Journal of Memory and Language 89 (2016) 179–199

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Production–comprehension asymmetries and the acquisition of evidential morphology

Ercenur Ünal *, Anna Papafragou

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

ARTICLE INFO

Article history:
Received 6 November 2014
revision received 25 November 2015
Available online 4 January 2016

Keywords:
Language acquisition asymmetries
Evidentiality
Turkish
Source monitoring
Theory of mind

ABSTRACT

Although children typically comprehend the links between specific forms and their meanings before they produce the forms themselves, the opposite pattern also occurs. The nature of these 'reverse asymmetries' between production and comprehension remains debated. Here we focus on a striking case where production precedes comprehension in the acquisition of Turkish evidential morphology and explore theoretical explanations of this asymmetry. We show that 3- to 6-year-old Turkish learners produce evidential morphemes accurately (Experiment 1) but have difficulty with evidential comprehension (Experiment 2). Furthermore, comprehension failures persist across multiple tasks (Experiments 3–4). We suggest that evidential comprehension is delayed by the development of mental perspective-taking abilities needed to compute others' knowledge sources. In support for this hypothesis, we find that children have difficulty reasoning about others' evidence in non-linguistic tasks but the difficulty disappears when the tasks involve accessing one's own evidential sources (Experiment 5).

Introduction

When learning their native language, children typically understand the mappings between forms and their meanings before they can produce these forms in speech. For instance, babies produce their first words around their first birthday but seem to have some understanding of spoken language a few months earlier; moreover, this asymmetry between comprehension and production is quite robust across different structures (Bates, Bretherton, & Snyder, 1988; Goldin-Meadow, Seligman, & Gelman, 1976). This makes sense since, logically, comprehension must precede production: one must know the meaning carried by a particular form and represent the form-meaning mappings in order to be able to use that form to convey an intended meaning through speech (Clark, 1993).

However, cases where the production of a form is more advanced that its comprehension are not uncommon (see Hendriks, 2013; Hendriks & Koster, 2010 for an overview). For instance, English-speaking children produce the third-person singular /*s*/ on novel verbs by age 2 or 3 (Theakston, Lieven, & Tomasello, 2003), but can reliably use it as a cue to subject number only after age 5 (Johnson, de Villiers, & Seymour, 2005). Similarly, young English learners do not have difficulty producing third person pronouns such as *him* (Bloom, Barss, Nicol, & Conway, 1994) and do not confuse pronouns and reflexives (e.g., *himself*) in production (de Villiers, Cahillane, & Altreuter, 2006); however, when presented with sentences such as "Ernie washed him" in the presence of two male referents, Ernie and Bert, children younger than 6 typically interpret the pronoun *him* as the reflexive *himself* (these difficulties are selective, since they do not extend to sentences such as "Ernie washed himself"; Chien & Wexler, 1990). To take another example, children do not seem to infelicitously produce the

* Corresponding author. Fax: +1 (302) 831 3645.
E-mail address: eunal@psych.udel.edu (E. Ünal).

http://dx.doi.org/10.1016/j.jml.2015.12.001
0749-596X/

quantifier *some* in cases where *all* would have been true and appropriate (Papafragou & Musolino, 2003), but fail to reject infelicitous sentences with *some* such as "Some elephants have trunks" in contexts where *all* would have been acceptable (Noveck, 2001, cf. Guasti et al., 2005). In a further demonstration, children between ages 4 and 5 consistently produced spatial modifiers (e.g., "the one on the napkin") when describing a particular agent's action in the presence of two potential referents (e.g., a frog sitting on a napkin and another frog sitting on a plate). Nevertheless, when presented with sentences such as "Put the frog on the napkin into the box" in the presence of two frogs, the same children consistently interpreted the prepositional phrase "on the napkin" as the destination of the action rather than the modifier of "the frog", i.e., they cannot use it as a cue to disambiguate between the referents in comprehension (Hurewitz, Brown-Bchmidt, Thorpe, Gleitman, & Trueswell, 2000).

