THE LANGUAGE-COGNITION INTERFACE:

EVIDENCE FROM THE DOMAIN OF EVIDENTIALITY

by

Ercenur Ünal

A dissertation submitted to the College of Arts and Sciences in partial
fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

Spring 2016

THE LANGUAGE-COGNITION INTERFACE:

EVIDENCE FROM THE DOMAIN OF EVIDENTIALITY

by

Ercenur Ünal

Approved: ______________________________
Robert F. Simons, Ph.D.
Chair of the Department of Psychological and Brain Sciences

Approved: ______________________________
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: ______________________________
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Anna Papafragou, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Helene Intraub, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Paul C. Quinn, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Brian P. Ackerman, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Roberta M. Golinkoff, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ____

Satoshi Tomioka, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I owe my deepest gratitude to my advisor Anna Papafragou for her generosity with her time, patiently teaching me how to think like a scientist, and all the guidance and support she provided over the past five years. Thanks to my committee members Helene Intraub, Paul Quinn, Brian Ackerman, Roberta Golinkoff and Satoshi Tomioka for their insightful feedback and thoughtful comments. All the tough questions they have asked throughout the process deepened my thinking and made the final state of this work stronger.

I extend special thanks to my undergraduate and master's advisor Aylin Küntay for encouraging me to pursue a career in academia. My interest in research, language and development started because of the intellectual stimulation she provided in her classes and laboratory. I am grateful for having her continuing support for the past ten years.

Thanks to the National Science Foundation (grant BCS0749870) for funding these studies. Thanks also to the Office of Graduate and Professional Education for the Graduate Fellowship that they awarded me and the Department of Psychological and Brain Sciences for their support in the form of a Dissertation Award.

Thanks to Adrienne Pinto, Alison McQuiston, Allison Witrock, Alyssa Kampa, Angela Yung, Didem Aykut, Taylor Dino, Taylor Maher, Sam Katz, Myeongyun Choi, Victoria Dillon, Yasemin Sandıkçı, and Yue Ji for their assistance with data collection and stimulus preparation.

Thanks to members of the Language and Communication Development Lab at Koç University for hosting me several times over the past five years and facilitating data collection for my studies. Thanks also to Koç University Department of Psychology for their generous support by allowing me to use their subject pool.

Thanks to the Early Learning Center at UD and Bianca Graves, and the preschools in Istanbul, Turkey and Newark, DE for helping me recruit children to participate in my studies. Thanks also to the hundreds of children and adults who participated in my studies.

Thanks to the present and past members of the Language and Cognition Lab for providing feedback and helpful discussions. Special thanks to Ann Bunger, Dimitris Skordos, Myrto Grigoroglou and Angeliki Athanasopoulou for being amazing colleagues and friends.

Thanks to my dear friends all over the world, especially Deniz, Yüksel, Eda, Berk, Ceren and Begüm, for providing love and support, and making distances and time zones not count. Also, thanks to all my friends who participated in my studies over the last five years.

Mom and dad, I have no words to express how grateful I am to have you as my parents. Without your unconditional love and endless support, I would not be able to do any of this.

Finally, I would like to dedicate my dissertation to my grandmother, Asuman Orhon who passed while I was writing my dissertation. She was a teacher and the way she guided her students will always inspire me.

PREFACE

This dissertation includes the dissertation author's previously published work that was completed at the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychological and Brain Sciences. Chapter 2 appeared as Ünal, E., Pinto, A., Bunger, A. & Papafragou, A. (2016). Monitoring sources of event memories: A cross-linguistic investigation. *Journal of Memory and Language, 87,* 157-176. doi: 10.1016/j.jml.2015.10.009. A modified version of Chapter 3 appeared as Ünal, E., & Papafragou, A. (2016) Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language, Special Issue on Speaking and Listening*. doi: 10.1016/j.jml.2015.12.001. The dissertation author is the first and corresponding author on both publications. For both publications, the dissertation author designed the experiments, collected the data, performed the statistical analyses and wrote the manuscripts. Reprint permissions for these publications are provided in Appendix E and F.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Languages vary in the way they encode different aspects of the world. Could such cross-linguistic differences affect how the world is represented in the minds of the speakers of different languages? And how do linguistic and conceptual representations make contact during language acquisition? This dissertation investigates the language-cognition interface cross-linguistically and developmentally with the goal of understanding which aspects of cognition are shared across speakers of different languages, and which aspects may be susceptible to linguistic influences.

The empirical focus of this dissertation is the relationship between *evidentiality* (the linguistic encoding of information source) and non-linguistic representations of *sources of information*. Languages like English offer speakers the option to differentiate the information gained through different sources, although encoding evidential distinctions is not obligatory. By contrast some languages such as Turkish mark evidential distinctions in their grammatical systems. From a cross-linguistic perspective, this dissertation asks whether differences in the way languages encode information sources affects source monitoring mechanisms in mature cognizers (Chapter 2). From a developmental perspective, this dissertation examines how linguistic evidentiality is acquired by learners of Turkish (Chapter 3), and how its acquisition is related to the development of non-linguistic representations of information sources (Chapters 3 and 4).

The results of the present investigation inform our understanding of the language-cognition interface in two ways. First, they suggest that linguistic evidentiality does not shape non-linguistic source concepts by showing that speakers of English and Turkish converge in their source monitoring abilities. Second, they confirm the tight relation between linguistic and conceptual representations of information sources by showing that the acquisition of linguistic evidentiality is constrained by the development of non-linguistic source concepts. Together, the present findings support a universalist view of the language-cognition interface according to which linguistic categories of evidentiality build on and reflect source concepts that are shared by speakers of different languages.

Chapter 1

INTRODUCTION

Language and Concepts

What is the nature and extent of the relationship between language and thought? This question has been of interest to several linguists, psychologists, anthropologists, and philosophers who have been studying language with the goal of uncovering the inner workings of the mind. For instance, according to Chomsky (1975):

> Language is a mirror of mind in a deep and significant sense. It is a product of human intelligence ... By studying the properties of natural languages, their structure, organization, and use, we may hope to learn something about human nature; something significant, if it is true that human cognitive capacity is the truly distinctive and most remarkable characteristic of the species. (p. 4)

The idea that language and thought are tightly related has also characterized theories of language acquisition. Most theories of language acquisition assume that learning a language, to some extent, involves mapping linguistic input onto preexisting conceptual representations (e.g., Gleitman, 1990; Landau & Gleitman, 1985). In this view, language builds on and *reflects* conceptual development. For instance, across several domains children's linguistic competence has been taken as a strong indicator of their understanding of the underlying concepts (Dromi, 1987; Gopnik & Meltzoff, 1997; Huttenlocher, Smiley, & Charney, 1983, Johnston & Slobin, 1979; Smiley & Huttenlocher, 1995). Furthermore, several commentators have attributed children's difficulties in language acquisition to the complexities in the conceptual prerequisites of the linguistic domain being acquired (Bartsch & Wellman, 1995; Gopnik & Meltzoff, 1997; Perner, 1991).

According to an alternative view, language has a more central role in conceptual development, rather than merely reflecting the underlying concepts. This view is also known as the *Sapir-Whorf hypothesis*, or the *Whorfian hypothesis*, after the thinkers who famously entertained the possibility that one's native language shapes cognition in the beginning of the 20th century (see, e.g., Sapir, 1924; Whorf, 1956). More recent proposals continuing this tradition claim that language is the medium through which concepts are *constructed* (Bowerman & Choi, 2001). Furthermore, language-specific encoding patterns might affect the salience of conceptual distinctions resulting in different timetables for the development of conceptual representations (Bowerman & Levinson, 2001).

Beyond its significance for language acquisition and development, this topic has implications for adult cognition. Whether long-term experience with the grammatical categories of one's native language influences cognitive processes in mature cognizers is widely debated. (For recent reviews see Gentner & Goldin-Meadow, 2003; Gleitman & Papafragou, 2005, 2012; Gumperz & Levinson, 1996; Landau, Dessalegn, & Goldberg, 2010; Levinson, 2003; Ünal & Papafragou, in press; Wolff & Holmes, 2011.)

Several commentators have proposed that linguistic effects on cognition may be ephemeral in nature (Gennari, Sloman, Malt, & Fitch, 2002; Landau, et al., 2010; Li, Dunham, & Carey, 2009; Papafragou, Massey, & Gleitman, 2002). Proponents of this view acknowledge that implicit (subvocal) linguistic encoding can be recruited online to support cognitive processing without permanently altering conceptual representations, but these language-related differences should 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).

Other researchers 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 the world across language communities may lead to stable differences in how members of these communities remember and reason about the world even when they are not explicitly using language (i.e., speaking or comprehending speech).

The predictions of these competing positions have been tested in various domains such as number, color, space, motion and navigation. However, these domains have strong visual and/or spatial components and might be more prone to bottom-up influences. One open possibility is that linguistic influences are more likely to emerge in abstract domains that involve higher-level processes that are removed from perception (Spelke & Tsivkin, 2001). We turn to one such domain in the present investigation.

Empirical Focus

This dissertation examines the relation between language and thought in the domain of source knowledge and its linguistic encoding. 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 knowledge* (Johnson, Hastroudi, & Lindsay, 1993). The process of attributing a piece of information to a specific source is known as *source monitoring* (Johnson, 1988). Sources of knowledge can be external (e.g., vision or audition) or internal (e.g., thoughts, imagination, dreams, or inferences; Johnson & Raye, 1981).

Speakers convey to their listeners how they know about something through *evidentiality* markers. The expression of evidentiality varies across languages (Aikenvald, 2004, 2014; Aikenvald & Dixon, 2001; L. Anderson, 1986; Chafe & Nichols, 1986; De Haan, 2001; DeLancey, 2001; Faller, 2001; Givón, 1982; Willett, 1988). For instance, in English, speakers do not need to encode the source of their information in their utterances. Thus, sentence (1) can be used whether the speaker has directly experienced the event or has only indirect information about it. However, speakers have the option to express evidentiality through lexical devices such as verbs (e.g., *see* and *hear*, etc.), or adverbials (e.g., *allegedly, reportedly,* etc.).

(1) Ali played.

In contrast, about a fourth of the world's languages obligatorily encode evidentiality through grammaticalized means, such as verb morphology (Aikenvald, 2004). For instance, in Turkish, two past tense markers, *-dı* and *–mış* (realized as *-dı*, *-di*, *-du*, *-dü*, *-tı*, *-ti*, *-tu*, *-tü* and *-mış*, *-miş*, *-muş*, *-müş*, respectively, depending on phonological factors) differentiate direct evidence from indirect evidence (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), *-mış* 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

Despite considerable cross-linguistic variation in the expression of evidentiality, the class of languages that encode evidentiality in their grammar are subject to certain constraints (Faller, 2001). According to Willet's (1988) cross-linguistic survey, the languages that have a grammaticalized evidential system typically encode three sources of information: direct/perceptual access, verbal report of others, or inferential access. Some languages have more detailed evidential systems that encode four or five distinctions. However, these distinctions emerge from subdivisions within the three source categories mentioned above. For instance, Tuyuka –a language spoken in parts of Colombia and Brazil– has a five way evidential system that encodes visual access, non-visual perceptual access, inferences based on direct evidence, verbal report of others and assumptions with different morphemes

(Barnes, 1984). Nevertheless, such systems are very rare and it is more typical for languages to have two-way evidential systems such as Turkish (Aikhenvald, 2004, 2014).

The relation between linguistic evidentiality and non-linguistic source monitoring is an excellent test bed for understanding the relationship between language and thought for several reasons. As noted earlier, information sources are unobservable, abstract concepts, which differ from each other in very subtle ways. As mentioned above, linguistic influences on cognition might be especially prominent in higher-level domains such as evidentiality. Thus the cross-linguistic variation in the expression of evidentiality raises the question whether speaking a language with grammaticalized evidentiality shapes source concepts and influences source monitoring mechanisms. Furthermore, such subtleties and complexities in the concepts encoded by evidentiality markers pose challenges for young learners (Papafragou, Cassidy & Gleitman, 2007; Snedeker & Gletiman, 2004). From a developmental perspective, this raises the question how children overcome those challenges and acquire the evidential distinctions in their language, and how their ability to do so is related to the development of corresponding source concepts. The aim of this dissertation is to address these issues.

Outline of the Dissertation

The goal of this dissertation is to understand the nature of the relation between language and thought using the domain of evidentiality as a test case. Using a cross-linguistic approach, this dissertation asks whether speakers of languages with different evidential systems differ in their non-linguistic representations of information sources. This approach allows us to understand whether or not language shapes thought. From a developmental perspective, this dissertation investigates how linguistic evidentiality is acquired, and how its acquisition is related to the development of source concepts. This approach allows us to understand whether there is a homology between language and thought, and whether language acquisition is constrained by conceptual development.

The dissertation is structured as follows. Chapter 2 asks whether long-term experience with the evidential categories of one's native language affects engagement of source monitoring mechanisms in mature cognizers. To test this, this chapter first empirically establishes the cross-linguistic differences in how English and Turkish speakers mark sources of information for events they have directly or indirectly found out about. Next, the chapter asks whether such cross-linguistic differences are reflected in non-linguistic source memory. There are two possibilities here that differ with respect to the direction of the causal flow between language and thought. One possibility is that Turkish and English speakers converge in their source monitoring abilities, in case linguistic evidentiality does not modify the underlying representations of information sources. An alternative possibility is that Turkish speakers are less prone to source monitoring errors compared to English speakers. Thus the features that distinguish different sources might be more salient or accessible to Turkish speakers as a consequence of having to make source distinctions in their language.

Chapter 3 asks whether there is evidence for a direct link between the acquisition of linguistic evidentiality and the development of source concepts within speakers of a single language. To address this question, this chapter tests young learners of Turkish between the ages of 3 and 6 on a series of carefully-matched experimental paradigms assessing linguistic and non-linguistic performance. The linguistic tasks include measures of production and comprehension of evidential markers. The non-linguistic tests include measures of children's understanding of their own and others' access to information. If there is a homology between language and thought, then the patterns observed in the linguistic domain with respect to the relationship between evidential production and comprehension should be mirrored in the non-linguistic counterparts of evidential production (i.e., accessing one's own information sources) and comprehension (i.e., accessing others' information sources). If confirmed, this hypothesis will inform theories of language acquisition by showing that the acquisition of linguistic evidentiality is constrained by the development of source concepts.

Building on the findings of Chapter 3, Chapter 4 aims to further explore the development of source concepts in young learners of Turkish. To do so, this chapter focuses on the relationship between gaining knowledge for oneself and attributing knowledge to others. The studies reported in this chapter explore developmental changes in the ability to connect evidence to knowledge and include a comparison to English learners.

Chapter 5 provides a summary of the results of the studies presented in Chapters 2, 3 and 4. This chapter also discusses theoretical implications of these results and ends with open questions and future directions for this line of research. To foreshadow the conclusions, the present findings provide evidence for a universalist view of the relation between language and cognition, according to which linguistic categories of evidentiality tightly map onto conceptual representations of information sources and reflect these possibly shared source concepts without shaping them.

Chapter 2

LINGUISTIC EVIDENTIALITY AND SOURCE MONITORING

Introduction

It has long been recognized that people do not readily bind their memories, knowledge or beliefs to the corresponding sources (Johnson et al., 1993; Johnson, 1997, 2006). 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 (Lindsay, Johnson, & Kwon, 1991; Johnson, Bransford, & Solomon, 1973; Mather, Johnson, & De Leonardis, 1999), and the subject's awareness of the cognitive operations that produced the memory (Durso & Johnson, 1980; Johnson & Raye, 1981; Intraub & Hoffman, 1992).

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 (R. E. 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; Lindsay & Johnson, 1989; Johnson et al., 1977, 1979, 1984). 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 (R. E. Anderson, 1984; Dallett & Wilcox, 1968).

Linguistic Influences on Source Monitoring

One question left open in the literature is whether source monitoring might be susceptible to linguistic influences. The two positions sketched in the introduction on the language-cognition interface make different predictions about whether and how language might become implicated in source monitoring. On one view long-term experience with the linguistic categories of one's native language shapes conceptual representations. On an alternative view, language may temporarily impact cognitive processes without modifying the underlying concepts.

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-Balanban, 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 Drummey & 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 (-*mış*) 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 Drummey 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, 2015; 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 (*-dı*) or the indirect past tense (*-mış*) 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 eye-witness memory (e.g., Strack & Bless, 1994; Hart & Albarricin, 2011). 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 chapter, 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. Fist, 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 chapter follows standard practice in experimental tests of the relation between language and thought in the literature (Gleitman & Papafragou, 2005).

Overview of the Current Chapter

In the current chapter, 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; Johnson, Kahan & Raye, 1984; Johnson, Raye, Wang & Taylor, 1979; Johnson, Taylor & Raye, 1977; Intraub & Hoffman, 1992; 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 (R. E. Anderson, 1984; Hayes-Roth & Thorndyke, 1979; Johnson et al., 1993, 1994; Johnson & Raye, 1981; 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, Kounios, & Reeder, 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 (Mage = 19.1, range = 18-22, 9 females) and 12 native speakers of Turkish (Mage = 19.6, range = 18-23, 8 females). In the Source Identification Task, participants were a new group of 24 English speakers (Mage = 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 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-inch 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.[1]

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 *–mış* (M = 0.25), ($t(15) = 9.49$, $p < .001$), and inferred events using the indirect morpheme *–mış* (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 *-mış* (and not variants such as *-mışmış* or *–mışdır*).

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 *–mış*. Using a median split on *–mış* 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 Figure 1). Mean use of *–mış* 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, $t(14) = 4.31$, $p < .001$.[2]

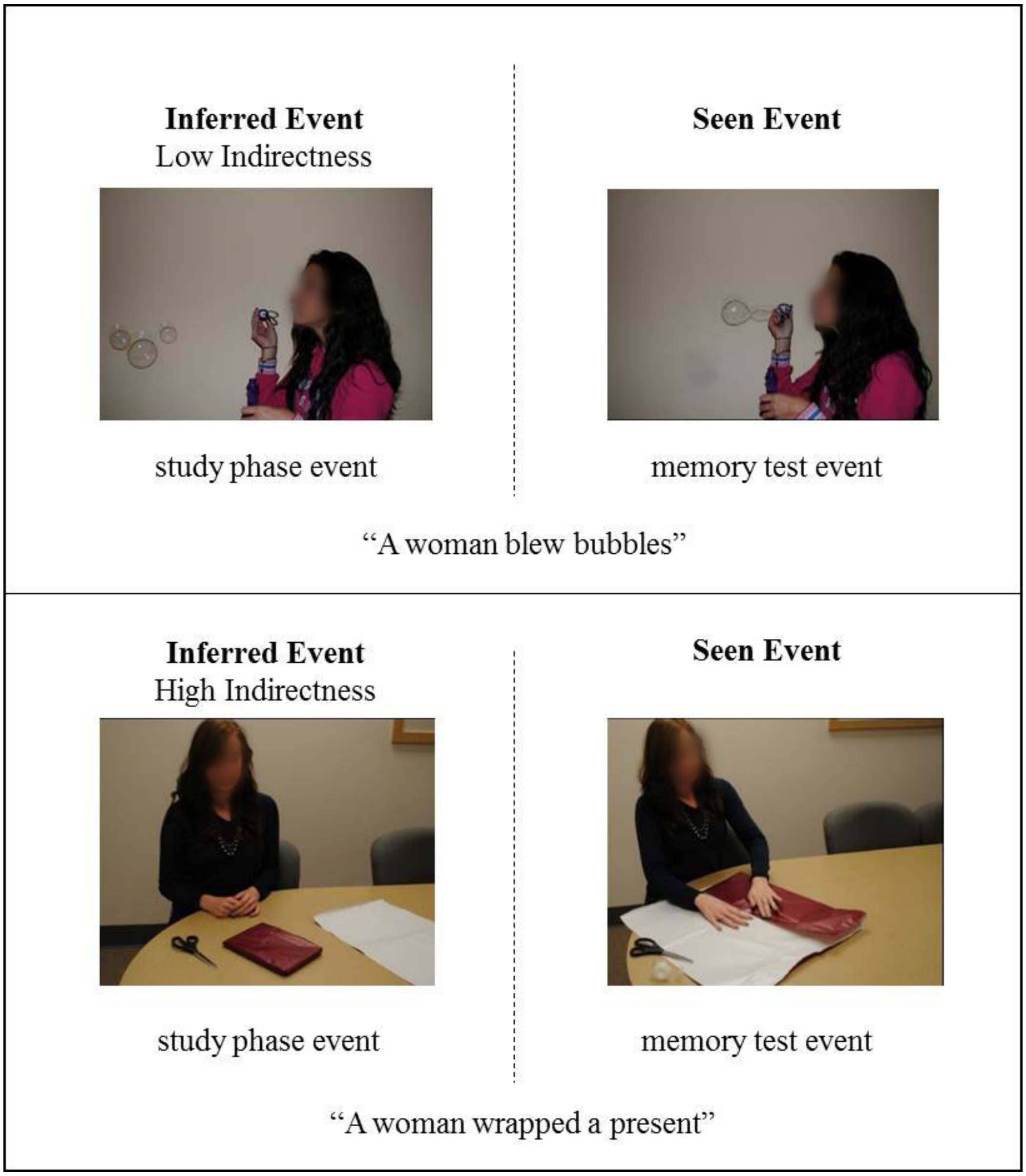

Figure 1 Examples of target events undergoing Source Changes

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 travelling 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' judgments draw the subtle boundary between perception and inference.

