people are much more likely to attribute an internally generated

representation (here, an inference) to perception as opposed to the other way around (see also Durso & Johnson, 1980; Johnson et al., 1977, 1979, 1984; Lindsay & Johnson, 1989).

General discussion

Languages differ in the way they encode information source, with some languages (e.g., Turkish) encoding evidentiality through obligatory grammatical morphemes and other languages (e.g., English) encoding evidentiality sporadically (and mostly lexically). In the experimental studies reported here, we tested two competing theoretical hypotheses about the relation between evidentiality and source monitoring. According to one hypothesis, evidentiality markers could enhance memory for sources in Turkish compared to English speakers by allowing people to recall or reconstruct source information based on the systematic differentiation of those sources at the linguistic level (Boroditsky, 2003; Imai & Gentner, 1997; Levinson, 2003; Lucy, 1992; Whorf, 1956). According to another hypothesis, evidential language might not produce deep attention-driven reorganization of source-monitoring processes (Gleitman & Papafragou, 2005, 2012; Landau et al., 2010; Papafragou et al., 2002). These competing hypotheses connect to a broader debate about the interface between language and thought (see references above). We began by exploring the cross-linguistic differences in evidential encoding between English and Turkish, paying special attention to the conditions under which the indirect evidential in Turkish is used to mark post-event inference; we also related the way Turkish speakers marked the perception-inference division in their language to explicit assessments of the same distinction in English speakers (Experiment 1). We then asked whether Turkish and English speakers differ in the speed and accuracy with which they recognize inference (as opposed to visual perception) as the source of their event memories, using different types of source monitoring cognitive tasks (Experiments 2 and 3). We concentrated on inference-to-perception source changes because prior studies (with English-speaking participants) had indicated that products of internal processes such as inference or imagination were more likely to be misattributed to perception than the other way around, hence these cases were highly likely to produce source monitoring errors (e.g., Brewer, 1977; Chan & McDermott, 2006; Durso & Johnson, 1980; Fazio & Marsh, 2010; Hannigan & Reinitz, 2001; Harris, 1974; Harris & Monaco, 1978; Johnson et al., 1973, 1977, 1979, 1984; Lindsay & Johnson, 1989; McDermott & Chan, 2006).

Cross-linguistic encoding of evidentiality

Our findings from Experiment 1 confirmed the presence of a strong cross-linguistic difference in how evidentiality is encoded: Turkish speakers—but not English speakers—encoded sources of events in ordinary event descriptions. Specifically, Turkish speakers used the direct past tense marker (*-dı*) for events they had seen and the indirect past tense marker (*-miş*) for events they had inferred on the

basis of visual evidence (see also Aksu & Slobin, 1986; Aksu-Koç, 1988, 2000; Aksu-Koç et al., 2009; Ozturk & Papafragou, *in press*; Ünal & Papafragou, *submitted for publication*). Moreover, a closer look at Turkish speakers' linguistic encoding of sources of inferred events revealed previously unnoticed differences in the extent to which they used the indirect evidential marking that depended on the quality of visual evidence. Specifically, Turkish speakers reserved the indirect evidential marking for highly indirect events yielding less secure inferences, but did not have a systematic preference between the two evidential forms for less indirect events for which the visual evidence made it highly likely that the event happened. These results establish that evidentiality is a good testbed for investigating the question of whether language might affect event source memory and motivate our later experiments.

Interestingly, Experiment 1 also revealed similarities across people of different language backgrounds: use of evidential marking in Turkish closely corresponds to English speakers' judgments about whether an event was seen or inferred when no memory task was involved. The events that were overwhelmingly marked by the direct past tense in Turkish were also judged to be "seen" by English speakers, and those marked by the indirect past tense were judged to be "inferred". Likewise, events that were equally likely to be marked with the direct or indirect past tense morphemes by Turkish speakers were also equally likely to receive judgments of "seen" and "inferred" by English speakers. Thus, subtleties in the use of grammaticalized evidentials map onto shared, potentially universal distinctions between perception and inference.