Two broad classes of explanations have been proposed for these lags between production and comprehension across different domains. One class of *methodological* explanations attributes the delay in comprehension to factors extrinsic to the specific language domain under study. Some researchers have proposed, for instance, that several comprehension measures are metalinguistic in nature, since they require children to reason explicitly about linguistic expressions (Davies & Katsos, 2010; de Villiers & Johnson, 2007; Johnson et al., 2005); others have hypothesized that measures of comprehension often tax children's domain-general processing (Grodzinsky & Reinhart, 1993; Reinhart, 2004) or pragmatic (Chien & Wexler, 1990) abilities. On this view, the delay in comprehension should diminish or disappear if task demands are modified to reduce challenges and/or to accommodate children's immature processing abilities. In support of this possibility, tasks with minimized memory and metalinguistic demands have revealed better comprehension performance in some areas. For instance, experiments using a head-turn procedure have demonstrated that 19-month-olds are sensitive to the third-person singular /s/ agreement morphology (Soderstrom, Wexler, & Jusczyk, 2002). Similarly, comprehension tasks that do not require children to explicitly evaluate the felicity of quantified statements such as *Some Xs Y* have revealed that children are, in fact, sensitive to the informativeness difference between *some* and *all* (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Guasti et al., 2005; Katsos & Bishop, 2011).

A second class of explanations takes asymmetries between production and comprehension to reveal a genuine psycholinguistic phenomenon, as opposed to an epiphenomenon that can be explained away by task demands. These *psycholinguistic* explanations begin with the assumption that the processes involved in production and comprehension are not simply the same computations executed in the reverse order (see, e.g., Hurewitz et al., 2000). In comprehension, the listener must unpack the meaning carried by the forms in incoming speech, and integrate the linguistic information with the information provided by the context and the speaker's intentions. Comprehension is thus a form of a mental guessing game. But in production, the speaker plans a message to convey an intended meaning through a particular form; and even though the hearer's needs may feed this process, the speaker plans and executes an utterance with her own goals, resources and perspective in mind. In adults, the mechanics underlying the production and comprehension processes have become exquisitely aligned – but in children, the same form embedded into these processes may be processed in radically different ways. Specific accounts in this broad class have explained children's failures to use modifiers (e.g., "the frog on the napkin") to disambiguate reference in comprehension as the result of the slow integration of weak, discourse-based cues such as modification into rapid parsing decisions (in combination with other factors; see Hurewitz et al., 2000). In other instantiations of psycholinguistic explanations, the production–comprehension gap in the third-person singular /s/ agreement and third-person pronouns such as *him* have been attributed to children's difficulty with the linguistic perspective-taking necessary to monitor the contribution of particular linguistic forms as opposed to other alternatives in comprehension (Hendriks, 2013; Smolensky, 1996; cf. Brandt-Kobele & Höhle, 2010; Hendriks & Koster, 2010; Hendriks & Spenader, 2005; Spenader, 2009).

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 describe the evidential system in Turkish and review past studies describing its acquisition.

The evidential system in Turkish: structure and acquisition

Languages indicate the source from which a piece of information was acquired through *evidentiality* markers. Some languages encode evidentiality through lexical means: in English, for instance, several verbs (*see*, *hear*, *infer*) and adverbials (*reportedly*, *allegedly*) encode evidential meanings. About one fourth of the world's languages encode evidentiality grammatically, mostly through morphological means (Aikhenvald, 2004, 2014; Aikhenvald & Dixon, 2001; Anderson, 1986; Chafe & Nichols, 1986; De Haan, 2001; DeLancey, 2001; Faller, 2001; Givón, 1982; Willet, 1988). For instance, in Turkish, two past-tense morphemes, *-dı* and *-mış* (realized as *-dı*, *-di*, *-du*, *-dü*, *-ti*, *-tı*, *-tu*, *-tü* and *-mış*, *-miş*, *-muş*, *-müş* respectively depending on phonological factors) have evidential meanings, i.e., they differentiate direct experience from indirect

experience (Aksu & Slobin, 1986; Aksu-Koç, 1988; Johanson, 2003; Kornfilt, 1997; Slobin & Aksu, 1982). All sentences in the past tense need to include one of these markers as verb suffixes. Thus, sentence (1a) indicates that the speaker has had first-hand experience of the event (typically, through visual perception), and sentence (1b) indicates that the speaker has indirectly acquired information about the event – either through hearsay or through inference.

(1)	a. Çocuk gel-di boy come-PAST dir.3sg The boy came (DIRECT) b. Çocuk gel-miş. boy come-PAST ind.3sg The boy came (INDIRECT)

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; de Villiers, Garfield, Gernet-Girard, Roeper, & Speas, 2009; Fitneva, 2009; Ozturk & Papafragou, 2007, 2015; Papafragou et al., 2007; Ünal & Papafragou, 2013a). 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 earlier discussion, and Ozturk & Papafragou, 2015). 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 & de Blois, 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.

