Cognition 197 (2020) 104194

Contents lists available at ScienceDirect

Cognition

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

# Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality

Dionysia Saratsli[a,*], Stefan Bartell[a], Anna Papafragou[b]

[a] Department of Linguistics and Cognitive Science, University of Delaware, United States of America
[b] Department of Linguistics, University of Pennsylvania, United States of America

ARTICLE INFO

*Keywords:*
Evidentiality
Learnability
Artificial language learning
Semantics
Typology
Pragmatics

ABSTRACT

It is often assumed that cross-linguistically more prevalent distinctions are easier to learn (*Typological Prevalence Hypothesis*; TPH). Prior work supports this idea in phonology, morphology and syntax but has not addressed semantics. Using Artificial Language Learning experiments with adults, we test predictions made by the TPH about the relative learnability of semantic distinctions in the domain of evidentiality, i.e., the linguistic encoding of information source. As the TPH predicted, when exposed to miniature evidential morphological systems, adult speakers of English whose language does not encode evidentiality grammatically learned the typologically most prevalent system (marking indirect, reportative information) better compared to less-attested systems (Experiments 1–2). Similar patterns were observed when non-linguistic symbols were used to encode evidential distinctions (Experiment 3). Our data support the conjecture that some semantic distinctions are marked preferentially and acquired more easily compared to others in both language and other symbolic systems.

## 1. Introduction

### 1.1. Cross-linguistic frequency and learnability of semantic distinctions

It is often assumed in the literature that cross-linguistically more frequent distinctions are easier to learn than less frequent ones (Bowerman, 1993; Clark, 1976; Jakobson, 1971; Pinker, 1984; Rosch, 1972; Slobin, 1985). This idea has been captured effectively by Gentner and Bowerman's (2009) Typological Prevalence Hypothesis (TPH): "All else being equal, within a given domain, the more frequently a given way of categorizing is found in the languages of the world, the more natural it is for human cognizers, hence the easier it will be for children to learn" (p. 467). Even though this idea is fundamental for how human beings learn to encode meaning in language, empirical tests of this hypothesis within the semantic domain have so far been rare.

Gentner and Bowerman (2009) themselves tested this hypothesis by comparing how English-speaking and Dutch-speaking children acquire their native language's spatial support prepositions. English encodes spatial support with the single preposition *on* but Dutch uses three different prepositions (*op*, *aan*, *om*) to encode the same semantic space. Because the English option is typologically more common, the TPH predicts that the English system should be more easily learned than the Dutch system. Gentner and Bowerman elicited spatial prepositions from young learners of the two languages and the results supported this prediction. However, the slower acquisition rate could be due to the higher number of prepositions (and/or the lower frequency of each individual preposition) in Dutch compared to English rather than the specific semantic distinctions encoded within each system. This fact complicates the interpretation of Gentner and Bowerman's results and hence the evidence in favor of TPH. Similar issues are bound to arise for all tests of TPH that involve comparisons of how young learners acquire semantic systems of variable complexity cross-linguistically.

In this paper, we offer a new test of TPH using an Artificial Language Learning Paradigm. This type of experimental design typically requires participants to learn different versions of a target language that differ minimally from each other in terms of a grammatical or lexical feature (see Folia, Uddén, Vries, Forkstam, & Petersson, 2010 for a review). During an initial learning phase, participants are exposed to the grammar/lexicon of the artificial language. The learning phase is followed by a test phase in which the extent to which participants learned the linguistic target is assessed. This paradigm offers a unique opportunity to explore the participants' learning process in relation to a specific linguistic feature of interest (Fedzechkina, Newport, & Jaeger, 2016). By having participants learn minimally different versions of the same artificial language, one can bypass the role of frequency or complexity in the learnability of actual systems in individual languages. Moreover, it is possible to test adults (as well as children) on the target artificial language, to the extent that learnability patterns could reveal

* Corresponding author at: Department of Linguistics and Cognitive Science, University of Delaware, 125 E Main St., Newark, DE 19716, United States of America.
*E-mail address:* dsaratsl@udel.edu (D. Saratsli).

https://doi.org/10.1016/j.cognition.2020.104194
Received 23 April 2019; Received in revised form 26 November 2019; Accepted 16 January 2020
Available online 25 January 2020

biases in both novice and more experienced learners.