Cross-linguistic differences and event source memory

Our source monitoring tasks in Experiments 2 and 3 revealed massive source monitoring errors in distinguishing inference from perception in event memory. When adults were provided with photographs of end states of events that would lead them to inferentially reconstruct the event and were later given a memory task, they reported having seen the point at which the event unfolded 35% of the time, i.e., they committed source-monitoring errors. Furthermore, two patterns provide evidence that these errors were specific to memory for sources, rather than a general memory limitation. First, in Experiment 2, our participants responded differently to items that involved changes to objects and events compared to items that involved changes to sources. That is, participants were more accurate in identifying whether they had seen a man reaching for apples or pears in a bowl, compared to when they had to identify whether they had seen a boy tear a paper towel or inferred that the boy had torn the paper towel. Moreover, in a control study reported in Experiment 3, our participants committed fewer errors when they originally studied the point at which an event unfolded and were tested on the end state of the same events. That shows that source monitoring errors with inference cannot simply be attributed to visual similarity between study and test items for events undergoing source changes. Our results are consistent with past work showing that people tend to attribute memories generated from internal processes

to perception (e.g., Brewer, 1977; Chan & McDermott, 2006; Durso & Johnson, 1980; Fazio & Marsh, 2010; Harris, 1974; Harris & Monaco, 1978; Johnson et al., 1973, 1977, 1979, 1984; Lindsay & Johnson, 1989; McDermott & Chan, 2006). Furthermore, our results extend prior work on post-event inference (Hannigan & Reinitz, 2001; Strickland & Keil, 2011) by showing that these errors can occur even after limited (single) exposures to post-event visual evidence.

Crucially, however, in both of our experiments, speakers of English and Turkish were equally prone to errors in identifying the sources of their memories. Moreover, the time course of the recovery of source information offered no support for the idea that Turkish speakers have an advantage in source monitoring compared to English speakers. There was a difference between the two language groups in the kinds of errors observed in the Specific Source Reporting condition of Experiment 2: even though, for Low-Indirectness events, both language groups were equally likely to make a certain type of error (i.e., to select "seen" over "not seen or inferred"), for High-Indirectness events, English speakers made more "seen" errors compared to Turkish speakers and Turkish speakers made more "not seen or inferred" errors compared to English speakers. This difference is not easily attributed to effects of evidential language encoding on memory, since such effects could not explain why Turkish speakers were more likely to treat source changes as event changes. Additionally, these differences disappeared when the task instructions (especially the nature of the three choices in the Specific Source Reporting condition and their mapping onto the response keys) were clarified: in Experiment 3, both language groups were equally likely to mischaracterize previously inferred events as "seen" regardless of how indirect the original event was. Furthermore, both language groups were equally fast to respond. Since the task in Experiment 3 offers the strongest and clearest test of source monitoring in the present battery, we conclude that the presence or absence of grammaticalized/obligatory source encoding in one's native language does not increase sensitivity to the distinction between sources in event memory. It is worth noting that the absence of a cognitive difference between two linguistic populations in our studies is not simply a null finding, since it occurs against the backdrop of deep and systematic cross-linguistic differences in encoding event sources for the very same stimuli (cf. Experiment 1).

Our findings seem to be at odds with the findings of a recent cross-linguistic study (Tosun et al., 2013) comparing English and Turkish speakers' memories for information presented in first-hand vs. non-first-hand forms. That study revealed that both language groups had equally accurate memories for information presented in first-hand form, but Turkish monolinguals and Turkish-English bilinguals had lower memory accuracy for information presented in non-first-hand forms. The discrepancy between Tosun et al.'s findings and ours can be attributed to a number of factors. First, as mentioned already, the stimuli used for English and Turkish speakers in Tosun et al.'s studies were not equivalent. English speakers had to report the presence or the absence of an evidential adverb, whereas Turkish speakers had to remember which