Source identification task

Participants correctly identified the target event 78% of the time (M_{seen} = 0.78, $M_{inferred}$ = 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, χ^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, χ^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 (*–mış*) 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, 2015; Papafragou, Li, Choi, & Han, 2007; Ünal & Papafragou, 2013; 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 Lindsay & Johnson, 1989; Hayes-Roth & Thorndyke, 1979; 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_{age} = 19.29, range = 18-29, 27 females) and 48 Turkish speakers (M_{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

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 Experiment 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 seconds 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

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 seconds 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

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 x 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 ($M = 0.76$).

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)

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.

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	
	"Inferred"	"Not seen or inferred"	"Seen"	"Inferred"
English	0.57 (0.35)	0.42 (0.35)	0.12 (0.26)	0.88 (0.26)
Turkish	0.61 (0.42)	0.39 (0.42)	0.24 (0.35)	0.76 (0.35)
	Source Change			
	High Indirectness		Low Indirectness	
	"Seen"	"Not seen or inferred"	"Seen"	"Not seen or inferred"
English	0.65 (0.37)	0.35 (0.37)	0.63 (0.39)	0.37 (0.39)
Turkish	0.41 (0.39)	0.59 (0.39)	0.63 (0.39)	0.37 (0.39)

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 x 2 x 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 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 (towards “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 x 2 ANOVA with Language (English, Turkish) and Condition (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 ($t(44) = 2.62$, $p = 0.012$). 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 ($M = 1139.31$).

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)

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 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.

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)

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; Lindsay & Johnson, 1989; Johnson, et al., 1977, 1979, 1984). 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_{age} = 18.60, range = 18-21, 31 females) and 50 native speakers of Turkish (M_{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 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 x 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 ($M = 0.83$).

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)

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

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	
	"Inferred"	"Neither"	"Seen"	"Inferred"
English	0.69 (0.38)	0.31 (0.38)	0.21 (0.32)	0.78 (0.32)
Turkish	0.72 (0.30)	0.28 (0.30)	0.15 (0.26)	0.85 (0.26)
	Source Change			
	High Indirectness		Low Indirectness	
	"Seen"	"Neither "	"Seen"	"Neither"
English	0.71 (0.33)	0.28 (0.37)	0.78 (0.28)	0.22 (0.28)
Turkish	0.60 (0.33)	0.40 (0.33)	0.71 (0.28)	0.29 (0.28)

Crucially, for Source Changes we analyzed error patterns with a 2 x 2 x 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; Lindsay & Johnson, 1989; Johnson, et al., 1977; 1979; 1984), 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 x 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.

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)

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; Lindsay & Johnson, 1989; Johnson et al., 1977, 1979, 1984).

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; Levinson, 2003; Lucy, 1992; Imai & Gentner, 1997; 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 in the introduction). 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; Harris, 1974; Hannigan & Reiniz, 2001; 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 (*–mış*) 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, 2015). 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 & Reiniz, 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 (Lindsay et al., 1991; Bransford & Johnson, 1973; 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 present studies focused on two memory sources that share several characteristics, namely visual perception and inference from visual cues. In the current studies, 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 studies, both sources were essentially composed of visual attributes and delivered similar information about the event, making source confusions more likely (cf. also Harris, 1978).

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; Levinson, 2003; Lucy, 1992; Imai & Gentner, 1997; 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, & Dosher, 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.

Chapter 3

ACQUISITION OF LINGUISTIC EVIDENTIALITY

Introduction

When learning their native language, children typically understand the mappings between forms and their meanings before they can produce these forms in speech. However, cases where the production of a form is more advanced that its comprehension are not uncommon (e.g., Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; de Villiers, Cahillane, & Altreuter, 2006; for an overview see Hendriks, 2013; Hendriks & Koster, 2010). In the studies reported here, we examine a striking case of a production-comprehension asymmetry in children' acquisition of evidentiality. This pattern has been previously noted in the developmental literature (Aksu-Koç, 1988; Ozturk & Papafragou, 2015; Papafragou, Li, Choi, & Han, 2007), but its origins have not been systematically explored. Our goal is to offer new, more robust empirical evidence for the asymmetry focusing on the acquisition of evidential morphology in Turkish; more importantly, we seek to evaluate the relative contribution of methodological demands and the psycholinguistic properties of evidentiality to children's comprehension difficulties. We aim to show that the facts about the acquisition of evidentiality are best understood as the result of an interaction between evidential meanings and the inherent perspective-taking asymmetries between production and comprehension. Before laying out the proposals that we are going to evaluate, we review past studies describing the acquisition of the Turkish evidential system.

Acquisition of Evidentiality

Recall that in Turkish, two past tense verbal affixes encode direct sources (perception; –*dı*) or indirect sources (hearsay/inference; –*mış*). Mothers of a nationally representative sample of Turkish children report that 83% of the children produce –*dı* and 48% of the children produce –*mış* by age 2, whereas these rates increase to 98% and 93% (respectively) by the time children are 3-years-old (Aksu-Koç et al., 2011). Similarly, naturalistic observations of children's spontaneous speech reveal that –*dı* and –*mış* emerge between the ages of 2 and 3 (Aksu-Koç, 1988). Nevertheless, these frequencies need to be interpreted with caution and supplemented with observations in more controlled contexts. In fact, experimental studies have revealed that, across languages, full semantic and pragmatic understanding of evidentiality does not develop until the end of the kindergarten years, and sometimes even later (see Matsui, 2014 for a review; cf. Aksu-Koç, 1988; Aksu-Koç, Ögel-Balaban, & Alp, 2009; Choi, 1995; Fitneva, 2009; Ozturk & Papafragou, 2007, 2015; Papafragou et al., 2007; Ünal & Papafragou, 2013; de Villiers, Garfield, Gernet-Girard, Roeper, & Speas, 2009). Importantly for present purposes, children in many of these studies consistently fail on comprehension tasks, despite the fact that they can reliably produce the evidential morphemes in speech (e.g., Aksu-Koç, 1988; Ozturk & Papafragou, 2015; Papafragou et al., 2007).

In a pioneering study, Aksu-Koç (1988) investigated the acquisition of evidential morphology by Turkish-speaking children. In an elicited production task, 3- to 6-year-old children accessed events acted out with toys from a witnessed/perceptual or a non-witnessed/inferential perspective and were asked to describe the events. Witnessed events were expected to be described with –*dı* and non-witnessed events were expected to be described with –*mış*. The same children were also given an comprehension task, in which they had to match evidentially marked utterances to characters in a story based on the characters' informational access (witnessed/inferred) to events (henceforth, the "who-said-it" task). In the production task, 3- and 4-year-olds produced the appropriate evidentials at levels at or higher than 70%. However, in comprehension the same level of performance emerged only at age 6 and only on *–dı* trials; furthermore, even 6-year-olds performed at chance level (56%) on *–mış* trials. Three- and 4-year-old children's comprehension performance was around 50% on *–dı* trials and 40% on *–mış* trials. In a follow-up task where children were asked to explain why they picked a particular character, only 25% of the children were able to correctly justify their choices by referring to the character's informational access to the event.

A similar asymmetry was obtained by Ozturk and Papafragou (2015). Turkish-speaking children between ages 5 and 7 had to describe short clipart animations that either fully depicted an event or depicted some evidence that would allow the children to infer what happened. Children consistently marked the events they saw with *–dı* (98% of the time), but were a lot less consistent in marking the events they inferred with *–mış* (52% of the time). Comprehension was measured with a version of the "who-said-it" task. Children were correct about 65% of the time for *–dı* trials, and about 60% of the time for *–mış* trials. A direct comparison between the production and the comprehension tasks confirmed that children performed better in the production task compared to the comprehension task.

A further study by Papafragou et al. (2007) showed that this asymmetry is not language-specific. In that study, 3- and 4-year-old Korean children were quite successful in producing both a direct (*–e*) and an indirect/reportative morpheme (*–tay*);

nevertheless, children had difficulty in a “who-said-it” task measuring evidential comprehension. Two further versions of the comprehension task were conducted. In one of these tasks, one character uttered a statement with either the direct or the indirect morpheme and children had to accept/correct the statement depending on the evidential basis of the speaker. In another task, one character uttered a sentence with a direct and another character a sentence with an indirect morpheme and children were asked either who saw or who was told. These versions also returned low comprehension accuracy.

Finally, there is evidence that the production-comprehension asymmetry is not limited to languages that encode evidentiality morphologically. In English, perception verbs such as *seem, sound, look, feel* can be used to syntactically encode evidentiality. The raised form of such verbs (e.g., “John looks like he is sick”) expresses that the speaker has direct evidence for the basic-level proposition asserted in the utterance (i.e., John is sick), whereas the unraised form of the verbs (e.g., “It looks like John is sick”) does not make such a commitment and thus the speaker’s evidence for the asserted proposition could either be direct or indirect. Rett and Hyams (2014) conducted a corpus study on the CHILDES database (MacWhinney & Snow, 1985, 1990) with children between the ages of 2 and 6 and confirmed that children begin modifying their utterances (i.e., use the raised or the unraised form of the verbs) depending on the type of evidence (direct or indirect) between the ages of 2 and 3. In a later study, Winans, Hyams, Rett, and Kalin (2014) measured comprehension using a felicity judgment task. Children and adults were presented with pictures that depict either direct or indirect evidence and were asked to evaluate whether the raised or the unraised forms were “a good or silly way of saying what is going on in the picture.”

Adults accepted unraised forms regardless of type of evidence, but they overwhelmingly rejected raised forms when presented in the context of indirect evidence. By contrast, children between ages 4 and 6 judged both raised and unraised forms to be acceptable in both direct and indirect evidence contexts and thus showed no sensitivity to the relation between the syntactic form and type of evidence.

Taken together, these studies point to a production-comprehension asymmetry in the domain of evidentiality that appears to emerge cross-linguistically. However, two issues remain open about the scope and nature of this asymmetry. A first issue is that prior work has typically used different stimuli to measure production and comprehension (Aksu-Koç, 1988; Ozturk & Papafragou, 2015; Papafragou et al., 2007; Rett & Hyams, 2014, Winans et al., 2014). It is therefore possible that at least some of the observed effects relate to superficial differences between experimental materials. A second, more substantive issue is that prior studies raised various theoretical possibilities to explain the observed production-comprehension difference (Aksu-Koç, 1988; Papafragou et al., 2007; Winans et al., 2014) but were not set up to adjudicate between them.

In general, the spectrum of explanations for the evidentiality facts can be organized into two classes that mirror explanations for production-comprehension asymmetries in other domains (see Ozturk & Papafragou, 2015 for a discussion). One class of explanations is methodological: it proposes that the lag in the comprehension of evidentiality is an artefact of the different methods used to test production and comprehension. Recall, for instance, that many prior studies used a version of the "who-said-it" task where an evidentially marked utterance had to be matched to one of two characters in accordance with their access to information. In order to perform successfully in that task, children had to identify the two characters' informational access, retain an evidentially marked utterance in their memory, unpack the meaning of the evidential morpheme, think about how each of the two characters would have described the event and pick the character whose informational access matched the evidentially marked utterance. Both memory and/or metalinguistic demands might have made this comprehension task more challenging than a production task. A similar argument can be made about felicity judgment tasks involving evidentials (Papafragou et al., 2007; Winans et al., 2014). Thus, cognitive-resource limitations might be responsible for children's comprehension failures.

A second class of explanations is psycholinguistic: it proposes that the same evidential form needs to be processed differently in production vs. comprehension because of the way the meaning of evidentials interacts with the self-other perspective difference between these two processes. Specifically, in the production task, the speaker encodes his/her own informational sources using evidential morphology, whereas in the comprehension task the hearer must consider someone else's (the speaker's) informational access in order to interpret an evidentially marked utterance. To the extent that evidential comprehension is an inherently metacognitive task that requires reasoning about someone else's information sources and perspective, it is reasonable to expect it to be more costly compared to evidential production. In support of this possibility, there is some evidence suggesting that representing one's own mental states, including the sources of one's knowledge, develops earlier than representing the mental states of others (Pillow, 2002; Pillow & Anderson, 2006; Pillow, Hill, Boyce, & Stein, 2000; Povinelli & deBlois, 1992; Sodian & Wimmer, 1987; Wimmer, Hogrefe, & Perner, 1988). However, most of this evidence comes from tasks that involved explicitly asking children about mental state contents and may have underestimated children's knowledge. Furthermore, these studies differed considerably in their methods and stimuli from the studies on linguistic evidentiality and so their results cannot be directly compared.

Overview of the Current Chapter

In this chapter, we present a series of five experiments that assessed Turkish learners' evidential production and comprehension. Throughout, we focus on the direct evidential (*–dı*) and the inferential interpretation of the indirect evidential (–*mış*). We begin by seeking more robust evidence for whether production precedes comprehension in the acquisition of evidentiality. In Experiment 4, we elicited evidential production using a task modeled after Aksu-Koç (1988). In Experiment 5, we developed a novel task inspired by the earlier "who said it" task to measure evidential comprehension using the same events. To foreshadow our findings, we replicated the asymmetry between correct production and comprehension.

In the remaining experiments, we tested competing predictions made by the two broad explanations of the evidential asymmetry. On the methodological hypothesis, the asymmetry should disappear if comprehension is assessed in tasks that minimize memory and metalinguistic demands; on the psycholinguistic hypothesis, the asymmetry should persist. To test these predictions, in Experiments 6 and 7 we used novel, simpler tasks to assess the comprehension of evidentiality and compared the results to both Experiment 4 and 5.

Furthermore, the methodological hypothesis expects children's difficulty to be tied to specific task demands, and hence does not expect similar patterns of difficulty to emerge in tasks in which children have to reason about sources of information in the absence of evidential morphology; however, the psycholinguistic hypothesis predicts that children's difficulty should extend to non-linguistic contexts in which children have to reason about others' information sources (but should diminish or disappear if children have to reason about evidence for information from their own perspective). To test these predictions, in Experiment 8 we removed evidential language from the comprehension task of Experiment 7, and asked children to reason about either someone else's or their own evidence for information.

Experiment 4

The goal of Experiment 4 was to elicit production of evidential morphemes for direct and indirect/inferential evidence for events. Of interest was whether children would modify their descriptions of the events based on the evidence they were presented with. For this and all subsequent tasks, we adopted a puppet theater set-up inspired by some of Aksu-Koç's (1988) tasks in which an event either takes place in full view of the child or occurs behind the curtains such that only its beginning and endpoint are observable. We reasoned that this set-up would highlight different types of access to an event (perceptual vs. inferential) that might not otherwise be salient to children.

Method

Participants

Participants were native speakers of Turkish distributed across four age groups: 3-year-olds (n = 12, 3;1-3;10, M_{age} = 3;7), 4-year-olds (n = 12, 4;0-4;8, M_{age} = 4;4), 5- to 6-year-olds (n = 12, 5;6-6;5, M_{age} = 6;0), and adults (n = 7, 25-27, M_{age} = 26). Children were recruited through preschools in Istanbul, Turkey. Adults were students at Koç University, Turkey and participated in the experiment to satisfy a course requirement.

Stimuli

There were two types of trials: Target trials and Filler trials. The stimuli for Target trials consisted of mostly change of state events in which a puppet performed actions in a puppet theater. There were two types of access depicted in these events: For Seen Events, the curtains of the puppet theater were open throughout the event, so that the participants witnessed the event (e.g., they saw the puppet stack some blocks). For Inferred Events, the curtains of the puppet theater were open for the beginning of the event (such that children saw, e.g., the puppet holding a balloon). Then, the curtains were drawn and the event unfolded. Then, curtains were pulled back so that the end state of the event (e.g., an inflated balloon) was observed. Thus, even though the participants did not see the event, they could infer what had happened based on available evidence. For Inferred Events, the agent was present for the beginning but not for the end state of the event. This is because pilot testing revealed that when the puppet was included at the end of the event alongside the object that went through a change, children sometimes described what object the puppet had (e.g., "He has a balloon") instead of describing what happened (i.e., the event). Thus, the puppet was excluded from the end-state of the events to elicit descriptions of the events. Examples of Seen and Inferred events are presented in Figure 2.

For Filler trials, the puppet showed an object that was fully visible to the participants (e.g., a giraffe) but did not perform any action on the object. The list of events used in the production task (and the subsequent comprehension tasks) is presented in Appendix B.

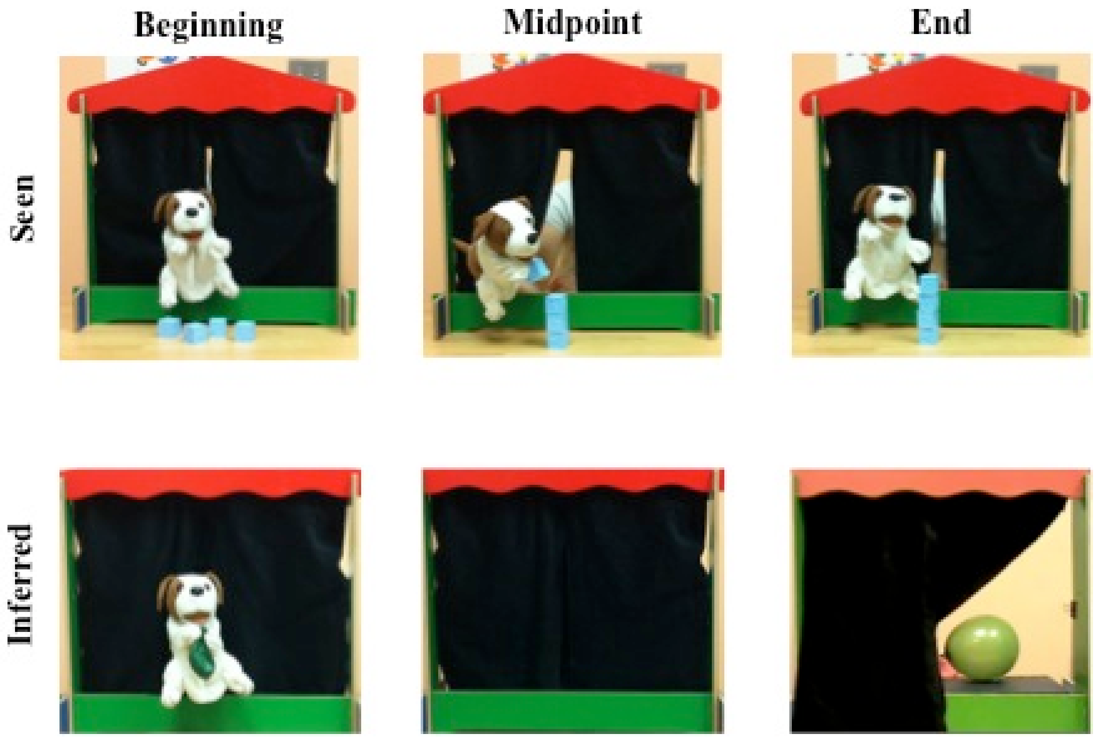

Figure 2 Snapshots of the videos for Seen and Inferred Events in the Production task (Experiment 4).

Four examples of each type of event were used for a total of 8 Target trials in addition to 8 Filler trials. Participants were presented with all 16 trials in the same semi-randomized order, with the constraint that there was a Filler trial in-between two Target trials.

Procedure

Each participant was tested individually in a quiet room at his/her preschool or university campus. Each child was seated in front of the puppet theater. Experimenter 1 (E1) always interacted with the child. Experimenter 2 (E2) acted out the puppet, Mr. Nut, from behind the puppet theater and was never seen by the child. E1 introduced

Mr. Nut to the child and asked the child to describe what Mr. Nut did. In Filler trials, the child was asked to tell what Mr Nut's toy was. We audio-recoded the child's descriptions. The procedure for the adults was exactly the same, except that they watched pre-recorded videos of the puppet performing the actions. We were interested in whether Seen events would lead to the production of direct past tense (–*dı*) and Inferred events would lead to the production of indirect past tense (–*mış*).

Results

Analytical strategy

Data from Experiment 4 (and all subsequent experiments) were analyzed using multi-level mixed logit modeling with crossed random intercepts for Subjects and Items (Baayen, 2008; Baayen, Davidson, & Bates, 2008). This analytical approach has two benefits. First, this approach allows subjects and items to be treated as random factors in a single model. Second, unlike traditional analyses of variance (ANOVA) on proportions of categorical outcomes obtained from subject and item means, this approach allows for better treatment of categorical data (Jaeger, 2008, cf. Barr, 2008, Fraundorf, Benjamin, & Watson, 2013).

Model fitting and results

All models were fit using *lmer* function of the *lme4* package (Bates, 2005; Bates, Maechler, & Bolker, 2011; Bates, Maechler, Bolker, & Walker, 2015; Bates & Sarkar, 2007) in *R* Project for Statistical Computing (R Development Core Team, 2007). The fixed effects that were investigated in Experiment 4 were Condition (Seen, Inferred), Age (Adults, 3-year-olds, 4-year-olds, 5- to 6-year-olds) and an interaction between Condition and Age. The fixed effect of Condition was coded with centered contrasts (-0.5, 0.5). The fixed effect of Age was analyzed with three planned comparisons using simple contrast coding (c_1: -0.25, 0.75, -0.25, -0.25; c_2: -0.25, -0.25, 0.75, -0.25; c_3: -0.25, -0.25, -0.25, 0.75) (Cohen, Cohen, West, & Aiken, 2003). This coding strategy allowed us to compare children in each of the three age groups to the adult reference group while the intercept corresponded to the grand mean. Fixed effects of Condition and Age in Experiments 5 and 6 were also coded using the same contrasts. For completeness, our tables report parameter estimates of all of the fixed effects that were tested (main effects of Condition and Age, and Condition by Age interaction), including the non-significant fixed effects.

Participants' descriptions were transcribed and coded. Beginning with Filler trials, both children and adults were highly accurate in labeling the objects ($M_{3\text{-}year\text{-}olds}$ = 0.93, $M_{4\text{-}year\text{-}olds}$ = 0.93, $M_{5\text{- to }6\text{-}year\text{-}olds}$ = 0.96, M_{adults} = 0.94).