Previous studies using an Artificial Language Learning paradigm have confirmed that cross-linguistically common distinctions are learned more easily than less common ones in the domains of syntax (Christiansen, 2000; Newport & Aslin, 2004; Hudson Kam & Newport, 2005, 2009; Thompson & Newport, 2007; Wonnacott, Newport, & Tanenhaus, 2008; Tily, Frank, & Jaeger, 2011; Culbertson, 2012; Culbertson & Smolensky, 2012; Culbertson, Smolensky, & Legendre, 2012), phonology (Finley & Badecker, 2009; Seidl & Buckley, 2005; Wilson, 2006) and morphology (Fedzechkina, Jaeger, & Newport, 2012; Merkx, Rastle, & Davis, 2011) but have drawn little upon semantics (Carstensen, Xu, Smith, & Regier, 2015; Maldonado & Culbertson, 2019; Xu, Dowman, & Griffiths, 2013). Here we present one of the first systematic tests of this hypothesis within the domain of semantics (which was the main emphasis of the TPH). We focus on evidentiality, i.e., the linguistic encoding of information source. Evidential distinctions are not grammaticalized in English and hence can be taught to English-speaking adults within an Artificial Language Learning paradigm without native language interference.

### *1.2. Evidentiality and the TPH*

Evidentiality refers to the way that language marks the speaker's source of information, for instance, whether the speaker had direct perceptual access to an event, inferred what happened based on some clues, or was told what happened by someone else. Languages differ in the way they encode evidentiality (Aikhenvald, 2004, 2014, 2018; Aikhenvald & Dixon, 2003, 2014; Chafe & Nichols, 1986; Cinque, 1999; Delancey, 2002; Ifantidou, 2001; Johanson & Utas, 2000; Kratzer, 1991; McCready, 2008; McCready & Ogata, 2007; Mushin, 2001; Willett, 1988). English and other languages make use of lexical means such as verbs (e.g., *see*, *hear*, *infer*) or adverbs (e.g., *allegedly*) to mark information sources. About a quarter of the world's languages, however, use grammatical morphemes to indicate information sources (Faller, 2012, 2014; Garrett, 2000; Izvorski, 1998; Matthewson, 2012; Speas, 2004, 2018). For example, Wanka Quechua has separate grammatical morphemes for three types of information source (Aikhenvald, 2004): *-mi* in (1) marks the speaker's visual experience, *-chr-* in (2) marks an inference drawn by the speaker, and *–shi* in (3) marks another person's report about what happened.

(1) Chay-chruu-**mi** achka wamla-pis walashr-pis alma-ku-lkaa-ña.
this-LOC-**DIR.EV** many girl-TOO boy-TOO bathe-REEL-IMPF.PL-NARR.PAST.
'Many girls and boys were swimming' (I saw them).

(2) Daañu pawa-shra-si ka-ya-n-**chr**-ari.
Field finish-PART-EVEN be-IMPF-3-**INFR**-EMPH.
'It (the field) might be completely destroyed' (I infer).

(3) Ancha-p-**shi** wa'a-chi-nki wamla-a-ta.
too.much-GEN-**REP** cry-CAUS-2 girl-1P-ACC.
'You make my daughter cry too much' (they tell me).

Semantically, the broad meaning dimensions that characterize grammatical evidentiality across languages involve perception, inference and verbal report (Aikhenvald, 2004, 2014, 2018; de Haan, 2013; Willett, 1988). Perception mostly includes visual access but also non-visual perception (information acquired through hearing that may also extend to other senses; Aikhenvald, 2018). Inference has several sub-types, including inference from visual premises, general knowledge or other types of reasoning. Verbal report covers cases where the source of the report is left unspecified (as in hearsay), and other cases where a specific entity is quoted. Languages vary in both the complexity of their evidential systems (i.e., the evidential categories they mark and hence the number of distinct evidential morphemes they have) and the way they group semantic distinctions (Aikhenvald, 2004; Speas, 2018). For instance, as discussed in more detail below, very often languages have a single evidential morpheme but leave other sources unmarked. Some languages have only two grammatical morphemes to mark information source, a direct/firsthand morpheme that may encode perceptual sources and an indirect/non-firsthand morpheme that may cover both inference and report (Aikhenvald, 2004). Others encode all three types of information source with dedicated morphemes, as in the examples (1)–(3) above. Yet other languages have more complex 4-way or 5-way evidential systems that subdivide these broad categories into more specialized distinctions (ibid.).