of the two evidential morphemes was included in the sentence—which might be harder than remembering lexical items. The difference in stimuli might account for the cross-linguistic differences observed by Tosun et al. Our findings suggest that Turkish speakers have no such disadvantage when remembering event sources when the two language groups are tested with the exact same *nonlinguistic* stimuli/task. Second, it is not clear whether the English and Turkish samples were comparable in Tosun et al.'s studies since there was no independent measure of equivalence. Our English and Turkish speakers are equivalent in terms of general memory accuracy (even though Turkish speakers were slower to respond sometimes and, as we hypothesized, may have been more uncertain about the task). In addition to these methodological issues, a critical difference between Tosun et al.'s and our studies was the kind of language effect that was under investigation. While Tosun et al. were interested in whether explicit linguistic framing affects memory, our studies tested whether habitual differences in how English and Turkish speakers encode event sources in language affects event source memory.

One might wonder whether the absence of cross-linguistic differences might be due to an overall difficulty or a lack of sensitivity of our measures of source memory. Two pieces of evidence suggest that this is unlikely. First, our English and Turkish speakers performed above chance level in both Coarse and Source Reporting conditions in both Experiments 2 and 3, suggesting that the task was not particularly difficult. Second, the error rates that we obtained from our participants (35%) is similar to the 30% error rate obtained in prior studies from English speakers (cf. Belli, Lindsay, Gales, & McCarthy, 1994; Intraub & Hoffman, 1992; McCloskey & Zaragoza, 1985). The overall performance in our task was much better compared to Tosun et al.'s studies where the overall accuracy rates were between 20% and 40%.

Similarities in source monitoring across language communities

Further support for the presence of shared source monitoring mechanisms across members of different language communities comes from two unexpected commonalities in source reasoning patterns in English and Turkish speakers. First, for both language groups and across both Experiments 2 and 3, memories for Low Indirectness events (that prompted less consistent use of the Turkish indirect past tense and elicited mixed judgments of being “seen” vs. “inferred” by English speakers in Experiment 1) were harder to discriminate from perception compared to High Indirectness events. Second, across language groups, the tendency to overattribute inferred events to perception persisted when memory for sources was tested with a more detailed source monitoring task as opposed to a simple recognition task. In Experiment 2, across both language groups, participants were less accurate and slower when reporting the exact sources of their memories for less indirect events that had undergone a source change than when they were simply reporting whether or not they had seen the events. In Experiment 3, performance in specific source

reporting improved compared to Experiment 2 after participants were trained to distinguish perception from inference. Nevertheless, it still was no better/faster than performance in coarse source reporting, a fact indicating that asking participants to make explicit source judgments did not increase their attention to the sources of their memories in our study.

Both of these patterns cohere with and extend prior literature on the circumstances under which source monitoring succeeds and fails. Beginning with the effects of Indirectness, we know that the accuracy of source monitoring decisions is affected by the similarity between memories whose sources need to be discriminated (Bransford & Johnson, 1973; Lindsay et al., 1991; Mather et al., 1999). Thus, inference from visual evidence is not qualitatively different from direct visual evidence, but instead, post-event decisions about whether an event was perceived or inferred depend on the quality of evidence about the event available in memory. For Low Indirectness events, inferences were based on higher quality post-event evidence and probably yielded event memory representations that shared several characteristics with the schematic characteristics of direct visual evidence. Because a memory is attributed to a particular source based on how well the qualities of the memory representation and that source match (Durso & Johnson, 1980; Johnson et al., 1993), source misattributions were especially high for these events. By contrast, for High Indirectness events, for which less secure inferences were made based on lower quality post-event visual evidence, the distance between what was seen and what was inferred was greater (such that one might have a different representation of the event by inferentially reconstructing it than one would have by actually seeing it). Since for such events the amount of overlap between the qualities of the event memory representation and qualities of direct visual evidence was smaller, the likelihood of source misattributions decreased.