For target trials, first we examined the proportion of non-past tense uses for Seen and Inferred events across the four Age groups (Table 8). Both children and adults sometimes used non-past tense descriptions, but this tendency was especially prominent in the youngest group of children.

Table 8 *Proportion of Non-Past Tense Uses for Seen and Inferred Events (Experiment 4)*

	Adults		3-year-olds		4-year-olds		5- to 6-year-olds	
	M	SE	*M*	SE	*M*	SE	*M*	SE
Seen	0.07	0.10	0.23	0.13	0.08	0.08	0.00	0.00
Inferred	0.00	0.00	0.05	0.06	0.00	0.00	0.02	0.04

Table 9 *Fixed effect estimates for multi-level model of Non-Past Tense Uses in Experiment 4*

Effect	Estimate	SE	t-value
Intercept	0.05	0.02	2.58*
Condition (Inferred vs. Seen)	0.08	0.02	3.73***
Age (Adults vs. 3-year-olds)	0.10	0.07	1.52
Age (Adults vs. 4-year-olds)	0.01	0.07	0.09
Age (Adults vs. 5- to 6-year-olds)	-0.02	0.07	-0.39
Condition (Seen): Age (3-year-olds)	0.11	0.06	1.71
Condition (Seen): Age (4-year-olds)	0.01	0.06	0.19
Condition (Seen): Age (5- to 6-year-olds)	-0.09	0.06	-1.45

* $p < .05$
** $p < .01$
*** $p < .001$
Formula in R: Non-Past~CONDITION * AGE + (1 | ID) + (1 | EVENT)

Table 9 presents fixed effect parameter estimates for the multi-level model of the non-past tense uses for the target events. The model was fitted using restricted maximum likelihood estimation (REML) of parameters. The dependent variable was binary values (present, absent) for the use of non-past tense descriptions at the item level. Subjects (ID) and Items (EVENT) were added as crossed random intercepts and Condition (Seen, Inferred) and Age (Adults, 3-year-olds, 4-year-olds, 5- to 6-year-olds) were added as fixed factors. The model revealed only a significant main effect of Condition ($p < .001$): proportion of non-past tense descriptions was higher for Seen events ($M = 0.10$) as opposed to Inferred events ($M = 0.02$). No other effects or interactions were significant.

Table 10 *Proportion of direct past tense (-dı) out of total past tense uses (Experiment 4)*

	Adults		3-year-olds		4-year-olds		5- to 6-year-olds	
	M	SE	*M*	SE	*M*	SE	*M*	SE
Seen	1.00	0.00	0.91	0.09	0.86	0.10	1.00	0.00
Inferred	0.11	0.12	0.07	0.08	0.15	0.10	0.28	0.13

Table 11 *Fixed effect estimates for multi-level model of direct past tense use in target event descriptions in Experiment 4*

Effect	Estimate	SE	t-value
Intercept	0.55	0.03	18.28***
Condition (Inferred vs. Seen)	0.79	0.04	19.41***
Age (Adults vs. 3-year-olds)	-0.06	0.08	-0.78
Age (Adults vs. 4-year-olds)	-0.05	0.08	-0.59
Age (Adults vs. 5- to 6-year-olds)	0.09	0.08	1.10
Condition (Seen): Age (3s)	-0.06	0.10	-0.57
Condition (Seen): Age (4s)	-0.17	0.09	-1.84
Condition (Seen): Age (5 to 6s)	-0.17	0.09	-1.84

* $p < .05$
** $p < .01$
*** $p < .001$
Formula in *R*: DI_PAST ~ CONDITION * AGE + (1 | ID) + (1 | EVENT)

For our main analysis, we excluded non-past tense descriptions such that direct past tense (-*dı*) and indirect past tense (-*mış*) uses added up to 1. Table 10 presents the proportion of descriptions marked with the direct past tense (-*dı*) out of total past tense descriptions for Seen and Inferred events across the four age groups. Scores closer to 1 indicate that participants were more likely to use direct past tense (-*dı*); and scores closer to zero indicate that participants were more likely to use indirect past tense (-*mış*). Both adults and children marked Seen events with the direct past tense (-*dı*); also, both age groups avoided marking Inferred events with the direct past tense (-*dı*) and instead marked them with the indirect past tense (-*mış*).

Table 11 summarizes fixed effect estimates for the multi-level model of the past-tense descriptions of target events. The model was fitted using REML estimates of parameters. The dependent variable was the binary values (present, absent) of the use of direct past tense (-*dı*) marking at the item level. As mentioned above, Subjects (ID) and Items (EVENT) were included as crossed random intercepts, and Condition (Seen, Inferred) and Age (Adults, 3-year-olds, 4-year-olds, 5- to 6-year-olds) were included as fixed factors. The model revealed only a main effect of Condition ($p < .001$). No other effects or interactions were significant. Adults and children in all three age groups were more likely to mark Seen events with the direct past tense (-*dı*) as opposed to the Inferred events. These findings confirm our prediction that Seen events would elicit –*dı* and Inferred events would elicit –*mış*.

Discussion

Our findings suggest that young learners of Turkish begin to successfully differentiate the two past tense markers on the basis of their evidential function at the age of 3. Our task revealed better performance compared to prior studies on Turkish, especially in terms of the indirect past-tense morpheme (Aksu-Koç, 1988; Ozturk & Papafragou, 2015; Ünal, & Papafragou, 2013). We speculate that our task provided naturalistic stimuli with rich contextual information for inferred events (e.g., closing the curtains while the event was unfolding), thereby facilitating the use of the indirect past tense (–*mış*).

Experiment 5

The goal of Experiment 5 was to test young Turkish learners' evidential comprehension using the events that reliably elicited evidential production from children of a comparable age in Experiment 4. To do so, we modified the "who said it" task of Aksu-Koç (1988). As in that task, we presented children with both "seen" and "inferred" depictions of the same event. Unlike the earlier task, however, where two characters gained access to the event, a single person (the experimenter) gained access to both versions of the event and later produced an utterance marked with either the direct (*–dı*) or the indirect (*–mış*) evidential. We were interested in seeing whether children would match utterances marked with the direct past tense (*–dı*) to the seen version and utterances marked with the indirect past tense (*–mış*) to the inferred version of the events.

Method

Participants

Data were collected from a new group of native speakers of Turkish in four age groups: 3-year-olds (*n* = 11, 3;0-3;10, M_{age} = 3;5), 4-year-olds (*n* = 11, 4;3-4;10, M_{age} = 4;6), 5- to 6-year-olds (*n* = 10, 5;7-6;7, M_{age} = 6;0), and adults (*n* = 9, 18-23, M_{age} = 19;6). Children were recruited through preschools in Istanbul, Turkey. Adults were students at Koç University, Turkey and received course credit for participation.

Stimuli

As in Experiment 4, there were two types of stimuli. For Target trials, stimuli consisted of videos of a subset of the target events in Experiment 4.[3] Each target event had two versions: Seen and Inferred. As in Experiment 4, for Seen Events, the curtains of the puppet theater were open for the beginning, midpoint and end of the event so that the event could be seen. For Inferred events, the curtains were open only for the beginning and end of the event, so that the event could be inferred on the basis of available evidence. Both versions of each event were performed by the same female experimenter, instead of a puppet. In each trial, the two versions of each event (Seen and Inferred) were arranged on a single screen (left-right position of versions within a slide was counterbalanced).

For Filler trials, stimuli consisted of videos of the same experimenter holding objects (e.g., a giraffe, a duck). In each trial, videos containing two different objects were paired together (e.g., the experimenter holding a giraffe was on the left was paired with the experimenter holding a duck). Again, these pairs were placed on a single display and the position of the video on the slide was counterbalanced. There was a total of 6 Target trials and 6 Filler trials. Items were arranged in a semi-randomized order, with the constraint that there was a Filler trial in-between Target trials. Half of the participants received the items in the reverse order.

Procedure

Each participant was tested individually in a quiet room at his/her preschool or university campus. Participants were tested using a 13-in. MacBook Pro laptop. Participants were tested by an experimenter who was different from the one that acted out the events in the videos. The experimenter and the participant sat across from the screen and next to each other. On Target trials, the experimenter said: “Look! There are two videos here. We are going to watch them one by one and then I’m going to describe only one of them.” Then, the experimenter and the participant watched videos of Seen and Inferred versions of the same event (e.g., the block-stacking episode) that were presented one after the other on two sides of the screen (see Figure 3 for a schematic depiction). Next, the experimenter uttered a description with either the direct past tense *–dı* as in (1) or the indirect past tense *–mış* as in (2). Then, the experimenter asked: "Which one did I describe?"

(1) Küp-ler-i diz-di.
Block-pl-ACC stack.PAST dir.3sg

(2) Küp-ler-i diz-miş.
Block-pl-ACC stack.PAST ind.3sg

On Filler trials, participants watched videos of the same person holding two different objects (e.g., a duck and a giraffe) that were shown side by side on the screen. Again, the videos were presented sequentially. The experimenter labeled one of the objects and participants were again asked to find the corresponding video. The video that the experimenter described (i.e., the correct response) was counterbalanced across participants.

If participants differentiated the two past tense morphemes on the basis of their evidentiality, the likelihood of picking the Seen event should change depending on the evidential marking in the sentence uttered by the experimenter. Specifically, participants should pick the Seen event upon hearing an utterance marked with the direct past tense (*–dı*) and avoid picking the Seen event and instead pick the Inferred event upon hearing an utterance marked with the indirect past tense (*–mış*). However, if participants failed to differentiate *–dı* and *–mış* in comprehension, then the likelihood of picking the Seen event should stay the same regardless of whether the experimenter utters a sentence with the direct or the indirect past tense.

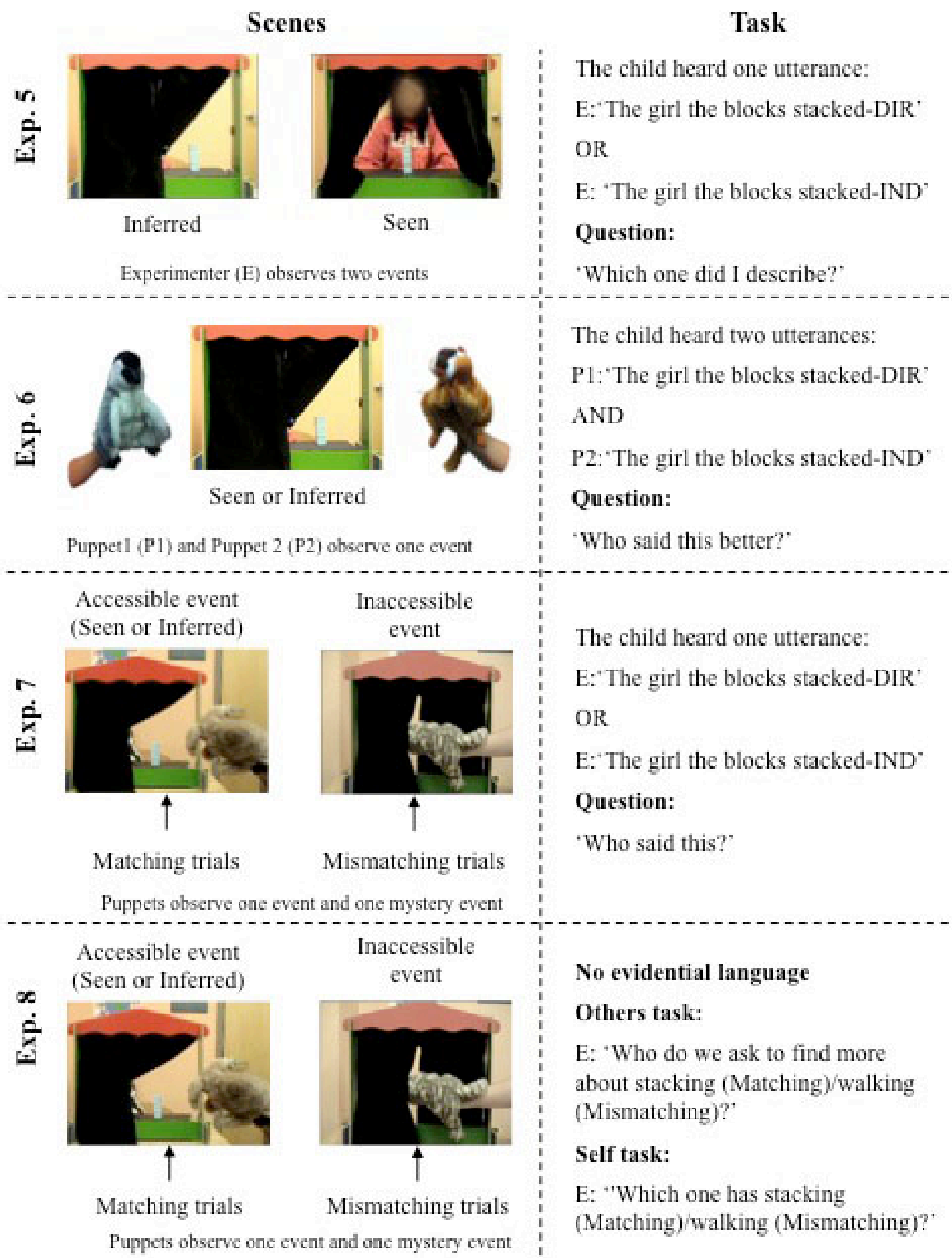

Figure 3 Summary of experimental design for Experiments 5-8.

Results

We followed the same model fitting and effect coding procedures as in Experiment 4 (see Model fitting and results above in Experiment 4 for more information). Beginning with the performance in Filler trials, we assessed performance with a multi-level mixed logit model using REML estimates of the parameters. The dependent variable was binary values (0, 1) for selecting the video that contained the labeled object at the trial level. Subjects (ID) and Items (Object) were included as crossed random intercepts and Age (Adults, 3-year-olds, 4-year-olds, 5- to 6-year-olds) was included as a fixed factor. The model revealed only a significant Intercept ($\beta = 0.95$, SE = 0.02, $p < .001$). The lack of an Age effect suggests that children's performance was at adult level even at age 3 ($M_{3\text{-year-olds}} = 0.88$, $M_{4\text{-year-olds}} = 0.97$, $M_{5\text{- to }6\text{-year-olds}} = 0.98$, $M_{adults} = 1.00$). Thus, the children in our experiment do not seem to be having difficulty with linking object labels to videos.

Next, we assessed performance on Target trials. Table 12 presents proportion of picking the Seen event across Age groups when participants were presented with utterances marked with the direct past tense (-*dı*) or the indirect past tense (-*mış*).

Table 12 *Proportion of picking the Seen event across Age groups (Experiment 5)*

	Adult		3-year-olds		4-year-olds		5- to 6-year-olds	
	M	SE	*M*	SE	*M*	SE	*M*	SE
Direct Past Tense (-*dı*)	0.96	0.06	0.57	0.16	0.61	0.15	0.67	0.17
Indirect Past Tense (-*mış*)	0.15	0.12	0.57	0.16	0.58	0.15	0.46	0.18

Table 13 *Fixed effect estimates for multi-level model of picking the Seen event in Target trials in Experiment 5*

Effect	Estimate	SE	t value
Intercept	0.56	0.07	8.44***
Condition (-*dı* vs. - *mış*)	-0.27	0.06	-4.88***
Age (Adults vs. 3-year-olds)	0.01	0.10	0.11
Age (Adults vs. 4-year-olds)	0.04	0.10	0.35
Age (Adults vs. 5- to 6-year-olds)	0.01	0.11	0.06
Condition (-*dı* vs. - *mış*) : Age (Adults vs. 3s)	0.81	0.16	5.15***
Condition (-*dı* vs. - *mış*) : Age (Adults vs. 4s)	0.79	0.15	5.12***
Condition (-*dı* vs. - *mış*) : Age (Adults vs. 5 to 6s)	0.60	0.17	3.59***

* $p < .05$
** $p < .01$
*** $p < .001$
Formula in *R*: Seen ~ CONDITION * AGE + (1 | ID) + (1 | EVENT)

Data were analyzed using multi-level mixed logit modeling with REML estimates of the parameters. The dependent variable was binary values (0,1) for picking the Seen event. Subjects (ID) and Items (EVENT) were included as crossed random intercepts, and Condition (Heard Direct Past Tense, -*dı*; Heard Indirect Past Tense, -*mış*) and Age (Adults, 3-year-olds, 4-year-olds, 5- to 6-year-olds) were included as fixed factors. Table 13 summarizes the fixed effect estimates for the multi-level model of picking the Seen event in Target trials. The model revealed a main effect of Condition ($p < .001$), and an interaction between Age and Condition ($p < .001$). Follow up analyses on the Age by Condition interaction with separate mixed models on picking the Seen event for each Age group with Subjects and Items as random intercepts and Condition as the fixed factor revealed that the likelihood of picking the Seen event differed as a function of the evidential marking in the utterance (-*dı* or -*mış*), but only for Adults ($p < .001$). Children were equally likely to pick the

Seen event upon hearing an utterance marked with either the direct or the indirect past tense (all $p > .05$).

Furthermore, the proportion of picking the Seen event was compared to chance level (0.50) with one-sample t-tests. As expected, adults selected the Seen event at levels significantly above chance when they heard an utterance marked with the direct past tense (-*dı*) ($t(8) = 12.64$, $p < .001$) and at levels significantly below chance when they heard an utterance marked with the indirect past tense (-*mış*) ($t(8) = -6.10$, $p < .001$). However, children in all three age groups performed at chance level regardless of evidential marking (-*dı* or -*mış*) (all $p > .05$).

Discussion

Taken together, Experiments 4 and 5 demonstrated a production-comprehension asymmetry in the domain of evidentiality. In production (Experiment 4), children reliably used the direct past tense (–*dı*) morpheme to describe events that they visually perceived and the indirect past tense (–*mış*) morpheme to describe events that they inferred from post-event visual evidence. In comprehension, however (Experiment 4), children's likelihood of picking the Seen event did not differ depending on whether they had previously heard utterances marked with the direct past tense (–*dı*) or the indirect past tense (–*mış*) morpheme. Furthermore, children's difficulties seemed to be selective, since children in all three age groups were highly successful in the Filler trials of the comprehension task that involved object labels. Our findings replicate the asymmetry documented in the literature in both Turkish and other languages (Aksu-Koç, 1988; Ozturk & Papafragou, 2015; Papafragou et al., 2007; Rett & Hyams, 2014; Winans et al., 2014) and extend prior findings by showing that the asymmetry persists even when the same events are used to compare production and comprehension.

At present, our results leave all theoretical options open. A first possibility is that the observed production-comprehension asymmetry is due to the specific task demands of our comprehension task. Even though our goal was to simplify the “who said it” task (Aksu-Koç, 1988; Ozturk & Papafragou, 2015; Papafragou et al., 2007) by having a single person (the experimenter) gain access to the same event from two perspectives (direct/visual vs. inferential), this step may not have helped young children. Furthermore, as in the earlier task, children heard a single evidentially marked sentence and presumably had to generate an unspoken alternative (i.e., the other evidential) so as to decide which version of the event was the best match. Both the generation of such alternatives (Chierchia et al., 2001; Barner, Brooks & Bale, 2011) and the subsequent memory and computational demands of aligning morphemes to events may be problematic for young children. Alternatively, our production-comprehension findings may be due to psycholinguistic factors that have to do with the perspective differences inherent in talking about one’s own access to information vs. unpacking others’ access to information from their speech. (Notice that both of these possibilities can account for the fact that children’s failures did not extend to the object label trials.)

Even though the current data cannot adjudicate between them, these two accounts diverge in their predictions about the robustness and extent of children’s comprehension difficulties. The methodological account predicts that the difficulties should diminish in less demanding measures of evidential comprehension and might entirely disappear in tasks that do not involve understanding evidential morphology.

The psycholinguistic account, however, predicts that the difficulty should persist in several versions of comprehension tasks, as long as they still involve reasoning about the available evidence from another's perspective; furthermore, the difficulty should extend to cases where children are asked to reason about others' knowledge sources in the absence of evidential morphology. In Experiments 6-8, we tested these predictions more fully.

Experiment 6

Experiment 6 introduced a new task that asked children to consider the speaker's informational access in understanding evidentials but was simpler than Experiment 5. In the new task, two puppets gained access to an event in the same way (i.e., they both either saw or figured out what happened). The puppets then went on to offer identical descriptions of the event, except that one was marked with the direct past tense (*–dı*) and the other with the indirect past tense (*–mış*). Children were asked "who said it better". This task had lower demands compared to Experiment 5 in two respects: first, children were presented with only one type of access (perceptual or inferential) to an event instead of seeing two perspectives on the same event; second, children were provided with the two contrastive (direct vs. indirect) descriptions of the event, such that they did not have to generate the other alternative. There is evidence that contrastive contexts are a good tool for revealing children's sensitivity to linguistic distinctions. In one particularly relevant demonstration (Ozturk & Papafragou, 2014), 4- to 5-year-old English learners were asked to evaluate epistemic modal statements in a hide-and-seek scenario. The majority of the children failed to reject a statement such as "The cow may be in the pink box" when the statement was weaker than the available evidence (e.g., when the cow *had to* be in the pink box).

However, when given a choice between *may* and *has to* versions of the same statement in the same context, children overwhelmingly chose the modal that was most appropriate based on the evidence available to the speaker (see also Chierchia et al., 2001; Hirst & Weil, 1982). In line with this evidence, the methodological account - but not the psycholinguistic account - predicts that Turkish learners' evidential comprehension should improve in the contrastive task of Experiment 6 compared to Experiment 5.