Pragmatically, the use of evidentially marked utterances can give rise to contextual implications about speaker certainty or reliability, even though evidential morphemes do not denote these dimensions as part of their core meaning (Speas, 2018). The pragmatic profile of evidentials is captured by current formal theories, according to which the use of evidentiality introduces a proposal to add the base proposition to the common ground but whether the addressee accepts this addition depends on how the addressee evaluates the nature of the evidence (Murray, 2011, 2014; cf. AnderBois, 2014). In this sense, linguistic evidentiality connects to the human ability to reason about sources of information (see Papafragou, Li, Choi, & Han, 2007; Ünal & Papafragou, 2018, for extensive discussion). In cognition, direct perceptual experience of an event is regarded as a more reliable source due to its assumed correspondence with reality; indirect sources of information such as inference or reports are often deemed less reliable in the sense that the former may be based on incomplete premises while the latter depends on the informant's reliability (Papafragou et al., 2007; Dancy, 1985; Matsui & Fitneva, 2009; Aikhenvald, 2018; Wiemer, 2018; Koring & De Mulder, 2015). These intuitions emerge early: even at the age of 3, young children know that seeing leads to knowing (Pratt & Bryant, 1990; Pillow, 1989) but identifying the contribution of other sources may be a more protracted development (Ozturk & Papafragou, 2016). Furthermore, children prioritize seeing over inference or hearsay when choosing whether someone else is knowledgeable about an event (Ozturk & Papafragou, 2016), even though these preferences are flexible (Brosseau-Liard & Birch, 2011; Fitneva, Lam, & Dunfield, 2013). Depending on the specific alternatives available within an evidential system in a language, then, the use of an indirect evidential might indicate that the speaker lacks direct evidence for an assertion (Speas, 2018) and might further convey tentativeness, doubt, or distance from the event. For the same reason, when evidence about an event is available from multiple sources, languages exhibit a preference for encoding it with visual evidentials, when possible, while using reportative evidentials is the least preferred option (Aikhenvald, 2018). As Aikhenvald notes, communicating an event that the speaker has experienced "is considered a better choice rather than reporting what they [the speaker] heard from someone else" (2004: 307). This emphasizes "the preference for visual information source – if it is available" and also speaks for the reliability of visual access, other things being equal (Aikhenvald, 2018: 27).

For present purposes, a key fact about grammatical evidentiality is that, across languages, there is widespread preference for evidential systems to mark indirect compared to direct access to information (de Haan, 2013; Aikhenvald, 2004, 2014, 2018). For instance, the World Atlas of Languages lists 161 languages worldwide that grammatically mark only indirect evidence, 71 languages that mark both direct and indirect evidence and no languages that mark direct evidence alone (de Haan, 2013). Similarly, according to Aikhenvald (2004, 2014, 2018), many languages have a dedicated reportative morpheme (or, less frequently, an indirect/non-first hand morpheme that can cover inference or reports) but leave other sources unmarked. Systems that mark only direct (visual) access with dedicated grammatical devices but do not grammatically encode other source types are unattested and thus in all likelihood rare (ibid.). As a result, in many languages, "the least formally marked verb in a language with evidentiality tends to acquire a visual, or firsthand reading" (Aikhenvald, 2018: 16, even though distinguishing such meanings from source-neutral meanings can be difficult).

The semantic domain of evidentiality offers an excellent test case for the TPH. Given the frequency patterns described above, the TPH predicts that a simple evidentiality system that only marks reported information would be more easily learnable compared to a system that only marks visually acquired information (with a system marking only inferred information falling in between). Furthermore, if the typological frequency of evidential systems is linked to broader, non-linguistic factors (e.g., the naturalness of drawing certain conceptual distinctions, as proposed by TPH, and/or the naturalness of communicating such distinctions), similar pressures might affect other symbolic (but non-linguistic) systems of marking information access.

### 1.3. Current experiments

The aim of the current experiments was two-fold. First, we used an Artificial Language Learning paradigm to assess the learnability of different evidential morphological systems and test predictions of the TPH (Experiments 1 and 2). Second, we explored whether the learnability patterns for evidential morphology would generalize to non-linguistic symbols to shed light on the nature and origins of the learnability facts (Experiment 3).