Turning to task effects, our data differ from prior studies showing that the likelihood of source misattributions can be reduced if participants are asked to provide specific information about the sources of their memories as opposed to performing a recognition task (e.g., Hayes-Roth & Thorndyke, 1979; Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989). We believe that the difference is due to the fact that the current study focused on two memory sources that share several characteristics, namely visual perception and inference from visual cues. In the current study, the event information that was acquired through inferential reasoning involved the same agent and the same action as the event information later presented visually. More generally, our stimuli involved veridical inferences from perceptual outcomes, where there was little distance between what was seen and what was supplied by inference. Since inference from visual evidence and visual perception in our stimuli delivered highly similar information in terms of semantic content, the manipulations in the specific source reporting condition may not have been powerful enough to reduce misattributions of inferred events to perception. By contrast, in past studies the distance between sources has typically been

greater. For instance, several studies investigated the relationship between visual memories of scenes and misleading information about the scenes embedded in a written text: in these studies, the distance between perceiving a scene and reading about a scene is greater than in the present studies (Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989). Specifically, in those earlier studies, the features that characterized the source of the original information were visual, and the features that characterized the source of the misleading information were verbal. Not only were these two sources different from each other in terms of their perceptual characteristics, but the information delivered from each source contradicted the other: the misleading text claimed the existence of objects that were not actually present in the initial scene. Thus, the participants could potentially benefit from the differences between the perceptual characteristics of the sources themselves, as well as the content of the information derived from each source when making source monitoring decisions. In the case of accessing the same event from visual perception and inference from visual evidence, as was the case in the present study, both sources were essentially composed of visual attributes and delivered similar information about the event, making source confusions more likely (cf. also Harris, 1974).

Conclusions: evidence, language and cognition

Together, our findings provide evidence against the position that language exerts strong, stable effects on source monitoring and on cognition more generally (e.g., Boroditsky, 2003; Bowerman & Choi, 2001; Imai & Gentner, 1997; Levinson, 2003; Lucy, 1992; Whorf, 1956). Rather, these data cohere with a long body of literature suggesting that cross-linguistic differences in event encoding do not necessarily lead to differences in how events are processed non-linguistically (see Gennari et al., 2002; Landau et al., 2010; Li et al., 2009; Munnich, Landau, & Doshier, 2001; Papafragou et al., 2002). Together with the commonalities in the way members of different language communities handle evidence and link it to event knowledge, our results offer support to the conclusion that language, to a considerable degree, reflects shared, perhaps universal conceptual representations (Bloom, 1994; Chomsky, 1975; Gleitman & Papafragou, 2005, 2012; Pinker, 1994).

Although these findings seem to conflict with prior work on the relation between evidentiality and source monitoring in Turkish (e.g., Aksu-Koç et al., 2009; Tosun et al., 2013), none of these studies directly compared two language populations on a truly *non-linguistic* measure. Our findings are consistent with a developmental study that investigated whether cross-linguistic differences in the encoding of evidentiality might affect the timetable of source monitoring development in young children (Papafragou et al., 2007). That study assessed whether 3- and 4-year-old learners of Korean, a language with grammaticalized evidentiality, would become able to accurately monitor the sources of their beliefs earlier than learners of English, and found no such differences between English and Korean learners. Just as the development of

source monitoring seems to proceed identically across children in such communities, the engagement of source monitoring mechanisms seems to proceed similarly across adults speaking languages with and without grammaticalized evidentiality.

Acknowledgments

This work has been supported by NSF Grant BCS0749870 to A. Papafragou. A. Pinto acknowledges support from a University of Delaware Undergraduate Supply and Expense Grant, a University of Delaware Undergraduate Summer Fellowship Grant, and a Psi Chi Society Summer Research Grant. We thank Myeongyun Choi, Alison McQuiston, Allison Witrock and Angela Yung for assistance with data collection, and Aylin Küntay and the Language and Communication Development Lab at Koç University in Istanbul-Turkey for hosting us and facilitating data collection for the Turkish portion of the experiments.