Current study

In this paper, 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 1, we elicited evidential production using a task modeled after Aksu-Koç (1988). In Experiment 2, 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 3 and 4 we used novel, simpler tasks to assess the comprehension of evidentiality and compared the results to both Experiment 1 and 2.

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 5 we removed evidential language from the comprehension task of Experiment 4, and asked children to reason about either someone else's or their own evidence for information.

Experiment 1

The goal of Experiment 1 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 evi-

dence 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.

Materials

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 Fig. 1.

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 A.

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 1 (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, 2012). The fixed effects that were investigated in Experiment 1 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 2 and 3 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 by the first author. Beginning with Filler trials, both children and adults were highly accurate in labeling the objects ($M_{3\text{-year-olds}} = 0.93$, $M_{4\text{-year-olds}} = 0.93$, $M_{5\text{- to 6-year-olds}} = 0.96$, $M_{adults} = 0.94$).

For target trials, first we examined the proportion of non-past tense uses for Seen and Inferred events across the four Age groups (Table 1). Both children and adults

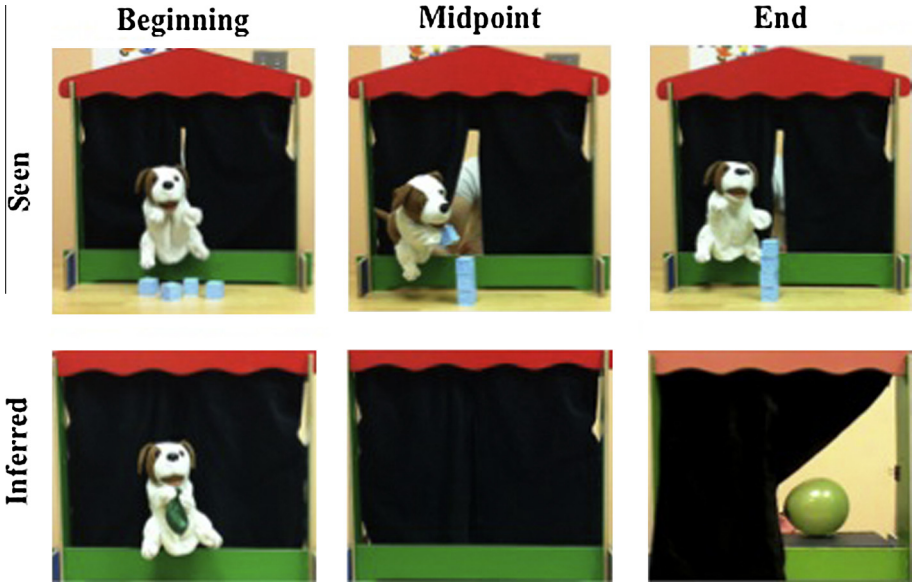

Fig. 1. Snapshots of the videos for Seen and Inferred Events in the Production task (Exp. 1). In the Seen event, the puppet first appears next to some blocks, then stacks the blocks, and at the end stands next to the stacked blocks. In the Inferred event, the puppet holds a deflated balloon, then goes behind the curtain; at the end, the curtain opens again and the balloon is inflated.

Table 1
Proportion of non-past tense uses for seen and inferred events (Experiment 1).

	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 2
Fixed effect estimates for multi-level model of non-past tense uses in Experiment 1.

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

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

sometimes used non-past tense descriptions, but this tendency was especially prominent in the youngest group of children.

Table 2 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.

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 3 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 4 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,

Table 3
Proportion of direct past tense (-dı) out of total past tense uses (Experiment 1).

	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 4
Fixed effect estimates for multi-level model of direct past tense use in target event descriptions in Experiment 1.

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 (3-year-olds)	−0.06	0.10	−0.57
Condition (Seen): Age (4-year-olds)	−0.17	0.09	−1.84
Condition (Seen): Age (5- to 6-year-olds)	−0.17	0.09	−1.84

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

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 ($p < .001$) (Table 2). 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, 2013a). 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 2

The goal of Experiment 2 was to test young Turkish learners' evidential comprehension using the events that reliably elicited evidential production from children of a comparable age in Experiment 1. 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.