Method

Participants

Data were collected from a new group of native speakers of Turkish across four age groups: 3-year-olds (n = 13, 3;4-3;11, M_{age} = 3;8), 4-year-olds (n = 13, 4;1-4;10, M_{age} = 4;7), 5- to 6-year-olds (n = 13, 5;5-6;1, M_{age} = 5;10), and adults (n = 7, 18-32, M_{age} = 22;6). Children were recruited through preschools in Istanbul, Turkey. Adults were students at Koç University, Turkey and participated in the experiment to satisfy a course requirement.

Stimuli

Stimuli were the same as in Experiment 5. Unlike the earlier study, only one version of a Target or Filler event was present on the screen at a time. Two lists were created and each version of a Target event (Seen, Inferred) was assigned to one of the lists. Fillers were the same across lists. Each list thus contained 3 Seen and 3 Inferred Target trials, as well as 6 Filler trials. Each list was arranged in a semi-randomized order, with the constraint that there was a Filler trial in-between two Target trials. Each participant saw one of the two lists.

Procedure

Each participant was tested in a quiet room at his/her preschool or university campus. Testing involved a 13-in. MacBook Pro laptop.

On Target trials, participants were presented with either a Seen or an Inferred event on the screen (see Figure 3 for a sample trial). Two puppets (a penguin and a squirrel) acted out by the same experimenter watched the event with the participants. The pair of puppets remained the same throughout the experiment. Each puppet offered a description of the event: one used the direct past tense (–*dı*) morpheme as in (1) above and the other one used the indirect past tense (–*mış*) morpheme as in (2) above. Participants were asked to choose the puppet that "said it better". In half of the trials, the penguin uttered the correct sentence; in the other half of the trials the squirrel uttered the correct sentence. The assignment of sentences to puppets (and thus the puppet that uttered the correct description) was counterbalanced across participants.

On Filler trials, participants were presented with videos of someone holding an object (e.g., a giraffe). Each puppet labeled the object differently (e.g., "This is a giraffe", "This is a duck"). Again, participants picked the puppet that "said it better".

If participants discriminated between the two past tense markers on the basis of evidentiality, then the likelihood of selecting the utterance marked with the direct past tense (-*dı*) as a better description of the event should change depending on type of access to the event. That is, participants should select the utterance marked with the direct past tense (-*dı*) when presented with a Seen event, and avoid picking the utterance marked with the direct past tense (-*dı*) (and instead pick the utterance marked with the indirect past tense -*mış*) when presented with an Inferred event.

Results

First, we assessed performance in Filler trials with multi-level mixed logit modeling using REML estimates of the parameters. We used the procedures of Experiments 4 and 5 to fit the models and code for fixed effects. The dependent variable was binary values (0, 1) for picking the puppet that correctly labeled the object in the video at the item level. Subjects (ID) and Items (Object) were added as crossed random intercepts and Age (Adults, 3-year-olds, 4-year-olds, 5- to 6-year-olds) as a fixed factor. As expected, the model revealed a significant Intercept (β = 0.95, SE = 0.02, $p < .001$) and no significant effect of Age. Even the youngest group of 3-year-olds was highly accurate on Filler trials ($M_{3\text{-year-olds}} = 0.87$, $M_{4\text{-year-olds}} = 0.96$, $M_{5\text{- to }6\text{-year-olds}} = 0.97$, $M_{adults} = 1.00$).

Next, we examined performance in Target trials. Table 14 presents the proportion of picking the utterance marked with the direct past tense (-*dı*) across Age groups when participants are presented with either Seen or Inferred events.

Table 14 *Proportion of selecting the utterance marked with direct past tense across Age groups (Experiment 6)*

	Adult		3-year-olds		4-year-olds		5- to 6-year-olds	
	M	SE	*M*	SE	*M*	SE	*M*	SE
Seen	0.95	0.13	0.56	0.14	0.72	0.14	0.62	0.14
Inferred	0.14	0.08	0.41	0.14	0.56	0.13	0.59	0.13

Table 15 presents the fixed effect estimates for the multi-level model of selecting the utterance marked with the direct past tense (-*dı*) in Target trials. The dependent variable was binary values (0,1) for selecting the utterance marked with the direct past tense (-*dı*). Subjects (ID) and Items (EVENT) were included as crossed random intercepts and Condition (Seen, Inferred) and Age (Adults, 3-year-olds, 4-year-olds, 5- to 6-year-olds) were included as fixed factors. The analysis revealed a main effect of Condition ($p < .001$) and an interaction between Condition and Age ($p < .001$). Follow up analyses on the Age by Condition interaction with separate mixed models on selecting the utterance marked with the direct past tense for each Age group with Subjects and Items as random intercepts and Condition as the fixed factor revealed that the likelihood of selecting the utterance marked with the direct past tense (-*dı*) as a better description of the event differed as a function of whether the event was Seen or Inferred, but only for Adults ($p < .001$). Children from all three age groups were equally likely to pick the utterance marked with the direct past tense (-*dı*) as a better description of either Seen or Inferred events (all $p > .05$) and therefore did not differentiate between the evidential function of the two past tense morphemes.

Table 15 *Fixed effect estimates for multi-level model of selecting the utterance marked with the direct past tense in Target trials in Experiment 6*

Effect	Estimate	SE	t value
Intercept	0.55	0.03	17.69***
Condition (Inferred vs. Seen)	0.25	0.06	4.23
Age (Adults vs. 3-year-olds)	-0.12	0.09	-1.39
Age (Adults vs. 4-year-olds)	0.08	0.09	0.85
Age (Adults vs. 5- to 6-year-olds)	0.05	0.09	0.61
Condition (Inferred vs. Seen) : Age (Adults vs. 3s)	-0.78	0.18	-4.36***
Condition (Inferred vs. Seen) : Age (Adults vs. 4s)	-0.67	0.18	-3.68***
Condition (Inferred vs. Seen) : Age (Adults vs. 5 to 6s)	-0.78	0.18	-4.36***

* $p < .05$
** $p < .01$
*** $p < .001$
Formula in *R*: Direct ~ CONDITION * AGE + (1 | ID) + (1 | EVENT)

Furthermore, adults selected the utterance marked with the direct past tense (-*dı*) at levels significantly different from chance when presented with a Seen ($t(6)=9.61$, $p < .001$) or an Inferred ($t(6) = -3.59$, $p = .01$) event. However, children's selection of the utterance marked with the direct past tense (-*dı*) did not differ from chance performance for both Seen and Inferred events (all $p > .05$).

Discussion

In Experiment 6, we considered the possibility that Turkish learners' evidential comprehension might improve in a contrastive task with lower demands, as predicted by the methodological account. An alternative possibility was that children's difficulty in evidential comprehension would persist in such a task, as long as children still had to reason about the meaning of evidential morphemes with respect to someone else's evidence for information. Our results supported the second possibility: children lacked a consistent preference for mapping the direct (–*dı*) or indirect past tense (–*mış*) to the relevant type of access to an event (perception vs. inference from observables). Finally, as in Experiment 5, children were highly successful in Filler trials and thus did not have a general difficulty in making comparative judgments. Together, our findings provide initial evidence against the methodological explanation of the production-comprehension asymmetry in the domain of evidentiality.

Experiment 7

The goal of Experiment 7 was to further test the theoretical accounts of the delay in evidential comprehension. According to the methodological account, reducing task demands further could reveal success in evidential comprehension that might have been masked by the comprehension tasks of Experiments 5 and 6. In this study, we presented children with an accessible (seen or inferred) event and a "mystery" event. One puppet had access to the accessible event and the other had access to the "mystery" event. Children then heard an utterance that described the accessible event but contained either a correct or an incorrect evidential; their task was to attribute the utterance to one of the puppets. To succeed, children needed to pick the accessible event when there was an evidential match but choose the inaccessible ("mystery") event in case of an evidential mismatch. Since there was no information about the inaccessible event, this task essentially relied on a single comparison between the evidential utterance and the accessible event. In this respect, it was simpler than prior paradigms that involved comparing a single evidential utterance to two events (Experiment 5) or two evidential utterances to a single event (Experiment 6). Thus the methodological hypothesis predicts that performance in evidential comprehension should improve compared to prior studies. By contrast, the psycholinguistic explanation predicts that performance should remain the same, since this task still requires reasoning about the speaker's knowledge source.

Method

Participants

Data were collected from a new group of native speakers of Turkish across two age groups: 4-year-olds (*n* = 11, 3;10-4;11, M_{age} = 4;6) and 5- to 6-year-olds (*n* = 9, 5;3-5;11, M_{age} = 5;7). We only looked at older children because they were more likely to benefit from task modifications. Children were recruited through preschools in Istanbul, Turkey.

Stimuli

We refilmed the Seen and Inferred versions of the 6 Target events in Experiments 5 and 6 with the same female experimenter who acted out the events in the earlier experiments; the only difference was that we added a puppet who was outside the puppet theater and gained access to the events. We supplemented them with Seen and Inferred versions of 6 additional events (with puppet observers) that were similar to the first set (see Appendix C for event lists). For the total set of 12 events (each with two versions: Seen and Inferred), we used 4 different puppets (a penguin, a cat, a bunny, and a squirrel) as the observer of the event. These events were accessible from both the puppet's and the child participant's perspective, and thus will be referred to as the accessible event. Each puppet served as the observer of the accessible event 3 times.

We also filmed a set of 4 "mystery" events. In each of these, one of the four puppets peeked behind the curtains of the puppet theater. The puppets were the same ones that served as the observers for the accessible events. These mystery events led to some unspecified knowledge on the part of the puppet (but were inaccessible to child participants). To create test trials, we paired one accessible event (in either the Seen or the Inferred version) with one "mystery" event on a single screen (see Figure 4). We used the following constraints. First, accessible events were always placed on the left and inaccessible events on the right side of the screen (and unfolded in that sequence). Second, within a trial, the puppet in the accessible event was always different from the puppet in the inaccessible event. Third, a given puppet pair within a trial (e.g., cat and bunny) was repeated only once but the assignment to the accessible vs. inaccessible event was switched the second time. Finally, we created two practice trials. These were similar to the test trials but the accessible events consisted of videos of the same agent holding an object (cf. the Filler trials of Experiment 6).

Two lists were created such that each version of an accessible Target event (Seen, Inferred) was assigned to one of the lists. Each list thus contained 12 test trials, with 6 of the accessible events were Seen and 6 of the accessible events were Inferred. "Mystery" (inaccessible) events, as well as the on-screen pairings of Target and mystery events were the same across lists. Each list was arranged in the same random order. The two practice trials were placed at the beginning of each list. Each participant saw one of the two lists.

Procedure

Each child was tested individually in a quiet room at his/her preschool with a 13-in. MacBook Pro laptop. As described above, on a given test trial the accessible event was presented on the left side of the screen and the inaccessible event was presented on the right side of the screen. The accessible event was either Seen or Inferred. The experimenter gave the following instructions: "Look, there are two screens here. A girl will be playing some games on these screens and these puppets will be watching. We can only see what the girl is doing in one of these screens. The curtains will be closed in the other one, but this puppet can look behind the curtains and watch the girl." Then, the experimenter played the videos one by one and presented children an evidentially marked utterance and asked: "Who said it?"

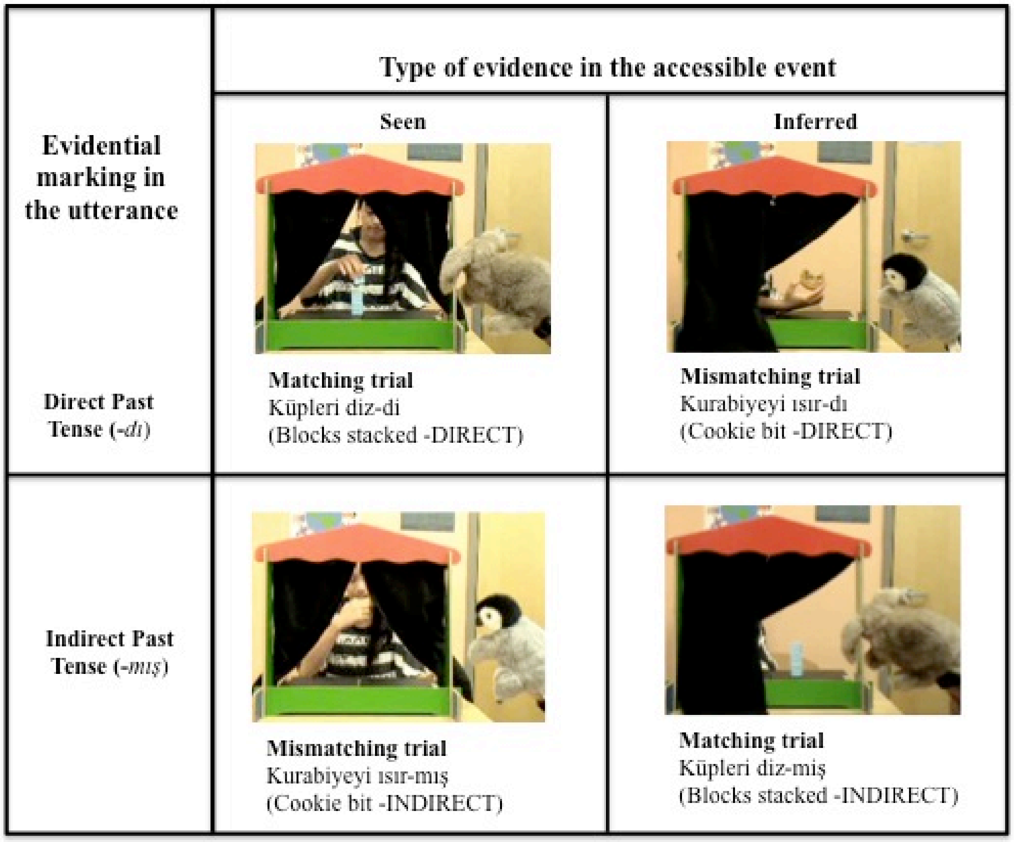

Figure 4 Examples of Matching and Mismatching Trials where the Accessible Event was Seen (right panel) or Inferred (left panel).

The main proposition in the utterance always matched the action in the accessible event but the evidential aligned with the puppet's (and the child's) access to this event only half of the time. Figure 4 presents examples of Matching and Mismatching Evidential trials for when the accessible event is Seen or Inferred. For example, when the accessible event involved the puppet *seeing* a girl stack blocks, the direct past-tense (cf. top-left panel) was a semantic match. But when the accessible event involved the puppet *seeing* that a girl ate a cookie, the indirect past-tense (as in bottom-left panel) was a mismatch. Conversely, if the accessible event involved the puppet *inferring* that a girl ate a cookie the direct past tense would be a mismatch (cf. top-right panel); but when the accessible event involved the puppet *inferring* that a girl stacked some blocks then the indirect past tense would be a match (as in bottom-right panel).

The order of trials was semi-randomized with the constraint that a given type of trial (Matching or Mismatching) did not repeat more than twice. There were 6 Matching and 6 Mismatching trials in total (with 3 Seen and 3 Inferred events within each type). The assignment of events to evidential matches vs. mismatches was fixed (see Appendix C).

As mentioned earlier, there were two practice trials in the beginning of the experiment. The accessible events consisted of videos of the same agent holding an object. The instructions were exactly same as the main experiment but children heard an utterance identifying an object (e.g., “There is a giraffe”) and were asked “who said it”. In the first practice trial, the label matched the identity of the object in the accessible event but in the second it did not. After children responded in each practice trial, the curtains of the theater in the mystery event were opened so the children could receive feedback about the accuracy of their response. The experimenter then told children: “Look, even if we cannot see what is behind the curtains, the puppet is seeing something”. This was done to confirm that inaccessible events were viable choices.

If children differentiated the evidential meaning of the two past tense morphemes, the likelihood of picking the accessible event should differ as a function of the evidential marking in the utterance. In other words, the children should (a) pick the accessible event when they heard an evidential that matched the type of evidence that the puppet had in that event (Matching Evidential trials) and (b) avoid picking the accessible event when they heard an evidential that did not match the type of access that the puppet had in the event and instead pick the inaccessible event (Mismatching Evidential trials). However, if children did not differentiate the two past tense markers on the basis of evidentiality, then the likelihood of picking the accessible event should not change across Matching Evidential and Mismatching Evidential trials. Since the base sentence (minus the evidential) always described the action in the accessible events correctly, if children failed to integrate evidential meaning into sentence interpretation, they might pick the accessible event regardless of whether the utterance had a Matching or a Mismatching Evidential.

Results

Table 16 presents the proportion of selecting the accessible event across types of evidence presented in the accessible event (Seen, Inferred) and type of evidential marking in the utterance (Matching, Mismatching) presented to the children across age groups.

Table 16 *Proportion of selecting the accessible event across types of evidence and evidential marking in Experiment 7*

		4-year-olds		5-year-olds	
		M	SE	*M*	SE
Seen					
	Matching Evidential	0.56	0.14	0.81	0.16
	Mismatching Evidential	0.64	0.14	0.78	0.14
Inferred					
	Matching Evidential	0.64	0.15	0.63	0.13
	Mismatching Evidential	0.53	0.15	0.67	0.16

Table 17 Fixed effect estimates for multi-level model of selecting the accessible event in Experiment 7

Effect	Estimate	SE	t-value
Intercept	0.65	0.06	10.82***
Evidential Type	-0.01	0.05	-0.13
Evidence Type	-0.03	0.05	-0.57
Age	-0.13	0.12	-1.09
Evidential Type * Evidence Type	-0.16	0.10	-1.54
Evidential Type * Age	-0.01	0.10	-0.13
Evidence Type * Age	-0.13	0.10	-1.28
Evidential Type * Evidence Type * Age	-0.27	0.21	1.28

* $p < .05$
** $p < .01$
*** $p < .001$
Formula in *R*: Accessible ~ EVIDENTIAL TYPE * EVIDENCE TYPE * AGE + (1 | ID) + (1 | EVENT)

As in previous experiments, data were analyzed with multi-level mixed logit modeling with REML estimates of the parameters. The same model fitting procedures of Experiments 4-6 were used. The dependent variable was binary values (0,1) for selecting the accessible event. Subjects (ID) and Items (EVENT) were included as crossed random intercepts. Evidence Type (Seen, Inferred) in the accessible event, Evidential Type (Matching, Mismatching) and Age (4-year-olds, 5-year-olds) were included as fixed factors. Because all three variables (Evidence Type, Evidential Type and Age) had two levels each, fixed effects for each variable were coded using centered contrasts (-0.5, 0.5).

Table 17 presents a summary of fixed effect estimates for the multi-level model of selecting the accessible event. The model revealed only a significant intercept ($p < .001$). No other effects or interactions were significant. Both age groups were more likely to pick the accessible event compared to the inaccessible event. Furthermore, children's likelihood of picking the accessible event did not change depending on whether the evidential marking in the utterance matched the type of evidence in the accessible event. Because the basic level proposition in the utterance (minus the evidential) correctly described the action in the accessible event, this response pattern suggests that children neglected the evidential and responded based on whether the rest of the sentence matched the accessible event. Importantly, the likelihood of picking the accessible event was the same regardless of whether the accessible event was Seen or Inferred, suggesting that children did not have a bias to pick either a Seen or an Inferred accessible event.

This conclusion is also supported by the fact that the selection of the accessible event was above chance level the evidential marking in the utterance matched the type of evidence in the accessible event ($t(21) = 2.63$, $p = .02$); and approached *above* chance levels the evidential marking in the utterance did not match the type of evidence in the accessible event ($t(20) = 2.01$, $p = 0.06$), while it should have been *below* chance level.

Discussion

Experiment 7 introduced an evidential comprehension task that did not require comparing two different evidentially marked utterances or two different sources of information for the same event. Participants only had to compare a single evidential statement to the way a puppet gained access to an event; children themselves had the same access to the event as the puppet. If there was a mismatch between the statement and the puppet's experience, children could attribute the statement to another puppet who gained access to an inaccessible ("mystery") event. Despite this difference from earlier tasks, the findings of Experiment 7 cohere with those of Experiments 5 and 6:

Turkish-speaking children showed no sensitivity to evidential meaning. Even when the evidential misrepresented the evidence present in the accessible event, children did not reliably switch to the inaccessible event. Together, the results from Experiments 5, 6 and 7 show that the delay in comprehension of evidential morphology in Turkish persists across alternative tasks with varying demands. This conclusion is unexpected on the hypothesis that evidential comprehension difficulties are tied to methodological factors but is entirely consistent with the psycholinguistic account of how evidentials are understood.[4]

Experiment 8

In Experiment 8, we turned to the conceptual underpinnings of linguistic evidentiality. We used the paradigm of Experiment 7 but replaced the evidentially marked utterances with non-past tense (infinitival) verb forms that either matched or did not match the accessible event. We used these verbs to ask whether Turkish learners can evaluate whether the evidence available to someone was sufficient for them to be knowledgeable about an event. We were especially interested in whether children's ability to reason about evidence for information differs when they reason about others' knowledge (Others task) and their own knowledge (Self task).

The outcome of this study bears directly on explanations of the delay in the comprehension of linguistic evidentials in Turkish learners. According to the methodological account, children might perform better in a knowledge-attribution task that does not involve evidential language. Importantly, this account does not predict any difference in performance depending on whether or not the task involves a Self-oriented or Others-oriented perspective. By contrast, according to the psycholinguistic account, children's performance should remain poor in a knowledge-attribution task that does not involve evidential language as long as that task requires reasoning about others' information sources. However, performance should improve when the perspective-taking component is removed and children have to reason about their own information sources.

Method

Participants

In the Others task, participants were 10 4-year-old (M_{age} = 4;6, 3;10-4;11) and 13 5- to 6-year-old (M_{age} = 5;8, 5;0-6;4) native speakers of Turkish. In the Self task, participants were 12 4-year-old (M_{age} = 4;6, 4;0-4;11) and 13 5- to 6-year-old (M_{age} = 5;7, 5;1-6;5) native speakers of Turkish. Participants were recruited through preschools in Istanbul, Turkey.