Appendix A. List of target events

	Target events
1	A woman pushed a chair
2	A man blew the candles on a cake
3	A woman peeled a clementine
4	A man kicked a garbage bin
5	A man opened a jar
6	A woman ate a slice of pizza
7	A man tore a paper towel
8	A woman wrapped a present
9	A woman braided her hair
10	A woman poured coffee into a cup
11	A woman cracked an egg
12	A woman knit a sweater
13	A man whisked eggs
14	A woman blew bubbles
15	A man cracked nuts
16	A woman rolled dough

References

- Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford, United Kingdom: Oxford University Press.
- Aikhenvald, A. Y. (2014). The grammar of knowledge: A cross-linguistic view of evidentials and the expression of information source. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *The grammar of knowledge: A cross-linguistic typology* (pp. 1–51). Oxford, United Kingdom: Oxford University Press.
- Aikhenvald, A., & Dixon, R. (Eds.). (2001). *Studies in evidentiality*. Amsterdam, The Netherlands: John Benjamins.
- Aksu, A., & Slobin, D. (1986). A psychological account of the development and use of evidentials in Turkish. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 159–167). Norwood, NJ: Ablex.
- Aksu-Koç, A. (1988). *The acquisition of aspect and modality: The case of past reference in Turkish*. Cambridge, United Kingdom: Cambridge University Press.
- Aksu-Koç, A. (2000). Some aspects of the acquisition of evidentials in Turkish. In L. Johanson & B. Utas (Eds.), *Evidentials: Turkic, Iranian and neighbouring languages* (pp. 15–28). Berlin, Germany: de Gruyter.
- Aksu-Koç, A., Ögel-Balaban, H., & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development*, 125, 13–28. <http://dx.doi.org/10.1002/cd.247>.