Materials

As in Experiment 1, there were two types of stimuli. For Target trials, stimuli consisted of videos of a subset of the target events in Experiment 1.[1] Each target event had two versions: Seen and Inferred. As in Experiment 1, 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

[1] 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 2 elicited the intended evidentiality markers, the proportion of descriptions marked with direct past tense (-*dı*) out of total past tense descriptions in Experiment 1 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 1.

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 Fig. 2 for a schematic depiction). Next, the experimenter uttered a description with either the direct past tense *-dı* as in (2) or the indirect past tense *-mış* as in (3). Then, the experimenter asked: "Which one did I describe?"

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

Results

We followed the same model fitting and effect coding procedures as in Experiment1 (see Model fitting and results above in Experiment 1 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 ($M_{3\text{-year-olds}} = 0.88$, $M_{4\text{-year-olds}} = 0.97$, $M_{5\text{- to 6-year-olds}} = 0.98$, $M_{\text{adults}} = 1.00$). 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 5 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*).

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 6 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 1 and 2 demonstrated a production–comprehension asymmetry in the domain of evidentiality. In production (Experiment 1), 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

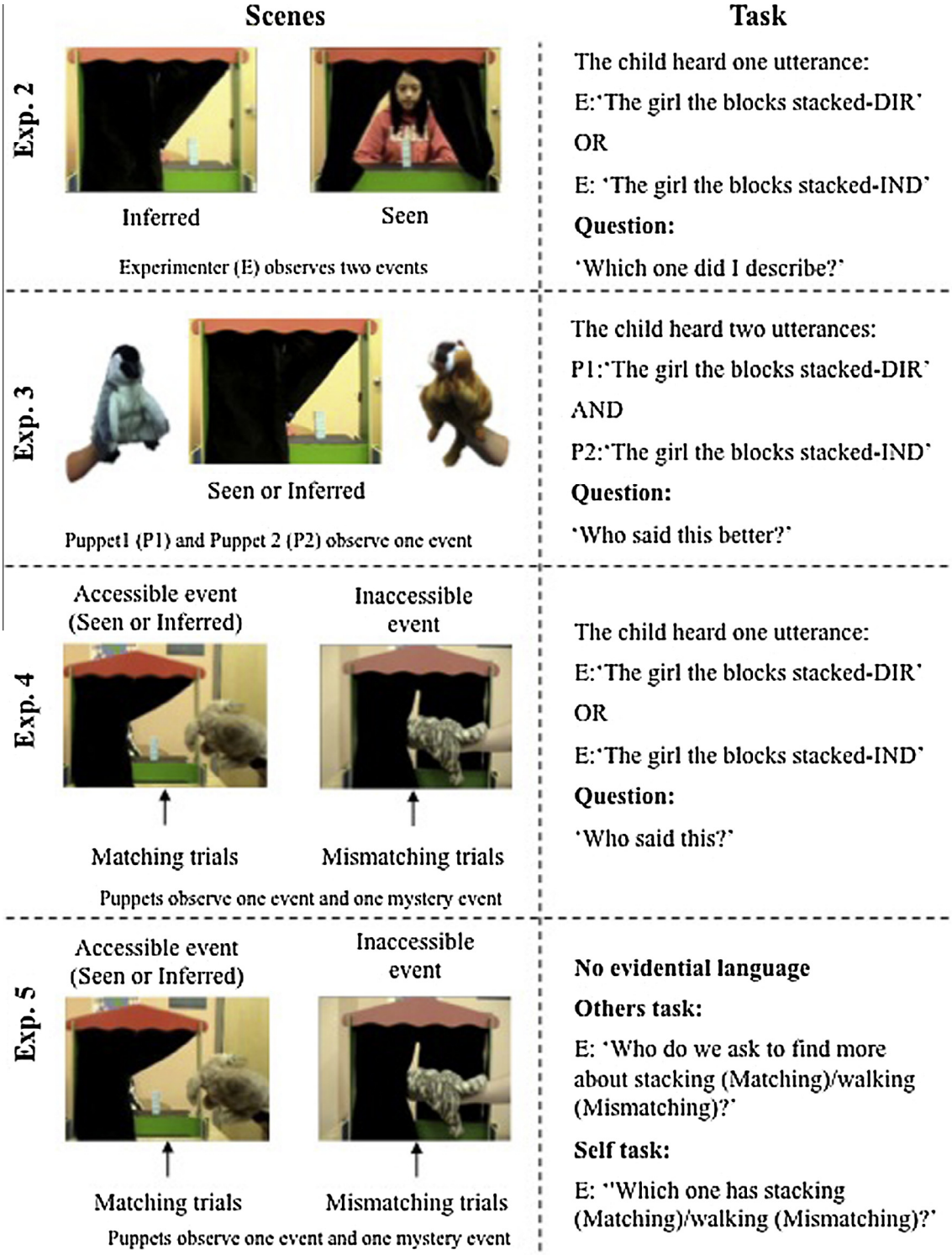

Fig. 2. Summary of experimental design for Exps. 2–5.