Stimuli

The visual stimuli were the same as in Experiment 7. A set of changes was made to the verbal stimuli (see Figure 3 for a summary). The evidentially marked utterances were replaced with verbs in the infinitive form ("to V", broadly equivalent here to the –*ing* form in English). We chose this form because it is the "unmarked" form of the verb that allows reference to events without encoding tense, aspect, or evidentiality. The resulting sentences were grammatical in Turkish. The Matching and Mismatching trials were kept the same, but instead of manipulating the match between evidential marking and type of access for the accessible event, we manipulated the match between the verb content and the accessible event. For Matching trials, the verbs used were the ones in the evidentially marked utterances of Experiment 7 (e.g., when the accessible event involved either seeing a stacking action or inferring that stacking had occurred, children were asked about "to stack"/"dizmek"). For the Mismatching trials, we devised verbs that would clearly be incorrect if applied to the accessible event (e.g., when the accessible event involved either seeing a biting action or inferring that biting had occurred, children were asked about "to wash"/"yıkamak"). The list of events and Matching vs. Mismatching Verbs is presented in Appendix C.

Procedure

The procedure was the same as Experiment 7 with the following exceptions. In the Others task, children's task was to find which of the two puppets knew about an event. Thus, at the end of each trial, the experimenter pointed to the two puppets and asked: "If you want to find out more about <infinitive>, which puppet should you ask?" The rationale was that children should (a) pick the puppet observing the accessible event when they were asked about a verb that matched that event (Matching Verb trials), and (b) avoid picking the puppet observing the accessible event and instead pick the puppet observing the inaccessible alternative when the verb did not match the accessible event (Mismatching Verb trials). In the Self task, the procedure was exactly the same but at the end of the trial, the experimenter pointed to the two videos and asked: "Which one has <infinitive>?" The rationale for responding was similar to the Others task but involved answering from one's own perspective instead of adopting the puppets' perspective.

Results

Table 18 presents the proportion of selecting the accessible event across types of evidence in the accessible event and types of verb in the Others and Self Tasks.

Data from both tasks were analyzed using mixed logit modeling with REML estimates of the parameters. The models were fitted using the same procedures as in previous experiments. The dependent variable was binary values (0,1) for selecting the accessible event. Subjects (ID) and Items (EVENT) were included as crossed random intercepts. Evidence Type (Seen, Inferred) in the accessible event, Verb Type (Matching, Mismatching) and Age (4-year-olds, 5-year-olds) were included as fixed factors. As in Experiment 7, all fixed factors were coded using centered contrasts (-0.5, 0.5).

Table 18 *Proportion of selecting the accessible event across evidence and verb types in the Others and Self Tasks in Experiment 8*

		Others Task				Self Task			
		4-year-olds		5-year-olds		4-year-olds		5-year-olds	
		M	SE	*M*	SE	*M*	SE	*M*	SE
Seen									
	Matching Verb	0.63	0.15	0.69	0.13	0.72	0.13	0.74	0.12
	Mismatching Verb	0.40	0.16	0.51	0.14	0.23	0.12	0.36	0.13
Inferred									
	Matching Verb	0.63	0.15	0.67	0.13	0.82	0.11	0.82	0.11
	Mismatching Verb	0.47	0.16	0.38	0.14	0.46	0.14	0.26	0.12

Table 19 provides a summary of the fixed effect estimates for multi-level model of selecting the accessible event in the Others Task. The model revealed only a main effect of Verb Type ($p < .05$). The likelihood of selecting the accessible event was higher for Matching Verbs ($M = 0.66$) than for Mismatching Verbs ($M = 0.44$).

No other effects or interactions were significant. Furthermore, the proportion of selecting the accessible event was significantly above chance level for Matching Verbs ($t(22) = 2.72$, $p = .01$) but did not differ from chance level for Mismatching Verbs ($t(22) = -1.04$, ns).

Table 19 *Fixed effect estimates for multi-level model of selecting the accessible event in the Others Task in Experiment 8*

Effect	Estimate	SE	t-value
Intercept	0.55	0.05	11.34***
Verb Type	-0.21	0.07	-3.09*
Evidence Type	-0.02	0.06	-0.4
Age	-0.03	0.09	-0.4
Verb Type * Evidence Type	-0.02	0.11	-0.14
Verb Type * Age	0.03	0.11	0.28
Evidence Type * Age	0.11	0.11	0.95
Verb Type * Evidence Type * Age	0.18	0.22	0.81

* $p < .05$
** $p < .01$
*** $p < .001$
Formula in *R*: Accessible ~VERB TYPE * EVIDENCE TYPE * AGE + (1 | ID) + (1 | EVENT)

Performance in the Others task was compared to performance in the evidential comprehension task of Experiment 4 using multi-level mixed logit modeling. The dependent variable was binary values (0,1) for selecting the accessible event; Subjects (ID) and Items (EVENT) were included as crossed random intercepts and Experiment (4, 5/Others) and Trial Type (Matching, Mismatching) were included as fixed factors. Including Age and/or Evidence Type as fixed factors did not reliably improve fit based on a chi-square test of the change in -2 restricted log likelihood compared to the model that included Experiment and Trial Type the fixed factors (all $p > .05$). Unsurprisingly, the model returned an Experiment by Trial Type interaction ($\beta = -0.21$, SE = 0.08, $p = .006$): the likelihood of selecting the accessible event did not differ between Matching ($M = 0.65$) and Mismatching ($M = 0.64$) Trials in the comprehension task of Experiment 7, but the likelihood of selecting the accessible event was significantly lower for Mismatching trials ($M = 0.44$) compared to Matching trials ($M = 0.66$) in the Others task of Experiment 8. Crucially, there was no effect of Experiment ($p = 0.9$) suggesting that the overall performance did not improve after evidential language was removed.

Table 20 *Fixed effect estimates for multi-level model of selecting the accessible event in the Self Task in Experiment 8*

Effect	Estimate	SE	t-value
Intercept	0.55	0.05	11.81***
Verb Type	-0.45	0.08	-5.42***
Evidence Type	0.08	0.05	1.71
Age	0.01	0.06	0.2
Verb Type * Evidence Type	-0.04	0.09	-0.37
Verb Type * Age	0.05	0.09	0.54
Evidence Type * Age	0.18	0.09	1.9
Verb Type * Evidence Type * Age	0.31	0.18	1.63

* $p < .05$
** $p < .01$
*** $p < .001$
Formula in *R*: Accessible ~VERB TYPE * EVIDENCE TYPE * AGE + (1 | ID) + (1 | EVENT)

Table 20 presents the fixed effect estimates for multi-level model of selecting the accessible event for the Self Task. As expected, the model revealed a main effect of Verb Type (*p* < .001). Both age groups were more likely to select the accessible event when asked about a Matching Verb (*M* = 0.77), compared to when they were asked about a Mismatching Verb (*M* = 0.31) for both Seen and Inferred events. As in the Others Task, there were no significant effects or interactions of Evidence Type (Seen, Inferred) and/or Age (4-year-olds, 5-year-olds). Also as expected, proportion of selecting the accessible event was significantly above chance level for Matching Verbs ($t(24)= 6.35$, $p < .001$) and below chance level for Mismatching Verbs ($t(24) = -3.81$, $p = .001$).

Performance in the Others and Self tasks were compared with a multi-level mixed logit model that had selecting the accessible event at the trial level as the dependent variable, Subjects (ID) and Items (EVENT) as crossed random intercepts and Task (Others, Self) and Verb Type (Matching, Mismatching) as fixed factors. Models that included Age and/or Evidence Type as fixed factors did not reliably improve fit based on a chi-square test of the change in -2 restricted log likelihood compared to the model that included Experiment and Trial Type the fixed factors (all $p > .05$) so they were omitted from the model. The model revealed a main effect of Verb Type ($\beta = -0.34$, SE = 0.06, $p < .001$) and an interaction between Task and Verb Type ($\beta = -0.25$, SE = 0.07, $p < .001$). In both tasks, the likelihood of selecting the accessible event was higher for Matching Verbs compared to Mismatching Verbs. However the difference in the likelihood of selecting the accessible event between Matching and Mismatching Verbs in the Self task (0.77 vs 0.31, respectively) was larger than the same difference in the Others task (0.66 vs. 0.44, respectively). That suggests that children were much more successful in differentiating between the Matching and Mismatching Verbs in the Self task than in the Others task, leading to better overall accuracy in the Self task (*M* = 0.73) compared to the Others task (*M* = 0.61). In fact, a similar mixed logit model that had accuracy at the trial level as the dependent variable only revealed a main effect of Task ($\beta = 0.12$, SE = 0.05, $p = .008$), confirming that children's performance improved in the Self task.

Discussion

Experiment 8 used the design of Experiment 7 but replaced the comprehension task involving evidential morphology with a task that assessed whether children understand how evidence and knowledge are linked in others (Others task) or oneself (Self task). In the Others task, children had difficulty linking evidence and event representations in others' minds; their performance was similar to Experiment 7. In the Self task, however, where children were asked to link evidence to event representations in their own mind, performance improved. Experiment 8 alleviates a concern with the design of Experiment 7, since it shows that children's failures in the earlier study were not simply due to a dispreference for inaccessible ("mystery") events. Most importantly, together with Experiment 7, the findings of Experiment 8 support the position that the comprehension lag in the domain of linguistic evidentiality is not explained by methodological factors but seems to be due to the psycholinguistic process of linking evidentials to others' knowledge sources.

General Discussion

Young learners typically comprehend the meaning of linguistic forms before they produce these forms in speech. However, cases where production precedes accurate comprehension have also been reported in various domains. In the studies reported here, we examined evidentiality, a domain that is known to give rise to a production-comprehension asymmetry across languages (Aksu-Koç, 1988; Ozturk & Papafragou, 2015; Papafragou et al., 2007; Rett & Hyams, 2014; Winans et al., 2014). Focusing on evidential morphology in Turkish, we sought to systematically explore the theoretical explanations of this asymmetry. Below we summarize our main findings and sketch their implications for the acquisition of evidentiality, as well as the relation between production and comprehension more generally.

Production and Comprehension of Evidential Morphology in Turkish

Turkish encodes evidential meanings in its past-tense morphemes -*dı* (direct) and -*mış* (indirect evidence; Aksu & Slobin, 1986; Johanson, 2003; Kornfilt, 1997; Slobin & Aksu, 1982). Prior work reported that the acquisition of these morphemes was protracted, with some aspects of evidential meaning being inaccessible even at the age of 6 or 7 (Ozturk & Papafragou, 2007, 2015; Papafragou et al., 2007; Ünal & Papafragou, 2013); furthermore, Turkish learners were reported to be more accurate in producing these morphemes than in understanding them in the speech of others (e.g., Aksu-Koç, 1988; Ozturk & Papafragou, 2015). In a first set of experiments, we revisited the age of acquisition of the Turkish evidential morphemes. We compared the direct morpheme (instantiated in scenes involving visual perception of an event) and the inferential interpretation of the indirect morpheme (instantiated in scenes involving backward causal inference about an event based on visual cues). To highlight the difference between seeing vs. inferring an event, we used naturalistic stimuli that included salient cues about how an event was accessed (i.e., a puppet theater where curtains remained open throughout the event vs. closed halfway). Unlike prior studies, we also used the same events to test production and comprehension of evidentials.

In Experiment 4, young learners of Turkish between the ages of 3 and 6 successfully differentiated the two past-tense morphemes on the basis of their evidential function. That is, they were more likely to mark the events they saw with the direct past tense (-*dı*) morpheme as opposed to the indirect past tense (-*mış*) morpheme. Conversely, they marked the events they inferred on the basis of available evidence with the indirect past tense (-*mış*) morpheme as opposed to the direct past tense (-*dı*) morpheme. It should be noted that children sometimes extended the direct past tense to non-witnessed events (see also Ozturk & Papafragou, 2015 for similar results with older children). For instance, the oldest group of 5- to 6-year-olds marked their descriptions of Inferred events with the direct past tense (-*dı*) morpheme for about 28% of the time. The percentages for 3- and 4-year-olds were much lower (7% and 15%, respectively) but these younger groups were also more likely to mark their descriptions with other non-past tense morphemes. These data suggest that the development of evidential production is not completed yet. Nevertheless, our findings lower prior estimates of evidential production in past work (Aksu-Koç, 1988; Ozturk & Papafragou, 2007, 2015; Ünal & Papafragou, 2013), presumably because of the contextual cues that increased the salience of the difference between inference and perception (cf. also Aksu-Koç, 1988, Aksu-Koç et al., 2009, for discussion of this point).

Yet findings of Experiment 5 demonstrated that, when children of similar ages had to match an evidentially marked utterance to one of two versions of the same event that differed only in terms of how the event was accessed (perception vs. inference), performance was not as robust. Even the oldest group of 5- to 6-year-old children performed worse in evidential comprehension compared to adults who were at ceiling. These data confirm the previously documented production-comprehension asymmetry in the domain of evidentiality and show that this asymmetry emerges even for the same events.

Origins of the Production-Comprehension Asymmetry

In the remaining set of experiments, we evaluated two alternative explanations of the asymmetry between evidential production and comprehension. According to the methodological hypothesis, the lag between evidential production and comprehension could be attributed to the higher memory, metalinguistic or other processing loads of comprehension compared to production tasks. According to the psycholinguistic hypothesis, evidential comprehension is delayed by the difficulty of embedding evidential forms into reasoning about someone else's information sources.

Each of the two hypotheses makes a unique prediction that is not shared by the other account. The methodological hypothesis uniquely predicts that Turkish learners' comprehension might improve in comprehension tasks with lower memory and processing demands. Contrary to this prediction, however, the difficulty with evidential comprehension persisted across several alternative tasks. Specifically, evidential comprehension was comparable across Experiment 5 (where a single evidential utterance had to be compared to two events), Experiment 6 (where two evidentially marked utterances had to be compared to one event) and Experiment 7 (where a single evidential utterance had to be compared to one event).

Turning to the psycholinguistic hypothesis, this hypothesis uniquely predicts that the asymmetry between evidential production and comprehension might also emerge in a non-evidential context, as long as the Self-Other perspectives inherent in encoding vs. decoding evidential meanings remained constant. This prediction was confirmed: Experiment 8 revealed that children's difficulty persisted in a version of Experiment 7 in which children had to reason about others' knowledge sources (Others task), but their performance improved when they had to reason about evidence for the same events from their own perspective (Self task).

Together, these results point to the conclusion that Turkish learners' difficulties with the comprehension of evidential morphology is tied to the perspective-taking demands of considering other cognizers' access to information. This conclusion coheres with prior developmental studies showing that linguistic knowledge of evidentiality builds on and closely follows conceptual knowledge about sources of information, and suggests an even tighter and more specific relation between linguistic knowledge of an evidential system and children's developing abilities to handle various information sources compared to those prior reports (Papafragou et al., 2007; Ozturk & Papafragou, 2015). This conclusion is likely to generalize to other languages beyond Turkish where the asymmetry has been observed (e.g., Korean, Papafragou et al., 2007; English, Winans et al., 2014). Naturally, the acquisition of the Turkish evidential system involves multiple factors, including mastering the complexities of mapping multiple information sources in the world to the more abstract, two-way direct-indirect distinction in the past tense (Ozturk & Papafragou, 2015). Nevertheless, our results clearly show that an important part of developing adult-like knowledge of the evidential system involves navigating the difference inherent in speaking about information access in one's own mind (in production) and other minds (in comprehension).

Our results are also consistent with prior developmental studies showing that, when judging the knowledge state of others, young children often fail to consider others' informational access, despite the fact that they can gain knowledge from different types of information sources such as visual access or verbal report themselves (Pillow, 2002; Pillow & Anderson, 2006; Pillow et al., 2000; Povinelli & deBlois, 1992; Sodian & Wimmer, 1987; Wimmer et al., 1988). For instance, in one study, when asked whether another child who had looked inside a container knew what was inside the container, the vast majority of the 3-year-olds and half of the 4-year-olds either overattributed knowledge to the second child or denied the second child knowledge altogether (Wimmer et al., 1988). By contrast, all of the 5-year-olds selectively attributed knowledge to the second child based on the second child's informational access. In another demonstration, 4-year-olds were able to draw a logical inference about the color of a set of balls but failed to attribute the same piece of knowledge to an adult, even though the adult had access to the critical premise that would allow her to draw the very same inference (Sodian & Wimmer, 1987). Six-year-olds in the same study correctly identified the adult as knowledgeable about the balls' color when she had access to the critical premise. In subsequent experiments, 4-year-olds kept neglecting the adult's inferential knowledge even when the adult shared the child's perspective (Sodian & Wimmer, 1987), or when the children were reminded of the critical premise that the adult knew (Keenan, Ruffman, & Olson, 1994; Pillow, 1989).

One question left open by the current findings is how children learn to produce the evidential morphemes of Turkish in the first place. After all, some understanding of the meaning of these forms in the speech of others is necessary for children to acquire productive command of them. We speculate that children begin with a coarse distinction between direct and indirect sources and implicitly associate these information sources with the corresponding evidentiality markers. However, this coarse distinction might not be sufficient to fully master the conditions under which these morphemes are used (and cannot support the more nuanced and specific reasoning about others' sources required to pass the present and all previous comprehension tasks).

Viewed within the broader context of the relation between speaking and understanding speech, our findings are consistent with the idea that the mechanisms involved in production and comprehension do not simply involve the same steps executed in the reverse order (Hurewitz et al., 2000). In production, speakers plan a message to convey an intended meaning (in this case, one's own informational access) through a particular form (evidentiality markers). In comprehension, the listener must unpack the meaning carried by the forms in incoming speech (and, in doing so, entertain concepts about information access in other minds). Even though in adults, these two processes are choreographed to align closely, in children, they diverge until independent developments in children's perspective-taking allow them to co-ordinate.

Finally, the present results have implications for the joint study of language production and comprehension in children, especially for cases where production seems to emerge before successful and complete comprehension. For instance, our approach suggests that task analysis, stepwise manipulations of task demands and comparisons to non-linguistic versions of the same experiments can help ascertain both the boundaries of comprehension difficulties and the theoretical nature of these difficulties.

Chapter 4

CONNECTING EVIDENCE TO KNOWLEDGE

Introduction

The experiments reported in Chapter 3 revealed an asymmetry in Turkish children's understanding of information sources for themselves and others that persist across linguistic and non-linguistic contexts. One question left open by the findings of Chapter 3 is at what age children become more successful in their knowledge attributions and perform close to adult levels. Since the experiments that assessed non-linguistic representations of sources had a limited age range of 4- and 5-year-olds, this developmental question could not be fully explored in Chapter 3. Thus, in the present chapter, we aim to gain a deeper understanding of development of source knowledge in Turkish children focusing on a wider age range, including older children. We also aim to address two limitations of the current literature on the development of source knowledge. Before we lay out the details of those issues, we provide a short review of the developmental literature on source monitoring.

Empirical evidence on children's understanding of knowledge sources comes predominantly from two lines of research. One line of studies asks children to judge whether they themselves or others are knowledgeable about the identity of an item based on some kind of informational access. These studies have shown that, in simple tasks, even young children seem to grasp the connection between seeing and knowing: 3-year-olds select the character who had visual access to an object hidden inside a box as the one who knows what is hidden inside the box over another character who simply lifted or pushed the box (Pillow, 1989; cf. Pratt & Bryant, 1990; see also Povinelli & deBlois, 1992; Wimmer, Hogrefe & Perner, 1988). Younger children's ability to link seeing to knowing in others improves when they share others' perceptual access (Ruffman & Olson, 1989) or receive training on how to interpret the task (Pratt & Bryant, 1990).

Nevertheless, understanding other types of access comes later: children do not fully realize that inference can be a source of knowledge until at least the age of 6. In one study (Sodian & Wimmer, 1987), the child and an adult observed an experimenter fill a transparent container with balls of the same color. Then, the experimenter transferred the balls to an opaque container. Importantly, this second transfer was observed by either the child or the adult. Both 4- and 6-year-olds were able to infer the color of the balls even when they did not see the transfer themselves using the critical premise. However, when asked whether the adult knew the color of the balls in the opaque container, 4-year-olds denied the adult's inferential knowledge. Six-year-olds, however, identified the adult as knowledgeable about the balls' color both when she did and when she did not see the transfer (but knew the critical premise). In subsequent experiments, 4-year-olds denied attributing inferential knowledge to the adult even when the adult shared the child's perspective (Sodian & Wimmer, 1987). Other studies have found that children become able to understand more subtle distinctions among types of inference and fully explain why these kinds of inferences lead to knowledge in the primary school years or sometimes even later (Pillow, 1999, 2002; Pillow & Anderson, 2006; Pillow, Boyce, & Stein, 2000; see also Brosseau-Liard & Birch, 2011; Fitneva, Lam, & Dunfield, 2013; Wimmer, Hogrefe, & Sodian, 1988; Keenan, Ruffman, & Olson, 1994; Ozturk & Papafragou, 2015; Robinson, Haigh, & Nurmsoo, 2008; Varouxaki, Freeman & Peters, 1999). Other studies have confirmed that 6-year-olds understand the role of evidence: 4- and 5-year-old children but not 6-year-olds often claim that an observer shown a small, uninformative region of a drawing will know what the drawing depicts (Taylor, 1988). Furthermore, 6- and 7-year old children are able to recognize that evidence can provide a basis for inference about unseen events; they are also able to evaluate the informativeness of potential evidence in deciding between a pair of hypotheses (Sodian, Zaitchik, & Carey, 1991). However, even children of this age do not always distinguish justified inference from mere guessing in others (Pillow, Hill, Boyce, & Stein, 2000).