- Anderson, R. E. (1984). Did I do it or did I only imagine doing it? *Journal of Experimental Psychology: General*, 113, 594–613. <http://dx.doi.org/10.1037/0096-3445.113.4.594>.
- Anderson, L. (1986). Evidentials, paths of change and mental maps: Typologically regular asymmetries. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 273–312). Norwood, NJ: Ablex.
- Belli, R. F., Lindsay, S. D., Gales, M. S., & McCarthy, T. T. (1994). Memory impairment and source misattribution experiments with short retention intervals. *Memory & Cognition*, 22, 40–54. <http://dx.doi.org/10.3758/BF03202760>.
- Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua*, 92, 297–329. [http://dx.doi.org/10.1016/0024-3841\(94\)90345-X](http://dx.doi.org/10.1016/0024-3841(94)90345-X).
- Boroditsky, L. (2003). Linguistic relativity. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 917–921). London, United Kingdom: MacMillan Press.
- Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge, United Kingdom: Cambridge University Press.
- Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331–350. [http://dx.doi.org/10.1016/0010-0285\(71\)90019-3](http://dx.doi.org/10.1016/0010-0285(71)90019-3).
- Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. Chase (Ed.), *Visual information processing* (pp. 383–438). New York, NY: Academic Press.
- Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition*, 5, 673–678. <http://dx.doi.org/10.3758/BF03197414>.
- Chafe, W. L., & Nichols, J. (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.
- Chan, J. C. K., & McDermott, K. B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology*, 20, 633–639. <http://dx.doi.org/10.1002/acp.1215>.
- Choi, S. (1995). The development of epistemic sentence-ending modal forms and functions in Korean children. In J. Bybee & S. Fleischman (Eds.), *Modality in grammar and discourse* (pp. 165–204). Amsterdam, The Netherlands: John Benjamins.
- Chomsky, N. (1975). *Reflections on language*. New York, NY: Pantheon.
- Dallett, K., & Wilcox, S. G. (1968). Remembering pictures vs. remembering descriptions. *Psychological Science*, 11, 139–140. <http://dx.doi.org/10.3758/BF03331012>.
- De Haan, F. (2001). The place of inference within the evidential system. *International Journal of American Linguistics*, 67, 193–219.
- DeLancey, S. (2001). The mirative and evidentiality. *Journal of Pragmatics*, 33, 369–382. [http://dx.doi.org/10.1016/S0378-2166\(01\)80001-1](http://dx.doi.org/10.1016/S0378-2166(01)80001-1).
- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, 24, 613–621. <http://dx.doi.org/10.1177/0956797612457374>.
- Drumme, A. B., & Newcombe, N. S. (2002). Developmental changes in source memory. *Developmental Science*, 5, 502–513. <http://dx.doi.org/10.1111/1467-7687.00243>.
- Durso, F. T., & Johnson, M. K. (1980). The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, 19, 416–429. [http://dx.doi.org/10.1016/S0022-5371\(80\)90294-7](http://dx.doi.org/10.1016/S0022-5371(80)90294-7).
- Faller, M. (2001). Remarks on evidential hierarchies. In D. Beaver, S. Kaufmann, B. Clark, & L. Casillas (Eds.), *Proceedings of the 'Semfest'* (pp. 37–59). Stanford, CA: CSLI Publications.
- Fazio, L. K., & Marsh, E. J. (2010). Correcting false memories. *Psychological Science*, 21, 801–803. <http://dx.doi.org/10.1177/0956797610371341>.
- Fitneva, S. (2009). Evidentiality and trust: The effect of informational goals. *New Directions for Child and Adolescent Development*, 125, 49–62. <http://dx.doi.org/10.1002/cd.249>.
- Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, 64, 74–92. <http://dx.doi.org/10.1016/j.cogpsych.2011.10.004>.
- Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, 83, 49–79. [http://dx.doi.org/10.1016/S0010-0277\(01\)00166-4](http://dx.doi.org/10.1016/S0010-0277(01)00166-4).
- Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.
- Givón, T. (1982). Evidentiality and epistemic state. *Studies in language*, 6, 23–49. <http://dx.doi.org/10.1075/sl.6.1.03giv>.
- Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633–661). Cambridge, United Kingdom: Cambridge University Press.
- Gleitman, L., & Papafragou, A. (2012). New perspectives on language and thought. In K. Holyoak & R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (2nd ed., pp. 543–568). New York, NY: Oxford University Press.
- Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, 59, 1366–1371. <http://dx.doi.org/10.1111/1467-8624.ep8589535>.
- Gumperz, J. J., & Levinson, S. C. (1996). *Rethinking linguistic relativity*. Cambridge, United Kingdom: Cambridge University Press.
- Hannigan, S. L., & Reinitz, M. T. (2001). A demonstration and comparison of two types of inference-based memory errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 931–940. <http://dx.doi.org/10.1037/0278-7393.27.4.931>.
- Harris, R. J. (1974). Memory and comprehension of implications and inferences of complex sentences. *Journal of Verbal Learning and Verbal Behavior*, 13, 626–637. [http://dx.doi.org/10.1016/S0022-5371\(74\)80050-2](http://dx.doi.org/10.1016/S0022-5371(74)80050-2).
- Harris, R. J., & Monaco, G. E. (1978). Psychology of pragmatic implication: Information processing between the lines. *Journal of Experimental Psychology: General*, 107, 1–22. <http://dx.doi.org/10.1037/0096-3445.107.1.1>.
- Hart, W., & Albarricín, D. (2011). Learning about what others were doing: Verb aspect and attributions of mundane and criminal intent for past actions. *Psychological Science*, 22, 261–266. <http://dx.doi.org/10.1177/0956797610395393>.
- Hayes-Roth, B., & Thorndyke, P. W. (1979). Integration of knowledge from texts. *Journal of Verbal Learning and Verbal Behavior*, 18, 91–108. [http://dx.doi.org/10.1016/S0022-5371\(79\)90594-2](http://dx.doi.org/10.1016/S0022-5371(79)90594-2).
- Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62, 169–200. [http://dx.doi.org/10.1016/S0010-0277\(96\)00784-6](http://dx.doi.org/10.1016/S0010-0277(96)00784-6).
- Intraub, H., & Hoffman, J. E. (1992). Reading and visual memory: Remembering scenes that were never seen. *American Journal of Psychology*, 105, 101–114. <http://dx.doi.org/10.2307/1422983>.
- Johanson, L. (2003). Evidentiality in Turkic. In Alexandra Y. Aikhenvald & R. M. W. Dixon (Eds.), *Studies in evidentiality (Typological studies in language 54)* (pp. 273–290). Amsterdam, The Netherlands & Philadelphia, PA: John Benjamins.
- Johnson, M. K. (1997). Source monitoring and memory distortion. *Philosophical Transactions: Biological Sciences*, 352, 1733–1745. <http://dx.doi.org/10.1098/rstb.1997.0156>.
- Johnson, M. K. (2006). Memory and reality. *American Psychologist*, 61, 760–771. <http://dx.doi.org/10.1037/0003-066X.61.8.760>.
- Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology*, 98, 203–205.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3–28. <http://dx.doi.org/10.1037/0033-2909.114.1.3>.
- Johnson, M. K., Kahan, T. L., & Raye, C. L. (1984). Dreams and reality monitoring. *Journal of Experimental Psychology: General*, 113, 329–344. <http://dx.doi.org/10.1037/0096-3445.113.3.329>.
- Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1409–1419. <http://dx.doi.org/10.1037/0278-7393.20.6.1409>.
- Johnson, M. K. (1988). Discriminating the origin of information. In T. F. Oltmanns & B. A. Maher (Eds.), *Delusional beliefs* (pp. 34–65). New York: Wiley.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67–85. <http://dx.doi.org/10.1037/0033-295X.88.1.67>.
- Johnson, M. K., Raye, C. L., Foley, M. A., & Kim, J. K. (1982). Pictures and images: Spatial and temporal information compared. *Bulletin of the Psychonomic Society*, 19, 23–26. <http://dx.doi.org/10.3758/BF03330029>.
- Johnson, M. K., Raye, C. L., Wang, A., & Taylor, T. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 229–240. <http://dx.doi.org/10.1037/0278-7393.5.3.229>.
- Johnson, M. K., Taylor, T., & Raye, C. L. (1977). Fact and fantasy: The effects of externally generated events on the apparent frequency of externally generated events. *Memory & Cognition*, 5, 115–122. <http://dx.doi.org/10.3758/BF03209202>.