inferred from post-event visual evidence. In comprehension, however (Experiment 2), 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

Table 5
Proportion of picking the Seen event across Age groups (Experiment 2).

	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 6
Fixed effect estimates for multi-level model of picking the Seen event in Target trials in Experiment 2.

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. 3-year-olds)	0.81	0.16	5.15***
Condition (*-dı* vs. - *mış*): Age (Adults vs. 4-year-olds)	0.79	0.15	5.12***
Condition (*-dı* vs. - *mış*): Age (Adults vs. 5- to 6-year-olds)	0.60	0.17	3.59***

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

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–c omprehension 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 (Barner, Brooks, & Bale, 2011; Chierchia et al., 2001) 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 3–5, we tested these predictions more fully.

Experiment 3

Experiment 3 introduced a new task that asked children to consider the speaker's informational access in understanding evidentials but was simpler than Experiment 2. 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 2 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 3 compared to Experiment 2.

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.

Materials

Stimuli were the same as in Experiment 2. 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 Fig. 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 (2) above and the other one used the indirect past tense (*-mış*) morpheme as in (3) 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 1 and 2 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 ($\beta = 0.95$, $SE = 0.02$, $p < .001$) and no significant effect of Age. Even the youngest group of 3-year-olds were highly accurate on Filler trials ($M_{3\text{-year-olds}} = 0.87$, $M_{4\text{-year-olds}} = 0.96$, $M_{5\text{- to 6-year-olds}} = 0.97$, $M_{adults} = 1.00$).

Next, we examined performance in Target trials. Table 7 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 8 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

Table 7
Proportion of selecting the utterance marked with direct past tense across Age groups (Experiment 3).

	Adult		3-year-olds		4-year-olds		5-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 8
Fixed effect estimates for multi-level model of selecting the utterance marked with the direct past tense in Target trials in Experiment 3.

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. 3-year-olds)	−0.78	0.18	−4.36***
Condition (Inferred vs. Seen): Age (Adults vs. 4-year-olds)	−0.67	0.18	−3.68***
Condition (Inferred vs. Seen): Age (Adults vs. 5- to 6-year-olds)	−0.78	0.18	−4.36***

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

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.

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 3, 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 2, 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 4

The goal of Experiment 4 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 2 and 3. 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 2) or two evidential utterances to a single event (Experiment 3). 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.

Materials

We refilmed the Seen and Inferred versions of the 6 Target events in Experiments 2 and 3 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 B 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