A separate line of studies has looked at children's ability to state "how they know" a piece of information. This paradigm have revealed that by age 3 children can accurately report the knowledge that they themselves have acquired (e.g., they can say what is in a tunnel after looking inside, being verbally informed, feeling inside, or figuring out the contents from a clue; Gopnik & Graf, 1988); however, children of this age, unlike older children, have difficulty reporting how they acquired such knowledge (ibid.; cf. O'Neill & Gopnik, 1991; Pillow, 1989; Pillow et al., 2000; Wimmer et al., 1988; Wooley & Bruell, 1996). These studies found that inference in particular was identified as a source later than perception and communication. Other tasks using deductive inference show that the tendency to explain one's own knowledge by spontaneous appeal to inference becomes robust only after the age of 5 (Sodian & Wimmer, 1987). In a study particularly relevant for present purposes, Ozturk and Papafragou (2015) found that 5-year-old Turkish children were able to successfully report that they had seen something but could not reliably report that they were told about or had inferred something until age 6. Similarly, 5-year-olds were also able to successfully report that someone else had seen something; however, even 7-year-olds had difficulty reporting that someone else was told about or had inferred a piece of knowledge.

In addition to producing many important results, prior literature on children's source monitoring has two limitations. First, as mentioned already, source monitoring paradigms typically require children to explicitly reflect on whether a certain experience led to knowledge acquisition (e.g., to say whether they themselves or someone else *knows X*; Pratt & Bryant, 1990; Sodian & Wimmer, 1987; a.o.), or, inversely, to identify which experience led to a piece of knowledge (e.g., to say whether they or others *saw/heard/figured out X*; Gopnik & Graf, 1998, and many subsequent studies). However, the use of mental-state terms in both cases might underestimate children's underlying source monitoring abilities. In support of this possibility, children have been shown to be sensitive to the distinction between a knowledgeable and ignorant interlocutor in deciding whom to trust, even when they cannot explicitly justify their choice (Robinson & Whitcombe, 2003). More generally, recent research on theory of mind suggests that the ability to explicitly report on what others know (e.g., people's false beliefs) might not accurately track the ability to reason about others' mental states (see Saxe, 2013). Thus, the linguistic material in the experimental scenarios above may present challenges to learners that are unrelated to the process of source monitoring *per se*. For instance, the difficulty of talking about one's access to inference might have to do with having acquired the verbs that refer to the act of hearing or inferring; such verbs are likely to be less transparent to children compared to the verb that refers to acts of seeing. It is possible that tasks that do not involve explicit reference to verbal report or inference but measure the basic processes of knowledge acquisition and knowledge attribution under different access conditions might reveal earlier success with some of the indirect sources that now pose the greatest difficulties for learners.

Second, our knowledge of children's understanding of different sources of knowledge mostly comes from situations in which children are asked to find out about the identity or a property of *objects* (e.g., contents of a container, color of a ball). This leaves open the question of whether the current estimates of children's source monitoring abilities generalize to other kinds of knowledge such as knowledge gained about events. Events are more complex than objects because they involve multiple participants. For example, the event of cutting may involve both multiple objects (e.g., a person that performs the action, an apple that is being cut, and a knife that is used for cutting) as well as the relational information between these components that result in the action.

In what follows, we report two experiments that study children's understanding of knowledge sources for themselves (Experiment 9) and for others (Experiment 10). Throughout, we use tasks that do not involve explicit mental state language (cf. the Self and Others tasks of Experiment 8). Our empirical focus is the understanding of visual perception and inference as sources of information about *events*. Specifically, we focus on probabilistic inferences from post-event visual premises, that elicit a backward inference to the point at which the event unfolds. Such probabilistic inferences from perceptibles are a good comparison point to visual perception because both types of informational sources are based on some visual evidence.

Experiment 9

The goal of Experiment 9 was to assess children's ability to use visual perception and inference from visual premises to gain knowledge. We presented children with photographs giving different types of access to change-of-state events. Children were given a verb that either matched or did not match the photograph, and were asked to find the picture of the verb. Of interest was whether children would be able to link the evidence provided by the photograph to the event described by the verb, and whether successful knowledge acquisition would differ depending on the type of access (i.e., perceptual, inferential).

A secondary goal was to assess whether the quality of visual evidence affects children's ability to derive inferences about events. Some visual premises might be enough to inferentially reconstruct what had happened, but other, perhaps less familiar visual premises might be less sufficient to trigger an inference about the main event. Here, we tested these possibilities by manipulating the familiarity of objects that served as post-event inferential cues.

Method

Participants

Data were collected from Turkish children in two age groups: 4-year olds (n = 16, mean age 4;6, range 4;0-4;11), and 5- to 6-year-olds (n = 16, mean age 6;0, range 5;6-6;10). Children were recruited through preschools in Istanbul, Turkey. Eight Turkish-speaking adults also participated as controls. Adults performed at ceiling level for all types of access ($M_{perception}$ = 1.00, $M_{inference\text{-}familiar}$ = 0.98, $M_{inference\text{-}unfamiliar}$ = 0.98) and thus adult results will not be reported.

Stimuli

Photographs giving different types of access to events were used as the visual stimuli. There were three types of access: Perception, Inference-Familiar, and Inference-Unfamiliar. For Perceptual access, the stimuli consisted of photographs of ongoing events including an animate agent (e.g., a woman drinking milk). The two types of indirect access differed in terms of the familiarity of objects used as visual cues. For Inference-Familiar access, the stimuli consisted of photographs of familiar objects in a state (e.g., an egg shell) that allowed the inferential reconstruction of an event (e.g., someone had cracked an egg). For Inference-Unfamiliar access, the stimuli were photographs of relatively unfamiliar objects in a state (e.g., an open walnut) that allowed the inferential reconstruction of an event (e.g., someone had cracked a walnut). Sample events for each type of access are presented in Figure 5.

1A. Perception 1B. Inference-Familiar 1C. Inference-Unfamiliar

Figure 5 Sample Stimuli for the three types of Access in Experiment 9.

There were a total of 18 events, which consisted of 6 examples of each of the three types of access. Each inferential event had two versions (Inference-Familiar and Inference-Unfamiliar; cf. Figure. 5). In order to counterbalance the familiarity of the objects for inferred events, two lists of events were created. The two versions of each inferred event were never assigned to a single list. Perceived events were the same for the two lists. The same random order of events was used for each list.

To select the verbs we would use to describe these events, and to ensure that type of access was reasonable, we asked a separate group of adults (n = 14) to describe each event with one verb and rate how certain they are that this is what has happened in the event on a 5-point scale (1-not certain at all, 5-very certain). Overall, the events were described by the matching verbs 80% of the time ($M_{perception}$ = 0.94, $M_{inference\text{-}familiar}$ = 0.79, $M_{inference\text{-}unfamiliar}$ = 0.66). Failures to use the matching verb to describe the event were mostly due to the tendency to use more general as opposed to more specific verbs. None of the events were described by the mismatching verbs that were assigned to the particular photographs. When events were identified with the matching verbs, mean certainty ratings were $M_{perception}$ = 3.92, $M_{inference\text{-}familiar}$ = 4.04, $M_{inference\text{-}unfamiliar}$ = 3.97. These means did not differ from each other.

As a final check, we consulted the Turkish Communication Inventory (TIGE) and found that 89% of the mothers of a representative sample of Turkish 3-year-olds reported that their children produced the verbs used to describe the events in our study (Aksu-Koç et al., 2011).

Procedure

Each child was tested individually in a quiet room at his/her preschool. The experimenter presented children with two upside down cards and told them that there was a picture under each card but they could look at only one of these pictures and the other card had to be upside down. Then the experimenter turned over one of the cards revealing a photograph that gave some type of access to an event (Perception, Inference-Familiar, or Inference-Unfamiliar). Then the experimenter said (in Turkish): “Now, I will tell you a word. I want you to find its picture for me. Don’t forget that there is a picture under this card (pointing to the face-down card) so, the picture for the word can be under this card as well.” Then the experimenter asked “which one has” and uttered a verb in the infinitive form (to “V”). The infinitive form of the verb is equivalent to the *–ing* form in English. We chose this form because it is an “unmarked” form of the verb that allowed us to refer to the events without encoding tense, aspect, or source of information. Furthermore, this form is already produced by 3-year-old Turkish learners according to 98.7% of the mothers surveyed in TIGE (Aksu-Koç et al., 2011). After the children made a choice, the experimenter went onto the next trial.

In half of the trials, the experimenter uttered a verb that matched the event depicted in the visible photograph (e.g., the photograph showed some egg shells and the experimenter said, “cracking”). In the other half of the trials, the experimenter uttered a verb that did not match the event depicted in the visible photograph (e.g., the photograph showed some egg shells and the experimenter said “hitting”). For each type of access, there were three matching verb and three mismatching verb trials. The assignment of verb types to events was counterbalanced across participants. For example, for the photograph showing the egg shells, half of the children heard “cracking” and the other half heard “hitting”.

At the beginning of the experiment there were three practice trials to familiarize the children with the task. For these trials, the visible photographs consisted of photographs of objects. The experimenter followed the same instructions as in the actual experiment. After the child responded in each practice trial, the experimenter gave feedback about his/her response by turning up the face-down card and showing the photograph under the card. Children were highly accurate (*M* = 0.96) in these practice trials.

If the children thought that the visible photograph provided sufficient evidence about the event encoded by the verb, then children should pick the visible photograph. Otherwise, they should pick the face-down alternative. Of interest was whether children's accuracy in linking matching verbs to visible photographs and avoiding linking mismatching verbs to visible photographs would differ across types of access and age groups. Specifically, we were interested in whether children's success in linking the events to the evidence would differ across direct/seen and indirect/inferred events. We were also interested in whether successful knowledge acquisition would be higher when the visual premises giving rise to an inference were familiar as opposed to when they were unfamiliar.

Results

Preliminary analyses showed that performance did not differ between Matching and Mismatching Verbs in either 4-year-olds (*M* = 0.81 vs. 0.83 respectively, *t*(15) = -0.21, *ns*) or 5- to 6-year-olds (*M* = 0.91 vs. 0.89 respectively, *t*(15) = 0.46, *ns*). We thus collapse across Verb Types in what follows.

For our main analysis, we assessed whether children's ability to derive information differed across Types of Access with a 3 x 2 mixed ANOVA with Access (Perception, Inference-Familiar, Inference-Unfamiliar) as the within-subjects factor and Age (4-year-olds, 5- to 6-year-olds) as the between-subjects factor (see Table 21). The dependent variable was accuracy. The analysis returned only a main effect of

Age (F(1,30)=10.915, p = .002, η^2 = .267), with 5-6-year-olds being more accurate (M = .89) than 4-year-olds (M = .82). Both groups performed at levels different from chance for all Types of Access ($p < .05$).

Table 21 *Proportion of accurate responses across types of access and age groups in Experiment 9*

	4-year-olds		5- to 6-year-olds	
	M	SE	M	SE
Perception	0.86	0.03	0.92	0.03
Inference-Familiar	0.79	0.04	0.89	0.04
Inference-Unfamiliar	0.80	0.03	0.90	0.03

Discussion

In Experiment 9 we tested whether children can identify events based on direct/perceptual evidence or post-event inferential cues. We used a novel task with minimal training and without any explicit mental state language (e.g., "Do you know…?"): in this task, children had to match a verb to either a corresponding picture or an upside-down picture. Our findings revealed a developmental difference: 5- to 6-year-olds were better than 4-year-olds in identifying events on the basis of the evidence available from the accessible picture. There was no difference in event identification between visual perception as opposed to post-event inferential evidence. Furthermore, within the class of post-event inferential evidence, children were equally successful in inferentially reconstructing events from both familiar and unfamiliar cues.

Experiment 10

The goal of Experiment 10 was to examine children's ability to attribute perception-based or inference-based knowledge to others. We used the paradigm of Experiment 9 but replaced the child's access to an event with someone else's (a puppet's) access. Furthermore, since there was no difference between the two types of Inference in Experiment 9, we only used the stimuli for Perception and Inference-Familiar events in Experiment 10.

Method

Participants

Data were collected from a new group of Turkish children in two age groups: 4-year olds (*n*=20, mean age 4;5, range 4;0–4;10) and 5- to 6-year-olds (*n*=16, mean age 6;0, range 5;6–6;10). Children were recruited through preschools in Istanbul, Turkey. Seven Turkish-speaking adults also participated as controls. As in Experiment 9, adult performance was at ceiling ($M_{perception}$ = 0.98, $M_{inference}$ = 0.94) and will not be reported.

Stimuli

The stimuli consisted of a subset of the stimuli for Experiment 9. Since there was no effect of familiarity of objects as visual cues for the inferential events, we excluded Inference-Unfamiliar events. Thus, the visual stimuli for Experiment 10 consisted of all of the Perception and Inference-Familiar (henceforth referred to as Inference) events of Experiment 9. There were 6 events per type of access, resulting in a total of 12 events.

Procedure

Each child was tested individually in a quiet room at his/her preschool. There were two experimenters: the first experimenter (E1) always interacted with the child and asked the questions, and the second experimenter (E2) only acted out the puppets. E1 presented the children with two cards that were face-down and said (in Turkish): "Look, I have two cards here. There is a picture under each card, but we can look at only one of them. The other has to be upside down." When E1 turned one of the cards up, E2 placed each of the puppets next to each card. E1 went on to say, "But these puppets can look at the picture under this card" (Figure 6a). Then one of the puppets looked at the visible photograph (the child could also see this card), and the other puppet looked under the face-down card (the child had no access to what was under the face-down card). After that, each puppet put the card they looked at into two different boxes (Figure 6b). The two boxes were next to the cards. E1 asked the children: "Who do we ask to find out more about <verb>?" (Figure 6c). As in Experiment 9, the verb was presented in the infinitive form, "to V". After the children made a choice the experimenter went onto the next trial.

As in Experiment 9, in half of the trials (3 per type of access, 6 in total), E1 asked about an event/verb that matched the visible photograph (e.g., the photographs showed some egg shells and E1 asked: "Who do we ask to find out about 'cracking'?"). In the other half of the trials E1 asked about an event/verb that did not match the visible photograph (e.g., the photographs showed some egg shells and E1 asked, "Who do we ask to find out about 'hitting'?"). The same counterbalancing procedure for verb types as in Experiment 9 was followed.

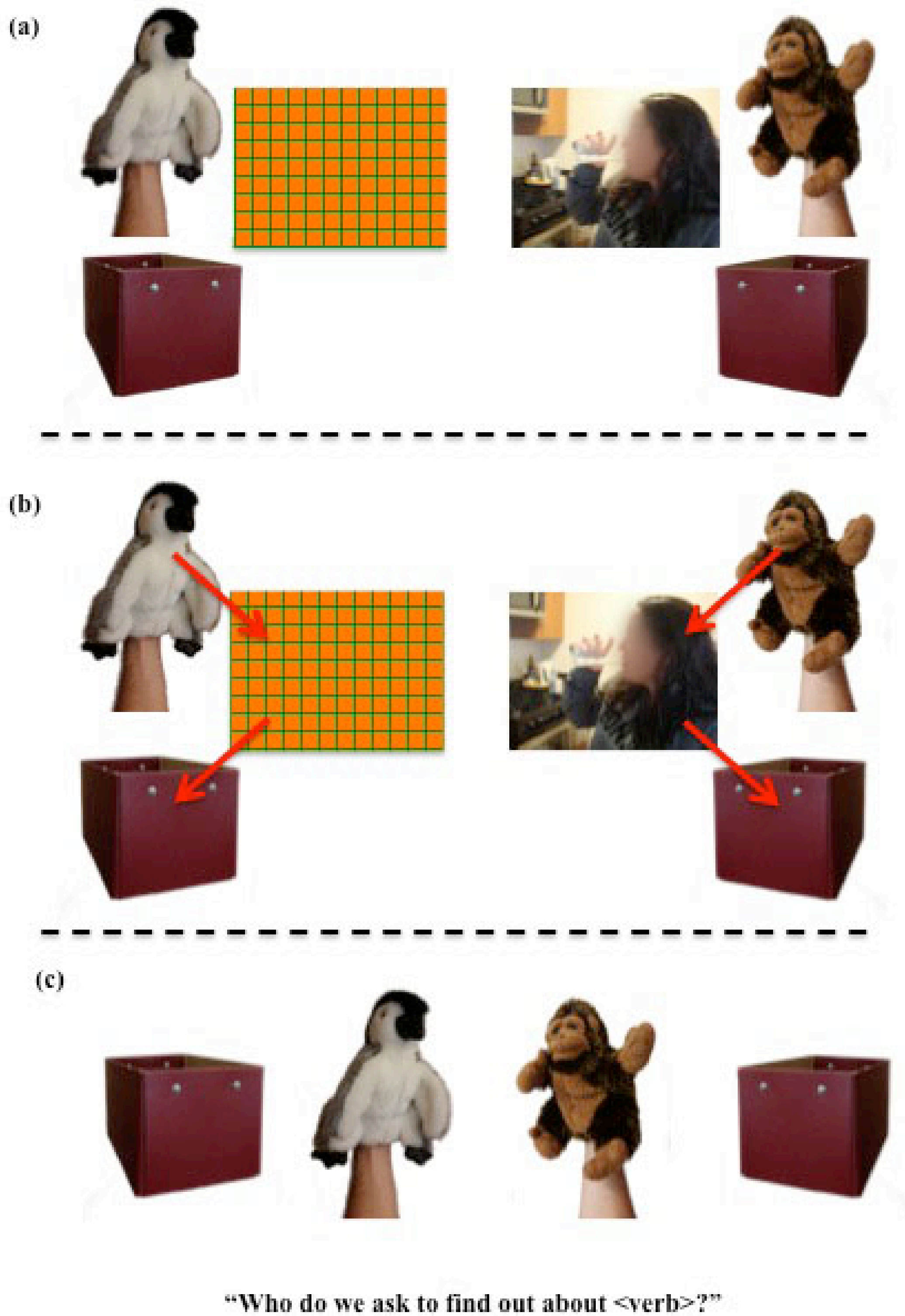

Figure 6 Summary of procedure of Experiment 10.

There were 3 practice trials at the beginning of the experiment. The visible photographs consisted of the same 3 photographs of objects as the practice trials of Experiment 9. E1 followed the same instructions as in the main experiment. After the child responded in each practice trial, E1turned up the face-down card and showed the photograph to give the child feedback about his/her response. Children made errors and received feedback in 22% of the filler trials.

Of interest was whether children would pick the puppet looking at the accessible event when they were asked about a verb that matched the accessible event and avoid picking the puppet looking at the accessible event and instead pick the puppet looking at the inaccessible alternative when the verb did not match the accessible event. Furthermore, we were interested in whether children's performance in attributing knowledge to others would differ from their performance in gaining knowledge for themselves in Experiment 9.

Results

As in Experiment 9, preliminary analyses showed that performance did not differ between Matching and Mismatching Verbs in either 4-year-olds (M = 0.52 vs. 0.67 respectively, $t(20) = -1.51$, *ns*) or 5- to 6-year-olds (M = 0.75 vs. 0.71 respectively, $t(15) = 0.50$, *ns*). We thus collapse across Verb Types in what follows.

Table 22 *Proportion of correct responses across types of access and age groups in Experiment 10*

	4-year-olds		5- to 6-year-olds	
	M	SE	M	SE
Perception	0.59	0.04	0.77	0.05
Inference	0.60	0.04	0.70	0.05

Table 22 presents the mean proportion of correct responses across Type of Access and Verb Type for each age group. We assessed whether children's ability to attribute knowledge to others differed across Types of Access with a 2 x 2 mixed ANOVA with Type of Access (Perception, Inference) as the within-subjects factor and Age (4-year-olds, 5- to 6-year-olds) as the between-subjects factor. The analysis returned only a main effect of Age ($F(1,35) = 8.94$, $p = .005$, $\eta^2 = .204$): 5- to 6-year-olds ($M = .73$) were more accurate than 4-year-olds ($M = .60$). There were no other significant effects or interactions. Both age groups performed at levels significantly different from chance for each Type of Access ($p < .05$).

Self vs. others in knowledge attribution

We compared the data from the Self Task (Experiment 9) to the data from the Others Task (Experiment 10) with a 2 x 2 x 2 mixed ANOVA with Type of Access (Perception, Inference) as the within-subject factor and Age (4-year-olds, 5- to 6-year-olds) and Perspective (Self, Other) as the between-subject factors. (We collapsed the data from Inference-Familiar and Inference-Unfamiliar trials in Experiment 9 to get a single accuracy score in Inference trials.) The analysis returned a main effect of Age ($F(1,65) = 15.16$, $p < .001$, $\eta^2 = .189$): 5- to 6-year-olds ($M = 0.82$) performed better than 4-year-olds ($M = 0.71$). There was also a main effect of Perspective ($F(1,65) = 55.22$, $p < .001$, $\eta^2 = .459$), such that performance was better in the Self task ($M = 0.87$) than in the Others task ($M = 0.67$). No other effects or interactions were significant. We conclude that attributing knowledge to others is more difficult than gaining knowledge for oneself.

Memory control

Notice that in the Others task of Experiment 10 children had to track which puppet accessed which event and the corresponding type of access and retain this information in memory until they heard the test question. One might hypothesize that the drop in performance might simply be attributed to higher memory demands of the Others task (as opposed to the need to take someone else's perspective). If that is the case, then children's performance should remain poor even if the perspective taking component is removed from the Others task and children are asked about their own knowledge as long as the memory demands remain the same. Alternatively, if the drop in performance in Experiment 10 stems from perspective taking demands, then performance should improve when children are asked about their own knowledge even when the memory demands of the task remain the same.