- Kornfilt, J. (1997). *Turkish*. London, United Kingdom: Routledge.
- Landau, B., Dessalegn, B., & Goldberg, A. M. (2010). Language and space: Momentary interactions. In P. Chilton & V. Evans (Eds.), *Language, cognition, and space: The state of the art and new directions. Advances in cognitive linguistics series* (pp. 51–78). London, United Kingdom: Equinox Publishing.
- Levinson, S. C. (2003). Language in mind: Let's get the issues straight! In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the issues of language and thought* (pp. 25–46). Cambridge, MA: MIT Press.
- Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology*, 58, 487–524. <http://dx.doi.org/10.1016/j.cogpsych.2008.12.001>.
- Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, 17, 349–358. <http://dx.doi.org/10.3758/BF03198473>.
- Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, 52, 297–318. [http://dx.doi.org/10.1016/0022-0965\(91\)90065-Z](http://dx.doi.org/10.1016/0022-0965(91)90065-Z).
- Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge, United Kingdom: Cambridge University Press.
- Lyons, K. E., Ghetti, S., & Cornoldi, C. (2010). Age differences in the contribution of recollection and familiarity to false-memory formation: A new paradigm to examine developmental reversals. *Developmental Science*, 13, 355–362. <http://dx.doi.org/10.1111/j.1467-7687.2009.00889.x>.
- Mather, M., Johnson, M. K., & De Leonardi, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, 16, 437–458. <http://dx.doi.org/10.1080/026432999380870>.
- Matsui, T. (2014). Children's understanding of linguistic expressions of certainty and evidentiality. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 295–316). Amsterdam, The Netherlands: John Benjamins.
- McCloskey, M., & Zaragoza, M. (1985). Misleading post event information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114, 1–16. <http://dx.doi.org/10.1037/0096-3445.114.1.1>.
- McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory & Cognition*, 34, 1273–1284. <http://dx.doi.org/10.3758/BF03193271>.
- Munnich, E., Landau, B., & Doshier, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition*, 81, 171–207. [http://dx.doi.org/10.1016/S0010-0277\(01\)00127-5](http://dx.doi.org/10.1016/S0010-0277(01)00127-5).
- Ozturk, O., & Papafragou, A. (in press). The acquisition of evidentiality and source monitoring. *Language Learning and Development*.
- Papafragou, A., Hulbert, J., & Trueswell, J. C. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108, 155–184. <http://dx.doi.org/10.1016/j.cognition.2008.02.007>.
- Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *Cognition*, 103, 253–299. <http://dx.doi.org/10.1016/j.cognition.2006.04.001>.
- Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, rattle 'n' roll: The representations of motion in language and cognition. *Cognition*, 84, 189–219. [http://dx.doi.org/10.1016/S0010-0277\(02\)00046-X](http://dx.doi.org/10.1016/S0010-0277(02)00046-X).
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York, NY: Morrow.
- Robinson, E. J. (2009). Commentary: What we can learn from research on evidentials. *New Directions for Child and Adolescent Development*, 125, 95–103. <http://dx.doi.org/10.1002/cd.252>.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-prime reference guide*. Pittsburgh, PA: Psychology Software Tools.
- Slobin, D., & Aksu, A. (1982). Tense, aspect, modality, and more in Turkish evidentials. In P. Hopper (Ed.), *Tense-aspect: Between semantics and pragmatics* (pp. 185–200). Amsterdam, The Netherlands: John Benjamins.
- Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language*, 33, 203–217. <http://dx.doi.org/10.1006/jmla.1994.1010>.
- Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition*, 121, 409–415. <http://dx.doi.org/10.1016/j.cognition.2011.04.007>.
- Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *Journal of Memory and Language*, 69, 121–134. <http://dx.doi.org/10.1016/j.jml.2013.03.004>.
- Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language*, 63, 64–82. <http://dx.doi.org/10.1016/j.jml.2010.02.006>.
- Ünal, E., & Papafragou, A. (submitted for publication). Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language* (Special Issue on the Relationship between Language Production and Comprehension).
- Ünal, E., & Papafragou, A. (2013). Linguistic and conceptual representations of inference as a knowledge source. In *Proceedings from the 37th annual Boston University conference on language development* (pp. 433–443). Somerville, MA: Cascadia Press.
- de Villiers, J. G., Garfield, J., Gernet-Girard, H., Roeper, T., & Spears, M. (2009). Evidentials in Tibetan: Acquisition, semantics, and cognitive development. *New Directions for Child and Adolescent Development*, 125, 29–47. <http://dx.doi.org/10.1002/cd.248>.
- Whorf, B. L. (1956). Language, thought and reality. In J. B. Carroll (Ed.), *Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.
- Willet, T. (1988). A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language*, 12, 51–97. <http://dx.doi.org/10.1075/sl.12.1.04wil>.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104, 7780–7785. <http://dx.doi.org/10.1073/pnas.0701644104>.
- Wolff, P. M., & Holmes, K. J. (2011). Linguistic relativity. *WIREs. Cognitive Science*, 2, 253–265. <http://dx.doi.org/10.1002/wcs.104>.
- Zaragoza, M. S., & Koshmider, J. W. I. I. (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 246–255. <http://dx.doi.org/10.1037/0278-7393.15.2.246>.