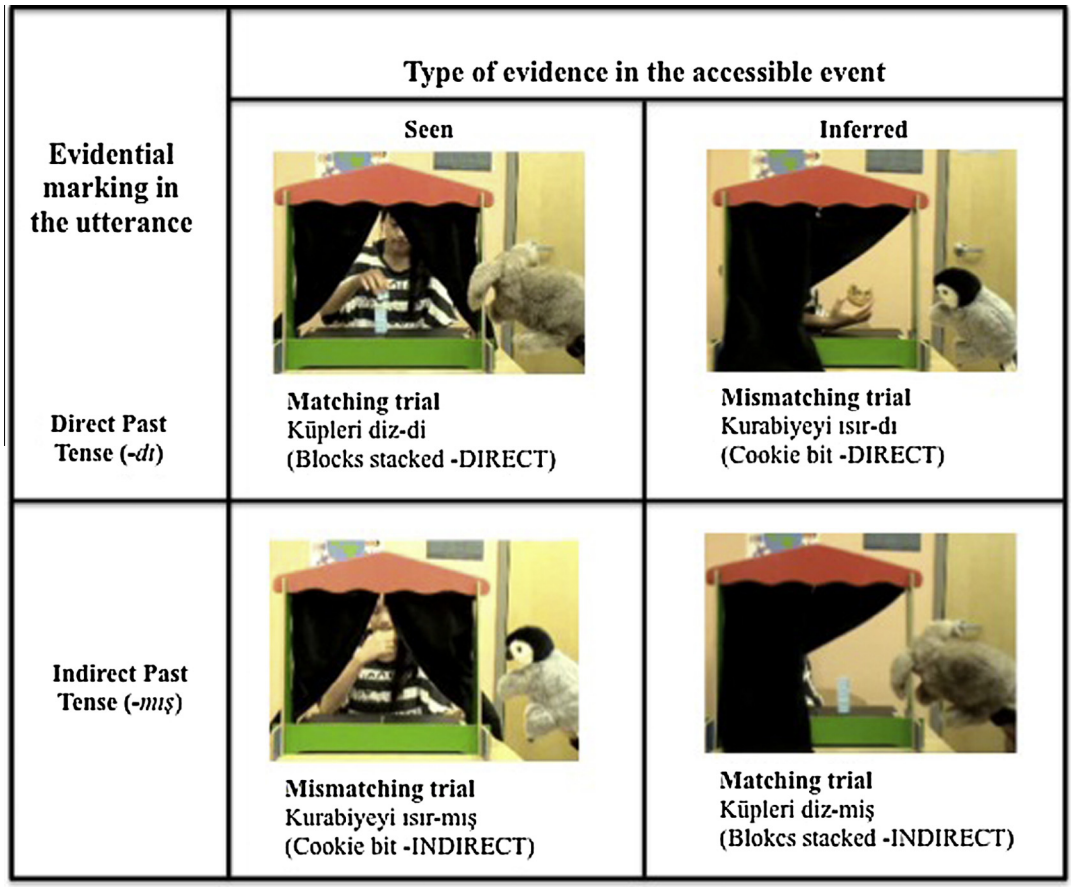

Fig. 3. Examples of matching and mismatching trials where the accessible event was Seen (right panel) or Inferred (left panel).

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 Fig. 3). 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 3).

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?".

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. Fig. 3 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 B).

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.

Table 9
Proportion of selecting the accessible event across types of evidence and evidential marking in Experiment 4.

	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 10
Fixed effect estimates for multi-level model of selecting the accessible event in Experiment 4.

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

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

Results

Table 9 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.

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 1–3 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 10 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 acces-

sible 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 ($t(20) = 2.01$, $p = .06$), while it should have been *below* chance level.

Discussion

Experiment 4 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 4 cohere with those of Experiments 2 and 3: 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 2, 3 and 4 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.[2]

Experiment 5

In Experiment 5, we turned to the conceptual underpinnings of linguistic evidentiality. We used the paradigm of Experiment 4 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; see Ünal & Papafragou, 2013a, 2013b).

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.

Materials

The visual stimuli were the same as in Experiment 4. A set of changes was made to the verbal stimuli (see Fig. 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 4 (e.g., when the accessible event involved either seeing a stacking action or inferring that stacking had occurred, children were asked about "to stack"/"dizm ek"). 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 B.

Procedure

The procedure was the same as Experiment 4 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

[2] A potential concern about Experiment 4 is that children resisted picking the "mystery" event because there was uncertainty or confusion about that option. Because the design and results of Experiment 5 address this concern, we postpone its discussion until the next section.

Table 11
Proportion of selecting the accessible event across evidence and verb types in the Others and Self Tasks in Experiment 5.

	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

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 11 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 4, all fixed factors were coded using centered contrasts (−0.5, 0.5).

Table 12 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).

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 4, but the likelihood of selecting the accessible event was significantly lower for Mismatching trials ($M = 0.44$) compared to Matching trials

Table 12
Fixed effect estimates for multi-level model of selecting the accessible event in the Others Task in Experiment 5.

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

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

Table 13
Fixed effect estimates for multi-level model of selecting the accessible event in the Self Task in Experiment 5.

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

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

($M = 0.66$) in the Others task of Experiment 5. Crucially, there was no effect of Experiment ($p = .9$) suggesting that the overall performance did not improve after evidential language was removed.