To investigate whether performance in Experiment 10 is affected by the higher memory demands of the task, we tested a new group of 4-year-olds (n = 6) using the paradigm of Experiment 10 but replaced the test question with "Which box has <verb>?", so that children would be asked about their own knowledge instead of the puppets' knowledge. Four-year-olds were highly accurate in this Memory Control task (Perception M = 0.75, Inference M = 0.81). We compared their performance to that of the 4-year-olds in Experiments 9 and 10 using a mixed ANOVA with Type of Access (Perception, Inference) as the within subjects factor and Task (Memory Control, Self, Others) as the between subjects factor. There was only a main effect of Task ($F(2,34) = 14.36$, $p < .001$, $\eta^2 = .458$): accuracy was lower in the Others Task (M = 0.62) than it was in both the Self Task (M = 0.83, $t((29) = 5.22$, $p < .001$) and the Memory Control Task (M = 0.78, $t(19) = 2.98$, $p = .017$), but there was no difference between the Self Task and the Memory Control Task ($t(20) = -0.98$, *ns*). These findings suggest that the drop in children's performance in Experiment 10 is not simply caused by the higher memory demands of the task but is best explained by the perspective-taking demands of attributing knowledge to others.

Language control

A possibility that remained open was that the relative difficulty of Experiment 10 compared to Experiment 9 might be linked to the Turkish form of the question used to probe others' knowledge states. Because Turkish has a flexible word order, the order in which information becomes available in an utterance may vary (Erguvanlı, 1984). As a result, in the Self task the target verb in the infinitive form appeared at the end of the sentence ("hangisinde <verb>" translated as: which one has <verb>). However, in the Others task the infinitive appeared at the beginning of the test question ("<verb> hakkında daha çok şey öğrenmek için kime sormalıyız" translated as: <verb> about more find out who do we ask). In order to ensure that the Self-Others asymmetry was not driven by such differences in the position of the critical information, we administered the same tasks in English, where the target verb appeared at the end of the test question in both the Self task ("which one has <verb>") and the Others task ("who do we ask to find out about <verb>").

In the English version, we replaced the infinitival form with the *–ing* form of the verb. Thirty-seven English-speaking 4-year-olds were recruited from daycares in Newark, Delaware. Each child was randomly age-matched with one of the original Turkish participants from Experiments 9 and 10, such that children from each language were compared for each of the Self task (n = 16, M_{age} = 4;6, range: 4;0-5;0) the Others task (n = 21, M_{age} = 4;5, range: 4;1-4;11). English speakers performed very similarly to Turkish speakers in both the Self task of Experiment 9 (Perception: M =

0.83, SE = .04; Inference-Familiar: *M* = 0.74, SE = .05; Inference-Unfamiliar: *M* = 0.78, SE = 0.04) and the Others task of Experiment 10 (Perception: *M* = 0.65, SE = .05; Inference: *M* = 0.63, SE = .04). Separate ANOVAs with Language and Type of Access as factors for each task revealed no effects or interactions involving Language. Furthermore, An ANOVA with Language, Task and Type of Access as factors revealed only a main effect of Task ($F(1,70) = 32.46$, $p < .001$, $\eta^2 = .317$). As expected, both language groups were more accurate in the Self task (*M* = 0.81) than in the Others task (*M* = 0.62). No other effects or interactions were significant. We conclude that the observed self-others asymmetry is not linked to surface features of the test probes in Turkish. In fact, children's ability to both gain knowledge about events and attribute such knowledge to others seems to proceed similarly across different language-learning communities.

Discussion

Experiment 10 examined children's ability to attribute knowledge to others who had either perceptual or inferential access to events. Our findings revealed a developmental difference, with 5- to 6-year-olds being better than 4-year-olds in attributing perception-based and inference-based knowledge about events to others. There was no difference between the two knowledge sources. A comparison between the findings of Experiments 9 and 10 showed an asymmetry between perspectives (i.e., self and others), such that success with gaining knowledge from perception and inference was higher than success with attributing perception-based and inference-based knowledge to others. These results are consistent with prior findings on children's understanding of perception and inference as knowledge sources (e.g., Sodian & Wimmer, 1987). We return to the significance of these findings in the General Discussion below.

General Discussion

The goal of the present investigation was to examine the development of children's understanding of knowledge sources using tasks that did not involve explicit mental state language, such as the verbs *think* and *know*. We compared attributing knowledge to oneself vs. others and focused on visual perception and inference from visual premises as ways of finding out about events.

Our findings revealed that 4- and 5- to 6-year-olds were highly successful in gaining knowledge about events from either direct/perceptual evidence or post-event inferential evidence (Experiment 9). Nevertheless, children in the same age groups made several errors when they had to attribute knowledge to others who had identical perceptual or inferential access to the very same events (Experiment 10). Furthermore, across both Self and the Others tasks, 4-year-olds made more errors compared to the older children. These findings cohere with and extend findings of prior research (Hogrefe et al., 1986; Sodian, 1988; Sodian & Wimmer, 1987, Wimmer et al., 1988, and Chapter 3 of this dissertation) in two ways. First, the self-others asymmetry extends to conditions in which children are asked to evaluate different ways of gaining knowledge about events, just like in prior studies in which children evaluated different ways of gaining knowledge about events. Second, the self-others asymmetry persists even when children are not asked to explicitly verbalize how they or someone else formed beliefs.

We argue that the asymmetry between representing knowledge for oneself and for others stems from the development of perspective-taking abilities needed to compute others' informational sources. In support for this idea, in the memory control experiment reported in Experiment 10, children responded differently when they were tested on their own knowledge sources (instead of others' sources) with the paradigm of Experiment 10. This suggests that children's difficulty in attributing knowledge to others cannot be simply attributed to a memory limitation.

Furthermore, the difficulty in attributing knowledge to others seems to characterize children from different communities: English-speaking 4-year-olds responded identically to Turkish-speaking 4-year-olds when tested with the current methods. Thus, it is unlikely that Turkish children's difficulty in attributing knowledge to others stems from the Turkish form of the question that assessed knowledge attribution to others. The fact that Turkish and English children's performance converged provides support for the idea that the ability to connect evidence to knowledge develops similarly in children acquiring languages with and without grammatizalized evidentiality systems.[5]

Finally, in both experiments there was a strong developmental trend such that children's success with knowledge attribution (both to oneself and others) increased with age. This developmental trend did not emerge in Experiment 8, possibly because of the narrower age range of 4- to 5-yearolds. In Experiment 9, success was very high in the oldest group of 5- to 6-year-olds who had 90% accuracy. However, in Experiment 10 even the oldest group of 5- to 6-year-old children did not perform at ceiling levels and had 75% accuracy. This suggests that the ability to reason about others' informational access does not catch up with the ability to access one's own information sources even at age 6 and continues to develop beyond the preschool years.

Chapter 5

GENERAL DISCUSSION

Languages vary in the way they encode several aspects of the world. Although it is widely recognized that there is a tight relationship between language and thought, the nature and extent of this relationship are widely debated. According to one view, language maps onto preexisting, possibly universal, conceptual categories, and reflects conceptual representations without modifying them. According to an alternative view, concepts are constructed through language, and acquiring language-specific distinctions alters conceptual representations. This dissertation tests these different views of the relationship between language and thought drawing on evidence from the domain of evidentiality and sources of knowledge.

As explained in the Introduction, there is considerable cross-linguistic variation in the encoding of information sources. Some languages such as Turkish mark evidential distinctions in their grammatical systems, and others such as English optionally encode evidentiality through lexical or syntactic devices. This variation raises the question whether speakers of languages with different evidential systems also differ in their source monitoring. From a developmental perspective, this variation raises the questions of how learners of a language such as Turkish acquire the evidential distinctions in their language, and how the acquisition of linguistic evidentiality relates to the development of source concepts. Thus, the study of the relation between evidentiality and non-linguistic sources of information provides unique insights about the relation between language and thought.

Summary of the Findings

The studies reported in Chapter 2 investigated whether speaking a language with obligatory evidentiality (such as Turkish) would make speakers less prone to source monitoring errors when distinguishing inference from perception in event source memory compared to speakers of a language (English) that lacks such encoding. These studies confirmed that Turkish and English speaking adults differed in how they mark source of information in language. Despite these cross-linguistic differences, however, Turkish and English speakers converged when monitoring the sources of their event memories. The two language groups made source monitoring errors to the same extent. Moreover, both groups were more likely to make source monitoring errors for post-event visual cues that yielded highly secure inferences as opposed to visual cues that yielded less secure inferences. Furthermore, their errors were of the same type: when speakers failed to report that they had inferred an event, both groups reported having seen the events. Together, these studies suggest that long-term experience with the evidential categories of one's native language does not shape conceptual representations of information sources.

Chapter 3 explored how young learners of Turkish acquire the evidential distinctions in their language. The studies reported in this Chapter revealed that Turkish learners between ages of 3 and 6 produced evidential morphemes accurately, but even 6-year-olds continued to have difficulty with evidential comprehension across multiple tasks with varying demands. This showed support for the idea that the difficulties with evidential comprehension may not be simply attributed to task demands. Importantly, children of similar ages also had difficulty reasoning about others' evidence even when the task did not involve knowledge of evidential language; but the difficulty disappeared when children were accessing their own information sources. The self-others asymmetry that persisted in linguistic tasks and their non-linguistic counterparts indicates that evidential comprehension is delayed by the development of perspective-taking abilities needed for considering other cognizers' access to information. These findings show a similarity between language and conceptual development and strongly suggest that the acquisition of linguistic categories is constrained by the development of corresponding conceptual representations.

Building on these findings, Chapter 4 further explored how Turkish children develop the ability to connect evidence to knowledge for themselves and others. These studies confirmed the asymmetry between accessing one's own vs. someone else's information sources in a new group of Turkish learners. Furthermore, these studies extended the findings reported in Chapter 3 in two ways. First, the Self and Others tasks in Chapter 3 tested only 4- and 5-year-olds. The Self and Others tasks in Chapter 4 included a wider age range of 4- to 6-year-olds —the typical age range in developmental investigations of source monitoring. The Self task revealed that the older group of children, including the 6-year-olds were highly successful and performed close to the ceiling level performance of adults. However, in the Others task, even the oldest group of 5- to 6-year-old children did not have ceiling-level performance. This suggests that children's understanding of how others gain knowledge from different kinds of informational sources is still developing. This has implications for future investigations of the acquisition of linguistic evidentiality (see Open Questions and Future Directions below). Second, the studies reported in Chapter 4 included several control manipulations aimed at eliminating alternative explanations of the Self-Others asymmetry. These control studies revealed that the asymmetry could not be merely attributed to the higher memory demands of the Others task, or the difference in the order in which the target verb appeared in the test questions of the Self and Others tasks (since English-speaking 4-year-olds performed similarly to their Turkish-speaking peers even though the test questions were somewhat different). Together these findings provide converging evidence for the prior conclusion that the acquisition of linguistic evidentiality depends – at least in part – on the development of source concepts.

Theoretical Implications

The present findings inform theories of the language-cognition interface by showing that the direction of the causal flow of the relationship between language and cognition is from cognition to language. This conclusion is supported by the findings of the cross-linguistic source monitoring studies in Chapter 2 (Experiments 2 and 3), which revealed commonalities in the way adult English and Turkish speakers remembered the sources of their event memories. Additional support comes from a developmental cross-linguistic comparison reported in Chapter 4, which showed that understanding the link between evidence and knowledge develops similarly across learners of English and Turkish. Although some prior work in cross-linguistic source monitoring has argued for an effect of linguistic evidentiality on source memory (Tosun et al., 2013), such effects only emerged in contexts where speakers were required to process linguistic material as part of a cognitive task. The present findings show that these cross-linguistic differences do not extend to contexts where adults (and children) are asked to perform a truly non-linguistic task. Together, the present findings strongly suggest that learned linguistic categories of evidentiality do not shape conceptual representations of information sources.

Viewed within the broader debate about the role of language on cognition, this conclusion coheres with prior work on further domains such as motion (Gennari et al., 2002; Papafragou et al., 2008; Trueswell & Papafragou, 2010), the object/substance distinction (Li, Dunham, & Carey, 2009), positional information in spatial scenes (Bosse & Papafragou, 2010), and spatial frames of reference (Li & Gleitman, 2002; Li, Abarbanell, Gleitman, & Papafragou, 2011). Across different domains, empirical findings point to some behavioral differences amongst speakers of different languages that can be attributed to cross-linguistic differences in encoding aspects of the world. However, these differences typically diminish or disappear when speakers are prevented from accessing verbal codes or are presented with other conceptual cues, a fact suggesting that language-driven differences are produced by temporary interactions between linguistic and visual and/or spatial representations. Furthermore, these investigations show that effects of language on cognitive processes are task-dependent and do not always surface in ordinary contexts. Thus, available findings in the domain of evidentiality and other domains support a universalist view of the language-cognition interface according to which linguistic categories exert their influence on how speakers cognize the world in a flexible way and possibly without modifying the underlying conceptual representations.

A further conclusion that emerges from the present findings is that there is a homology between linguistic and conceptual representations of sources of knowledge. Evidence in support for this idea comes from the findings of the studies reported in Chapter 3, which show that the asymmetry between the production and comprehension of evidentiality markers in Turkish is mirrored in non-linguistic counterparts of evidential production (i.e., accessing one’s own information sources) and comprehension (i.e., reasoning about someone else's information sources). This shows that there is a direct link between source representations in language and cognition and strongly suggests that the acquisition of linguistic evidentiality is constrained by the development of underlying source concepts.

This conclusion coheres with findings of prior work with young leaners of languages with evidential morphology showing a tight relation between conceptual representations of information sources and linguistic evidentiality (Aksu-Koç, 1988, Ozturk & Papafragou, 2015; Papafragou et al., 2007; Uzundag, Tasci, Küntay, & Aksu-Koç, 2016). Specifically, these studies have shown that the course of acquisition of different evidential morphemes follows the patterns in children's ability to identify the corresponding sources of information in non-linguistic tasks (Ozturk & Papafragou, 2015; Papafragou et al., 2007). The present findings extend prior findings by showing an asymmetry between accessing one's own vs. someone else's evidential sources that also persists in language and cognition.

The present findings have implications for theories of language acquisition by showing that part of the difficulty in the acquisition of linguistic categories can be attributed to the complexities in the underlying concepts. Converging evidence from other linguistic domains supports this conclusion: mental state verbs like '*think*' or '*know*' are acquired much later than verbs referring to actions like '*jump*' or '*run*' partly due to the challenges posed by the abstractness of the concepts expressed by these mental state terms (Bartsch & Wellman, 1995; Gopnik & Meltzoff, 1997; Perner, 1991; Smiley & Huttenlocher, 1995).

One might argue that the similarities between linguistic evidentiality and non-linguistic source concepts can be taken as evidence for an alternative view of language acquisition, according to which concepts are constructed through language. However, the fact that English-speaking and Turkish-speaking 4-year-olds performed identically when tested with the very same tasks suggest that this is unlikely. Furthermore, there is evidence that acquiring evidential morphology lags behind the ability to reason about information sources in several respects (Ozturk & Papafragou, 2015; Papafragou et al., 2007). This asymmetry offers evidence against the possibility that evidential distinctions in language serve as pacesetters for cognitive development. Together, findings from these studies confirm the tight mapping between language and concepts, and support the idea that source monitoring proceeds similarly across speakers of different languages and that language builds on (rather than shapes) the ability to reason about different types of informational access.

Open Questions and Future Directions

Several directions for future research remain open at this point. As the present findings show, both the full acquisition of evidential systems and the development of source monitoring follow a lengthy timetable. In order to fully understand the factors behind such protracted development, future work should focus on a wider age range including older children, given that the majority of the available studies focus on younger children between the ages of 3 and 7. Studies with older children can also reveal whether factors such as the indirectness of evidence affect the use of indirect evidential morpheme in learners of Turkish or other languages with grammaticalized evidentiality.

The present research has focused on the contrast between grammatical vs. lexical encoding of information source. However, as mentioned in the Introduction, there is considerable variation even within the class of languages that encode evidentiality in their grammar, with some languages having detailed distinctions between types of direct and indirect access (see Aikhenvald, 2004, 2014). A possible direction for future research would be to compare speakers of languages with richer evidential systems to speakers of languages with two-way distinctions such as Turkish to test whether such systems differentially impact source monitoring processes. Although this possibility is tempting, it is currently hard to evaluate because little is known about how such detailed systems are actually used in ordinary contexts to encode informational access. Even though richer systems with four- or five-way evidential distinctions exist, they are in fact quite rare (Aikhenvald, 2004). Furthermore, many grammatical systems of evidentiality seem to be subject to some regularities and respect the broad semantic distinction between direct and indirect information sources (Faller, 2001; Willet, 1988) in ways that seem to affect the learnability of evidential systems (Bartell & Papafragou, 2015).

The present findings offer support for the idea that the development of source monitoring follows a similar timetable across distinct language learning groups. However, in the past some studies have argued for a source-monitoring advantage for Turkish learners over English learners (Aksu-Koç et al., 2009; Lucas et al., 2013). These studies lacked important controls and suffered from several interpretative issues. The present research offers a new methodological approach for future studies aiming to establish such cross-linguistic differences (or the lack thereof). As a first step, future studies should empirically establish cross-linguistic differences using carefully matched stimuli. Second, researchers should seek to establish cognitive equivalence among language groups using independent measures. Finally, the non-linguistic tasks should be identical across language groups and closely correspond to the linguistic domain under investigation.

ENDNOTES

[1] 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).

[2] 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*).

[3] We did not include the full set for brevity since children now had to watch both versions of each event (see Procedure below). To ensure that the events included in Experiment 5 elicited the intended evidentiality markers, the proportion of descriptions marked with direct past tense (-*dı*) out of total past tense descriptions in Experiment 4 was recalculated based on this subset of events. Mean use of – *dı* was 94% for Seen events, and 19% for Inferred events (compared to 94% and 16%, respectively based on the full set of 8 events). This suggests that sampling from the original set of events did not substantially affect the comparison to the production results in Experiment 4.

[4] A potential concern about Experiment 7 is that children resisted picking the "mystery" event because there was uncertainty or confusion about that option. Because the design and results of Experiment 8 address this concern, we postpone its discussion until the next section.

[5] This interpretation appears to conflict with the findings of a recent cross-linguistic investigation comparing English, Turkish and Chinese speakers between the ages of 3 and 4 on false belief and flexible trust tasks (Lucas Lewis, Pala, Wong, & Berridge, 2013). The flexible trust task, that was most relevant for the present purposes, required tracking two speakers' accuracy in labeling objects in order to be able select which speaker to trust when learning a label for a novel object. The results revealed that only older Turkish children performed above chance level in the false belief task and Turkish children outperformed the other groups in the flexible trust task. Although, the authors argued that Turkish children's superior performance in false belief tasks and flexible trust can be attributed to learning a language with grammaticalized evidentiality, this hypothesis could not be tested directly as Lucas et al. did not include a measure of Turkish children's knowledge of evidential language. Furthermore, Lucas et al. did not directly test whether the relation between language and flexible trust was mediated by false belief performance, so the mechanism that might transmit language effects on flexible trust remains unknown. This is especially important given that the mapping between the meaning conveyed by evidentiality markers (direct, hearsay or inference) and the information sources in the task (Speaker A vs. Speaker B) was not straightforward as in the studies of Aksu- Koç and colleagues (2009; see Chapter 2 for a detailed discussion).

REFERENCES

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford, United Kingdom: Oxford University Press.

Aikhenvald, A., & Dixon, R., Eds. (2001). *Studies in evidentiality*. Amsterdam, Netherlands: John Benjamins.

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.

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, Ögel-Balaban, H., & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development*, *125*, 13–28. doi: 10.1002/cd.247

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., Küntay, A.C., Acarlar, F., Mavis¸ İ., Sofu, H., Topbaş, S. & Turan, F. (2011). *Türkçe'de erken sözcük ve dilbilgisi gelişimini ölçme ve değerlendirme çalışması. Türkçe İletşim Gelişimi Envanterleri. Tige-I and Tige-II (Measurement and evaluation of early vocabulary and grammar development in early Turkish. Turkish Communicative Development Inventories. Tige-I and Tige-II)* (Research Report No. 107K058). Ankara, Turkey. Retrieved from http://uvt.ulakbim.gov.tr/proje/

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.

Anderson, R. E. (1984). Did I do it or did I only imagine doing it? *Journal of Experimental Psychology: General, 113*, 594–613. doi: 10.1037/0096-3445.113.4.594

Baayen, R. H. (2008). *Analyzing linguistic data a practical introduction to statistics using R.* New York, NY: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition, 188*, 87–96. doi:10.1016/j.cognition.2010.10.010

Barnes, J. (1984). Evidentials in the Tuyuca Verb. *International Journal of American Linguistics*, *50*, 255–271. Retrieved from http://www.jstor.org/stable/1265549

Barr, D. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59,* 457-474. doi: 10.1016/j.jml.2007.09.002

Bartsch, R., & Wellman, H. (1995). *Children talk about the mind*. Oxford, United Kingdom: Oxford University Press.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, *5*, 27–30. Retrieved from http://cran. rproj ect. org/doc/Rnews/RnewsfZO05-1.pdf

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes,* R package version 0.99875-6.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.99375-39. Retrieved from http://CRAN.R-project.org/package-lme4.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67,* 1-48. doi:10.18637/jss.v067.i01.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar:*

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*(1), 40-54. doi: 10.3758/BF03202760

Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua*, *92*, 297-329. doi:10.1016/0024-3841(94)90345-X

Boroditsky, L. (2003). Linguistic Relativity. In Nadel, L. (Ed.) *Encyclopedia of cognitive science* (pp. 917-921). London, United Kingdom: MacMillan Press.

Bosse, S., & Papafragou, A. (2010). Spatial position in language and visual memory: A cross-linguistic comparison. *Proceedings from the 32nd Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

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.