Table 13 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 5 used the design of Experiment 4 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 4. In the Self task, however, where children were asked to link evidence to event representations in their own mind, performance improved. Experiment 5 alleviates a concern with the design of Experiment 4, 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 4, the findings of Experiment 5 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. This unusual pattern has been previously attributed to either the processing/metalinguistic demands of comprehension tasks (Chien & Wexler, 1990; Davies & Katsos, 2010; de Villiers & Johnson, 2007; Grodzinsky & Reinhart, 1993; Johnson et al., 2005; Reinhart, 2004) or to psycholinguistic properties of comprehension itself (Hurewitz et al., 2000; cf. Hendriks, 2013; Smolensky, 1996 for a different approach). 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, 2013a); 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 the-

ater 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 1, 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, 2013a), 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).[3]

Yet findings of Experiment 2 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 significantly worse in evidential comprehension compared to adults who were at ceiling. These data confirm the previously documented production–comprehen sion 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 2 (where a single evidential utterance had to be compared to two events), Experiment 3 (where two evidentially marked utterances had to be compared to one event) and Experiment 4 (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 5 revealed that children's difficulty persisted in a version of Experiment 4 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 (Ozturk & Papafragou, 2015; Papafragou et al., 2007). 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;

[3] Recent data from Turkish-speaking adults show that the use of the indirect past tense is sensitive to the 'distance' between the representation formed by partial visual cues vs. the full visual perception of an event (Ünal, Pinto, Bunger, & Papafragou, 2016). Future work should assess whether similar factors affect the use of past-tense morphology in Turkish-learning children.

Povinelli & de Blois, 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). In those studies, children had to explicitly reflect on and verbalize the reasoning underlying others' knowledge acquisition. In our own studies (Experiment 5), we found that children had difficulty reasoning about others' knowledge sources even when they did not have to explicitly talk about others' knowledge states themselves.

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.

Acknowledgments

This work was supported by NSF Grant BCS0749870 to AP. EÜ was supported by a Graduate Fellowship awarded by University of Delaware Office of Graduate and Professional Education. We thank members of the Language and Cognition Lab at the University of Delaware for comments, Myeongyun Choi and Yasemin Sandıkçı for assistance with data collection and stimulus preparation, the Language and Communication Development Lab at Koç University and the preschools in Istanbul for facilitating data collection, and all of the children and adults that participated in these studies.

Appendix A

List of events/objects used in Experiment 1.

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 2–3. 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 4 and 5 alongside additional stimuli (see Appendix B).

Appendix B

List of Events and Verbs Used in Experiments 4 and 5. In Experiment 4, all verbs matched the events (but half of the evidential morphemes on the verbs did). In Experiment 5, only half of the verbs matched the events.

		Verbs (Experiment 4)	Verbs (Experiment 5)
Matching trial events			
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

References

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

Aikhenvald, A. Y. (2014). The grammar of knowledge: A cross-linguistic view of evidentials and the expression of information source. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *The grammar of knowledge: A cross-linguistic typology* (pp. 1–51). Oxford, United Kingdom: Oxford University Press.

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

Aksu, A., & Slobin, D. (1986). A psychological account of the development and use of evidentials in Turkish. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 159–167). Norwood, NJ: Ablex.

Aksu-Koç, A. (1988). *The acquisition of aspect and modality: The case of past reference in Turkish*. Cambridge, United Kingdom: Cambridge University Press.

Aksu-Koç, A., 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. <http://uvt.ulakbim.gov.tr/proje/>.

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

Anderson, 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.

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.cognition.2010.10.010.

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

Bates, D. M. (2005). Fitting linear mixed models in R. *R News, 5*, 27–30. <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, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York, NY: Cambridge University Press.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.99375-39. <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. http://dx.doi.org/10.18637/jss.v067.i01.

Bloom, P., Barss, A., Nicol, J., & Conway, L. (1994). Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language, 70*, 53–71. http://dx.doi.org/10.2307/416740.

Brandt-Kobele, O. C., & Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: The case of verb inflection in language acquisition. *Lingua, 120*, 1910–1925. http://dx.doi.org/10.1016/j.lingua.2010.02.008.

Chafe, W. L., & Nichols, J. (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.

Chien, Y. C., & Wexler, K. (1990). Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1*, 225–295. http://dx.doi.org/10.1207/s15327817la0103_2.

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.

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.

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua, 120*, 1956–1972. http://dx.doi.org/10.1016/j.lingua.2010.02.005.

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

DeLancey, S. (2001). The mirative and evidentiality. *Journal of Pragmatics, 33*, 369–382. http://dx.doi.org/10.1016/S0378-2166(01)80001-1.

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.

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

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. http://dx.doi.org/10.1016/j.jml.2013.06.003.

Givón, T. (1982). Evidentiality and epistemic state. *Studies in Language, 6*, 23–49. http://dx.doi.org/10.1075/sl.6.1.03giv.

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

Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry, 24*, 69–102. http://dx.doi.org/10.1162/089892998562708.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children sometimes (but not always) compute implicatures. *Language and Cognitive Processes, 20*, 667–696. http://dx.doi.org/10.1080/01690960444000250.

Hendriks, P. (2013). *Asymmetries between language production and comprehension*. London, United Kingdom: Springer. http://dx.doi.org/10.1007/978-94-007-6901-4.

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

Hendriks, P., & Spenader, J. (2005). Pronouns when production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition, 13*, 319–348. http://dx.doi.org/10.1207/s15327817la1304_3.

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

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. http://dx.doi.org/10.1023/A:1026468209238.

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. http://dx.doi.org/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)* (Vol. 54, pp. 273–290). Amsterdam, Netherlands & Philadelphia, PA: John Benjamins.

Johnson, V. E., de Villiers, J. G., & Seymour, H. N. (2005). Agreement without understanding? The case of third person singular /s/. *First Language, 25*, 317–330. http://dx.doi.org/10.1177/0142723705053120.

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicatures. *Cognition, 120*, 67–81. http://dx.doi.org/10.1016/j.cognition.2011.02.015.

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. http://dx.doi.org/10.1016/0885-2014(94)90010-8.

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

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

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language, 17*, 457–472. http://dx.doi.org/10.1017/S0305000900013866.

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

Noveck, I. A. (2001). When children are more logical than adults. *Cognition, 78*, 165–188. http://dx.doi.org/10.1016/S0010-0277(00)00114-1.

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. (2014). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, 1–24. http://dx.doi.org/10.1080/15475441.2014.905169.

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

Papafragou, A., Li, P., Choi, Y., & Han, C.-H. (2007). Evidentiality in language and cognition. *Cognition, 103*, 253–299. http://dx.doi.org/10.1016/j.cognition.2006.04.001.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics/pragmatics interface. *Cognition, 86*, 253–282. http://dx.doi.org/10.1016/S0010-0277(02)00179-8.

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

Pillow, B. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development, 73*, 779–792. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1037//0012-1649.36.2.169.

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. http://dx.doi.org/10.1037/0735-7036.106.3.228.

R-Core-Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria.

Reinhart, T. (2004). The processing cost of reference-set computation: Acquisition of stress shift and focus. *Language Acquisition, 12*, 109–155. http://dx.doi.org/10.1207/s15327817la1202_1.

Rett, J., & Hyams, N. (2014). The acquisition of syntactically encoded evidentiality. *Language Acquisition, 21*, 173–197. http://dx.doi.org/10.1080/10489223.2014.884572.

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.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry, 27*, 720–731. <http://www.jstor.org/stable/4178959>.

Soderstrom, M., Wexler, K., & Jusczyk, P. (2002). English-learning children's sensitivity to agreement morphology in receptive grammar. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th annual Boston University conference on language development* (pp. 643–652). Somerville, MA: Cascadilla Press.

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

Spenader, J. (2009). Coherent discourse solves the pronoun interpretation problem. *Journal of Child Language, 36*, 23–52. http://dx.doi.org/10.1017/S0305000908008854.

Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2003). The role of the input in the acquisition of third person singular verbs in English. *Journal of Speech, Language and Hearing Research, 46*, 863–877. http://dx.doi.org/10.1044/1092-4388(2003/067).

Ünal, E., & Papafragou, A. (2013a). Linguistic and conceptual representations of inference as a knowledge source. In S. Baiz, N. Goldman, & R. Hawkes (Eds.), *Proceedings of the 37th annual Boston University conference on language development* (pp. 433–443). Somerville, MA: Cascadilla Press.

Ünal, E., & Papafragou, A. (2013b). Perception and inference as sources of knowledge in oneself and others. *Poster presented at 8th biennial meeting of cognitive development society*, Memphis, TN.

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.

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. http://dx.doi.org/10.1002/cd.248.

de Villiers, J. G., & Johnson, V. E. (2007). The information in third-person /s/: Acquisition across dialects of American English. *Journal of Child Language, 34*, 133–158. http://dx.doi.org/10.1017/S0305000906007768.

Willet, T. (1988). A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language, 12*, 51–97. http://dx.doi.org/10.1075/sl.12.1.04wil.

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

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. <http://linguistics.ucla.edu/people/grads/lwinans/papers/NELS-paper.pdf>.