Bowerman, M., & Levison, S. (2001). *Language acquisition and conceptual development*. Cambridge, United Kingdom: Cambridge University Press.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*(4), 331–350. doi: 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*(6), 673–678. doi:10.3758/BF03197414

Brosseau-Liard, P.E., & Birch, S. A. J. (2011). Epistemic states and traits: Preschoolers appreciate the differential informativeness of situation-specific and person-specific cues to knowledge. *Child Development*, 82, 1788-1796. doi: 10.1111/j.1467-8624.2011.01662.x

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*(5), 633–639. doi:10.1002/acp.1215

Chierchia, G., Crain, S. Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In A. H.J. Do, L. Domínguez, & A. Johansen (Eds.), *Proceedings of the 25th annual Boston University Conference on Language Development,* (pp.157-168), Somerville, MA: Cascadilla Press.

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, Netherlands: John Benjamins.

Chomsky, N. (1975). *Reflections on language*. New York, NY: Pantheon.

Clark, E., (1993). *The lexicon in acquisition*. Cambridge, United Kingdom: Cambridge University Press.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dallett, K., & Wilcox, S.G. (1968). Remembering pictures vs. remembering descriptions. *Psychological Science, 11*(4), 139-140. doi: 10.3758/BF03331012

De Haan, F. (2001). The place of inference within the evidential system. *International Journal of American Linguistics*, 67, 193-219. Retrieved from http://www.jstor.org/stable/1265889

DeLancey, S. (2001). The mirative and evidentiality. *Journal of Pragmatics, 33*, 369–382. doi: 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. doi: 10.1177/0956797612457374

Dromi, E. (1987). *Early lexical development*. Cambridge, United Kingdom: Cambridge University Press.

Drummey, A, B. & Newcombe, N. S. (2002). Developmental changes in source memory. *Developmental Science*, *5*, 502–513. doi: 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. doi: 10.1016/S0022-5371(80)90294-7

Erguvanlı, E. E. (1984). *The function of word order in Turkish grammar*. Berkeley, CA: University of California Press.

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(6), 801–803. doi:10.1177/0956797610371341

Fitneva, S. (2009). Evidentiality and trust: the effect of informational goals. *New Directions for Child and Adolescent Development*, *125*, 49–62. doi:10.1002/cd.249

Fitneva, S. A., Lam, N. H. L., & Dunfield, K. A. (2013). The Development of Children's Information Gathering: To Look or to Ask? *Developmental Psychology, 49*, 533-542. doi: 10.1037/a0031326

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, *64*, 74–92. doi: 10.1016/j.cogpsych.2011.10.004

Fraundorf, S.H., Benjamin, A.S., &Watson, D.G. (2013). What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language. 69*, 196-227. doi: 10.1016/j.jml.2013.06.003

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition*, *83*, 49–79. doi: 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. doi: 10.1075/sl.6.1.03giv

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55. doi: 10.1207/s15327817la0101_2

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak and 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 and R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (2nd edition) (pp.543-568). New York, NY: Oxford University Press.

Goldin-Meadow, S., Seligman, M. E.P., & Gelman, R. (1976). Language in the two-year-old. *Cognition, 4*, 189–202. doi:10.1016/0010-0277(76)90004-4

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. doi: 10.1111/1467-8624.ep8589535.

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts and theories*. Cambridge, MA: MIT Press.

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(4), 931–940. doi: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*(6), 626–637. doi: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), 1–22. doi: 10.1037/0096-3445.107.1.1

Hart, W., & Albarricin, D. (2011). Learning about what others were doing: Verb aspect and attributions of mundane and criminal intent for past actions. *Psychological Science*, *22*(2), 261–266. doi: 10.1177/0956797610395393

Hayes-Roth, B., & Thorndyke, P. W. (1979). Integration of knowledge from texts. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 91-108. doi: 10.1016/S0022-5371(79)90594-2

Hendriks, P. (2013). *Asymmetries between Language Production and Comprehension*. London, United Kingdom: Springer. doi:10.1007/978-94-007-6901-4

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*, 1887–1897. doi:10.1016/j.lingua.2010.02.002

Hirst, W., & Weil, J. (1982). Acquisition of the epistemic and deontic meaning of modals. *Journal of Child Language, 9*, 659–666. doi: http://dx.doi.org/10.1017/S0305000900004967

Hogrefe, G.J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in the attribution of epistemic states. *Child Development, 57*, 567-582. doi: 10.2307/1130337

Huang, Y., Spelke, E., & Snedeker, J. (2013). What exactly do number words mean? *Language Learning and Development, 9*, 105-129. doi:10.1080/15475441.2012.658731.

Hurewitz, F., Brown-Bchmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*, 597–627. doi: 10.1023/A:1026468209238

Huttenlocher, J., Smiley, P., & Charney, R. (1983). The emergence of action categories in the child: evidence from verb meanings. *Psychological Review*, *90*, 72–93. doi: 10.1037/0033-295X.90.1.72

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition, 62*, 169-200. doi: 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. doi: 10.2307/1422983

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446. doi:10.1016/j.jml.2007.11.007

Johanson, L. (2003). Evidentiality in Turkic. In: A. Y. Aikhenvald, & R. M. W. Dixon, (Eds.), *Studies in Evidentiality (Typological Studies in Language 54),* (pp.273-290). Amsterdam, Netherlands & Philadelphia, PA: John Benjamins.

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. (1997). Source monitoring and memory distortion. *Philosophical Transactions: Biological Sciences*, *352*, 1733–1745. doi: 10.1098/rstb.1997.0156

Johnson, M. K. (2006). Memory and Reality. *American Psychologist*, *61*, 760-771. doi: 10.1037/0003-066X.61.8.760

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28. doi: 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. doi: 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. doi: 10.1037/0278-7393.20.6.1409

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85. doi: 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. doi: 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. doi: 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*(1), 115-122. doi: 10.3758/BF03209202

Johnson, M.K., Bransford, J.D., & Solomon, S.K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology, 98*, 203-205

Johnston, J. R., & Slobin, D. I. (1979). The development of locative expressions in English, Italian, Serbo-Croation and Turkish. *Child Language*, *6*, 529-545. doi: 10.1017/S030500090000252X

Keenan, T., Ruffman, T., & Olson, D. (1994). When do children begin to understand logical inference as a source of knowledge? *Cognitive Development*, *9*, 331-353. doi:10.1016/0885-2014(94)90010-8

Kornfilt, J. (1997). *Turkish.* London, United Kingdom: Routledge.

Landau, B., Dessalegn, B., & Goldberg, A. M. (2010) Language and space: Momentary interactions. In P. Chilton and 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.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

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., Abarbanell, L., Gleitman, L., & Papafragou, A. (2011). Spatial reasoning in Tenejapan Mayans. *Cognition, 120*, 33-53. doi: 10.1016/j.cognition.2011.02.012

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology, 58*, 487–524. doi: 10.1016/j.cogpsych.2008.12.001

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83*, 265-294. doi: 10.1016/S0010-0277(02)00009-4

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1077-1083. doi: 10.1037/0278-7393.16.6.1077

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17*, 349–358. doi: 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. doi: 10.1016/0022-0965(91)90065-Z

Lucas, A. J., Lewis, C., Pala, F. C., Wong, K., & Berridge, D. (2013). Social-Cognitive Processes in Preschoolers' Selective Trust: Three Cultures Compared. *Developmental Psychology, 49*, 579-590. doi: 10.1037/a0029864

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*, 355–62. doi:10.1111/j.1467-7687.2009.00889.x

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An Update, *Journal of Child Language*, *17*, 457-472. doi: http://dx.doi.org/10.1017/S0305000900013866

MacWhinney, B., & Snow, C. E. (1985). The child language data exchange system. *Journal of Child Language*, *12*, 271–295. doi:http://dx.doi.org/10.1017/S0305000900006449

Mather, M., Johnson, M. K., & De Leonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, *16,* 437–458. doi: 10.1080/026432999380870

Matsui, T. (2014). Children's understanding of linguistic expressions of certainty and evidentiality. In: Matthews, D. (Ed.), *Pragmatic development in first language acquisition*. (pp. 295–316). Amsterdam, The Netherlands: John Benjamins. doi: 10.1075/tilar.10.17mat

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. doi: 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*, 1273–84. doi: 10.3758/BF03193271

Miller, S. A., Hardin, C. A., & Montgomery, D. E. (2003). Young children's understanding of the conditions for knowledge acquisition. *Journal of Cognition and Development, 4*, 325-356. doi: 10.1207/S15327647JCD0403_05

Munnich, E., Landau, B., & Dosher, B. A. (2001). Spatial language and spatial representation: a cross-linguistic comparison. *Cognition*, *81*, 171-207. doi: 10.1016/S0010-0277(01)00127-5

O'Neill, D., & Gopnik, A. (1991). Young children's ability to identify the sources of their beliefs. *Developmental Psychology, 27(3)*, 390-397. doi: 10.1037/0012-1649.27.3.390

Ozturk, O, & Papafragou, A. (2014). The acquisition of epistemic modality: from semantic meaning to pragmatic interpretation. *Language Learning and Development*, 00, 1-24. doi:10.1080/15475441.2014.905169

Ozturk, O., & Papafragou, A. (2007). Children's acquisition of evidentiality. *Proceedings from the 31st Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Ozturk, O., & Papafragou, A. (2015). The acquisition of evidentiality and source monitoring. *Language Learning and Development.* doi**:** 10.1080/15475441.2015.1024834

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition, 105*, 125–165. doi:10.1016/j.cognition.2006.09.008

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*(1), 155**–**184. doi: 10.1016/j.cognition.2008.02.007

Papafragou, A., Li, P., Choi, Y., & Han, C.-H. (2007). Evidentiality in language and cognition. *Cognition, 103,* 253–99. doi: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. doi: 10.1016/S0010-0277(02)00046-X

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Pillow, B. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*, 116-129. doi:10.1016/0022-0965(89)90066

Pillow, B. (1999). Children's understanding of inferential knowledge. *Journal of Genetic Psychology, 160*, 419–428.

Pillow, B. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development*, 73, 779-792. doi:10.1111/1467-8624.00438

Pillow, B. H., & Anderson, K. L. (2006). Children's awareness of their own certainty and understanding of deduction and guessing. *British Journal of Developmental Psychology, 24*, 823–849. doi:10.1348/026151005X83360

Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception, and guessing. *Developmental Psychology, 36,* 169-179. doi:10.1037//0012-1649.36.2.169

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York, NY: Morrow.

Povinelli, D., & de Blois, S. (1992). Young children's (*Homo sapiens*) understanding of knowledge formation in themselves and others. *Journal of Comparative Psychology, 106*, 228-238. doi:10.1037/0735-7036.106.3.228

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, 61, 973-982. doi: 10.1111/j.1467-8624.1990.tb02835.x

R-Core-Team. (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Rett, J. & Hyams, N. (2014). The acquisition of syntactically encoded evidentiality. *Language Acquisiton, 21*, 173-197. doi:10.1080/10489223.2014.884572

Robinson, E. J. (2009). Commentary: What we can learn from research on evidentials. *New Directions for Child and Adolescent Development, 125,* 95–103. doi: 10.1002/cd.252

Robinson, E. J., & Haigh, S. N. & Nurmsoo, E. (2008). Children's working understanding of knowledge sources: Confidence in knowledge gained from testimony. *Cognitive Development, 23,* 105-118. doi: 10.1016/j.cogdev.2007.05.001

Robinson, E. J., & Whitcombe, E. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development*, *74(1)*, 48-62. doi: 10.1111/1467-8624.t01-1-00520

Ruffman, T. & Olson, D. R. (1989), Children's Ascriptions of Knowledge to Others. *Developmental Psychology, 25,* 601-606. doi: 10.1037/0012-1649.25.4.601

Saxe, R. (2013). The new puzzle of Theory of Mind development. In M. Banaji & S. Gelman (Eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 107-112). New York, NY. Oxford University Press.

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, Netherlands: John Benjamins.

Smiley, P., & Huttenlocher, J. E. (1995). Conceptual development and the child's early words for events, objects, and persons. In M. Tomasello & W. Merriman (Eds.), *Beyond names for thing: Young children's acquisition of verbs* (pp. 21–62). Hillsdale, NJ: Erlbaum.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In G. Hall & S. Waxman (Eds.), *Weaving a lexicon* (pp. 257–293). Cambridge, MA: MIT Press.

Sodian, B. (1988). Children's attributions of knowledge to the listener in a referential communication task. *Child Development, 59*, 378-385. doi: 10.2307/1130317

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development, 58*, 424-433. doi:10.2307/1130519

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development,* 62, 753-766. doi: 10.1111/j.1467-8624.1991.tb01567.x

Spelke, E., & Tsivkin, S. (2001). Initial knowledge and conceptual change: space and number. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 70–100). Cambridge, United Kingdom: Cambridge University Press.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33*, 203–217. doi: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. doi: 10.1016/j.cognition.2011.04.007

Taylor, M. (1988). The development of children's understanding of the seeing-knowing distinction. In J. Astington, P. Harris & D. Olson (Eds.), *Developing theories of mind* (pp. 207-225). New York, NY: Cambridge University Press.

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. doi: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. doi: 10.1016/j.jml.2010.02.006

Uzundag, B, Tasci, S, Küntay, A. C., Aksu-Koç, A. (2016). Functions of evidentials in Turkish child and child-directed speech in early child-caregiver interactions. In *Proceedings of the 40th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press

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: Cascadilla Press.

Ünal, E., & Papafragou, A. (in press). Interactions between language and visuo-spatial representations. *Language Learning.*

Varouxaki, A. Freeman, N.H., Peters, D., & Lewis, C. (1999). Inference neglect and ignorance denial. *British Journal of Developmental Psychology*, *17*, 483–499. doi: 10.1348/026151099165401

de Villiers, J. G., Cahillane, J., & Altreuter, E. (2006). What can production reveal about Principle B? In: K.U. Deen, J. Nomura, B. Schulz, & B.D. Schwartz (Eds), *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America*, (pp. 89-100). Honolulu, HI.

de Villiers, J. G., Garfield, J., Gernet-Girard, H., Roeper, T., & Speas, M. (2009). Evidentials in Tibetan: acquisition, semantics, and cognitive development. *New Directions for Child and Adolescent Development*, *125*, 29–47. doi: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. doi:10.1075/sl.12.1.04wil

Wimmer, H., Hogrefe, G. J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development*, *59,* 386-396. doi:10.1111/1467-8624.ep8589275.

Wimmer, H., Hogrefe, G. J., & Sodian, B. (1988). A second stage in children's conception of mental life: Understanding informational access as origins of knowledge and belief. In J. Astington, P. Harris & D. Olson (eds.), *Developing theories of mind* (pp. 173-192). Cambridge: Cambridge University Press.

Winans, L., Hyams, N., Rett, J., & Kalin, L. (2014). *Children's comprehension of syntactically encoded evidentiality*. Paper presented at the 45th annual meeting of the North East Linguistic Society, Cambridge, MA. Retrieved from http://linguistics.ucla.edu/people/grads/lwinans/papers/NELS-paper.pdf

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. doi: 10.1073/pnas.0701644104

Wolff, P. M., & Holmes, K. J. (2011). Linguistic relativity. *WIREs Cognitive Science, 2*, 253–265. doi: 10.1002/wcs.104

Woolley, J., & Bruell, M. (1996). Young children's awareness of the origins of their mental representations. *Developmental Psychology*, *32(2)*, 335-346. doi: 10.1037/0012-1649.32.2.335

Zaragoza, M. S., & Koshmider, J. W. III. (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 246–255. doi: 10.1037/0278-7393.15.2.246

Appendix A

LIST OF TARGET EVENTS (EXPERIMENTS 1-3)

	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

Appendix B

LIST OF EVENTS/OBJECTS USED IN EXPERIMENT 4

Trial Type	Event/Object
Seen	1 Closing the cover of a box
Seen	2 Dropping three small objects into a jar
Seen	3 Stacking blocks
Seen	4 Pushing a car
Inferred	5 Coloring a star
Inferred	6 Cutting a piece of cardboard
Inferred	7 Inflating a balloon
Inferred	8 Tearing a shirt
Filler	9 Giraffe
Filler	10 Bird house
Filler	11 Panda
Filler	12 Octopus
Filler	13 Turtle
Filler	14 Cat
Filler	15 Horse
Filler	16 Butterfly

Note. Events (1-3) and (5-7) were also used in Experiments 5-6. For these later studies, we created both Seen and Inferred versions of each event. Slightly modified incarnations of these new Seen and Inferred versions were also used in Experiments 7 and 8 alongside additional stimuli (see Appendix C).

Appendix C

LIST OF EVENTS AND VERBS USED IN EXPERIMENTS 7 AND 8

Matching Trial Events		Verbs (Experiment 7)	Verbs (Experiment 8)
1	Closing the cover of a box	Closing	Closing
2	Dropping three small objects into a jar	Dropping	Dropping
3	Stacking blocks	Stacking	Stacking
4	Coloring a star	Coloring	Coloring
5	Cutting a piece of cardboard	Cutting	Cutting
6	Inflating a balloon	Inflating	Inflating
Mismatching Trial Events			
7	Biting a cookie	Biting	Washing
8	Dressing a toy frog	Dressing	Sleeping
9	Wearing a glove	Wearing	Jumping
10	Sticking band aid on hand	Sticking	Walking
11	Pushing a car into a box	Pushing	Combing
12	Opening the cover of a book	Opening	Tying

Note. In Experiment 7, all verbs matched the events (but half of the evidential morphemes on the verbs did.) In Experiment 8, only half of the verbs matched the events.

Appendix D

UNIVERSITY OF DELAWARE IRB APPROVAL FOR RESEARCH ON HUMAN SUBJECTS

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	June 27, 2011
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-2] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	June 27, 2011
EXPIRATION DATE:	July 24, 2012
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # 7

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

University of Delaware

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: June 5, 2012

TO: Anna Papafragou, PhD
FROM: University of Delaware IRB

STUDY TITLE: [180476-4] Evidential Categories in Language and Cognition

SUBMISSION TYPE: Amendment/Modification

ACTION: APPROVED
APPROVAL DATE: June 5, 2012
EXPIRATION DATE: July 24, 2012
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 7

Thank you for your submission of Amendment/Modification materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that <u>informed consent</u> is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: June 25, 2013

TO: Anna Papafragou, PhD
FROM: University of Delaware IRB

STUDY TITLE: [180476-7] Evidential Categories in Language and Cognition

SUBMISSION TYPE: Continuing Review/Progress Report

ACTION: APPROVED
APPROVAL DATE: June 25, 2013
EXPIRATION DATE: July 24, 2014
REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # 7

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

UNIVERSITY OF DELAWARE

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	July 14, 2014
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-10] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	July 14, 2014
EXPIRATION DATE:	July 24, 2015
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

UNIVERSITY OF DELAWARE
RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE:	July 17, 2015
TO:	Anna Papafragou, PhD
FROM:	University of Delaware IRB
STUDY TITLE:	[180476-13] Evidential Categories in Language and Cognition
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	July 17, 2015
EXPIRATION DATE:	July 24, 2016
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # *(7)*

Thank you for your submission of Continuing Review/Progress Report materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

Appendix E

REPRINT PERMISSION FOR CHAPTER 2

ELSEVIER LICENSE TERMS AND CONDITIONS

Mar 28, 2016

This is an Agreement between Ercenur Unal ("You") and Elsevier ("Elsevier"). It consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Ercenur Unal
Customer address	105 The Green Room 108 NEWARK, DE 19711
License number	3837790021790
License date	Mar 28, 2016
Licensed content publisher	Elsevier
Licensed content publication	Journal of Memory and Language
Licensed content title	Monitoring sources of event memories: A cross-linguistic investigation
Licensed content author	Ercenur Ünal, Adrienne Pinto ,Ann Bunger, Anna Papafragou
Licensed content date	April 2016
Licensed content volume number	87
Licensed content issue number	n/a

Number of pages	20
Start Page	157
End Page	176
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	full article
Format	electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	The Language-Cognition Interface: Evidence from the Domain of Evidentiality
Expected completion date	Apr 2016
Estimated size (number of pages)	184
Elsevier VAT number	GB 494 6272 12
Price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	**0.00 USD**

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable

acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The
Lancet, Vol. number, Author(s), Title of article, Pages No.,
Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license atthe end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All

content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies.

Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor author communications.

Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript

via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group

directly by providing copies to their students or to research collaborators for their personal use

for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement

- after the embargo period

  via non-commercial hosting platforms such as their institutional repository via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.
The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee. Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.8

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Appendix F

REPRINT PERMISSION FOR CHAPTER 3

ELSEVIER LICENSE TERMS AND CONDITIONS

Mar 29, 2016

This is an Agreement between Ercenur Unal ("You") and Elsevier ("Elsevier"). It consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Ercenur Unal
Customer address	105 The Green Room 108 NEWARK, DE 19711
License number	3838100816035
License date	Mar 28, 2016
Licensed content publisher	Elsevier
Licensed content publication	Journal of Memory and Language
Licensed content title	Production–comprehension asymmetries and the acquisition of evidential morphology
Licensed content author	Ercenur Ünal, Anna Papafragou
Licensed content date	Available online 4 January 2016
Licensed content volume number	n/a
Licensed content issue number	n/a
Number of pages	1

Start Page	None
End Page	None
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	The Language-Cognition Interface: Evidence from the Domain of Evidentiality
Expected completion date	Apr 2016
Estimated size (number of pages)	
Elsevier VAT number	GB 494 6272 12
Price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	**0.00 USD**

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR

APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The
Lancet, Vol. number, Author(s), Title of article, Pages No.,
Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage

for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies.

Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor author communications.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use

for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement

- after the embargo period

  via non-commercial hosting platforms such as their institutional repository via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only.

You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee. Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.8

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.