Specifically, in our first two experiments, we tested evidential distinctions in an Artificial Language Learning task using adult native speakers of English. Participants were exposed to scenarios in which characters experienced a series of events. Access to the events was manipulated within-subjects such that a character might (a) see an event directly, (b) see visual clues and infer that a certain event had occurred, and (c) be informed by an observer about the occurrence of an event. The character described each event in an artificial language containing a novel evidential morpheme (*ga*). Because prior studies with adults had indicated that mental state meanings for novel predicates are hard to infer from observing events in the world without specific syntactic support (Gillette, Gleitman, Gleitman, & Lederer, 1999; Papafragou, Cassidy, & Gleitman, 2007), and the acquisition of evidentiality in children is protracted cross-linguistically (Aksu-Koç, 1988; Aksu-Koç & Alici, 2000; Aksu-Koç, Ögel-Balaban, & Alp, 2009; de Villiers, Garfield, Gernet Girard, Roeper, & Speas, 2009; Fitneva, 2008, 2018; Lee & Law, 2000; Matsui, Miura, & McCagg, 2006; Matsui, Yamamoto, & McCagg, 2006; Ozturk & Papafragou, 2016; Papafragou, Cassidy, & Gleitman, 2007; Ünal & Papafragou, 2018; Uzundag, Taşçı, Küntay, & Aksu-Koç, 2018; Winans, Hyams, Rett, & Kalin, 2014), we anticipated that acquiring evidential meanings would be challenging. We therefore simplified the artificial language that participants were exposed to by preserving the lexicon of English but changing the morphosyntax (the new language lacked determiners and had a Subject-Object-Verb word order). The evidential morpheme appeared verb-finally (*The girl cups stackedga*). The type of access marked by the morpheme was manipulated between-subjects such that the morpheme marked a single access type and the other two types received null marking. Thus, depending on the marked meaning, each participant was exposed to a Visual, Inferential, or Reportative evidential system (see Table 1), and had to learn what the morpheme meant. Of interest

**Table 1**
Structure of evidential systems in Artificial Language Learning studies (Exps. 1 and 2).

| Evidentiality system | Speaker's access to event | | |
|---|---|---|---|
| | Visual (Visual Perception) | Inferential (Inference from Visual Premises) | Reportative (Testimony from Others) |
| Visual | *ga* | – | – |
| Inferential | – | *ga* | – |
| Reportative | – | – | *ga* |

was whether, as predicted by the TPH, participants would succeed with Reportative systems more than others.

In our final experiment, we focused on Artificial Symbol Learning. The design was similar to our earlier paradigms but the character's description of the event was in English instead of the artificial language, and the participants' task was to acquire the meaning of a non-linguistic symbol that served as an evidentiality marker (instead of the morpheme *ga*). If the results of both the morpheme and the symbol learning experiments converge, we will have evidence that both the typological frequency and learnability facts about evidentials are motivated by broader considerations of which distinctions are 'natural' to encode or convey.

## 2. Experiment 1

### 2.1. Methods

#### 2.1.1. Participants

We recruited 111 participants in total, between the ages of 18 and 22 (10 male and 101 female). All participants were undergraduate students at the University of Delaware and were enrolled in an Introductory Psychology course that awarded credit for their participation. They were native speakers of English. Nine participants reported speaking a different language at home but none of these languages included grammaticalized evidentials.

#### 2.1.2. Stimuli

Our stimuli consisted of 39 scenarios showing a puppet (henceforth, the "Agent") enact an event on the stage of a puppet theater. Another puppet (henceforth, the "Speaker") experienced this event in more or less direct ways (sometimes from the other side of the stage or through an "Informant") and later described it. Each scenario involved a change of state and unfolded over four static pictures presented simultaneously for 20 s (see Fig. 1). There were 3 versions of each scenario depending on the Speaker's access to the event.

(A) In the Visual Access version, the stage curtains remained open and the Speaker could see the event in its entirety. Specifically, the Speaker (e.g., the dog) saw the Agent stand next to some objects (e.g., the gorilla next to the board; panel A1 in Fig. 1), manipulate the objects (color the star on the board, panels A2-A3) and stand next to the event aftermath (the gorilla standing next to the colored star, panel A4).

(B) In the Inferential Access version, the Speaker (e.g., the dog) could initially see the Agent and objects, just as in the Visual Access version (e.g., the gorilla and the board with the star, see panel B1). Then the stage curtains closed (panels B2-B3) and later opened to reveal (what presumably was) the aftermath of the event (panel B4; the gorilla next to the colored star). Thus, the Speaker could reconstruct the main event via visual cues.

(C) In the Reportative Access version, an Informant (e.g., the dog in panel C1) watched the entire event from the other side of the stage. The Informant saw the Agent stand next to the objects (e.g., the gorilla next to the board, panel C1), and manipulate the objects (coloring the star, panel C2 and C3). Later the Informant verbally informed the Speaker (e.g., the cat) about what had happened (panel C4; see the speech bubble with a snapshot of the event). During the act of reporting the stage curtains were closed so the Speaker would only have verbal information about the event.

All scenarios and versions followed the above basic structure. A complete list of the events in our stimuli can be found in Appendix A. For each of the roles of Agent, Speaker and Informant we used between 5 and 7 different animal puppets.

We also created a description of each event that appeared in a speech bubble next to the Speaker. The speech bubble appeared after

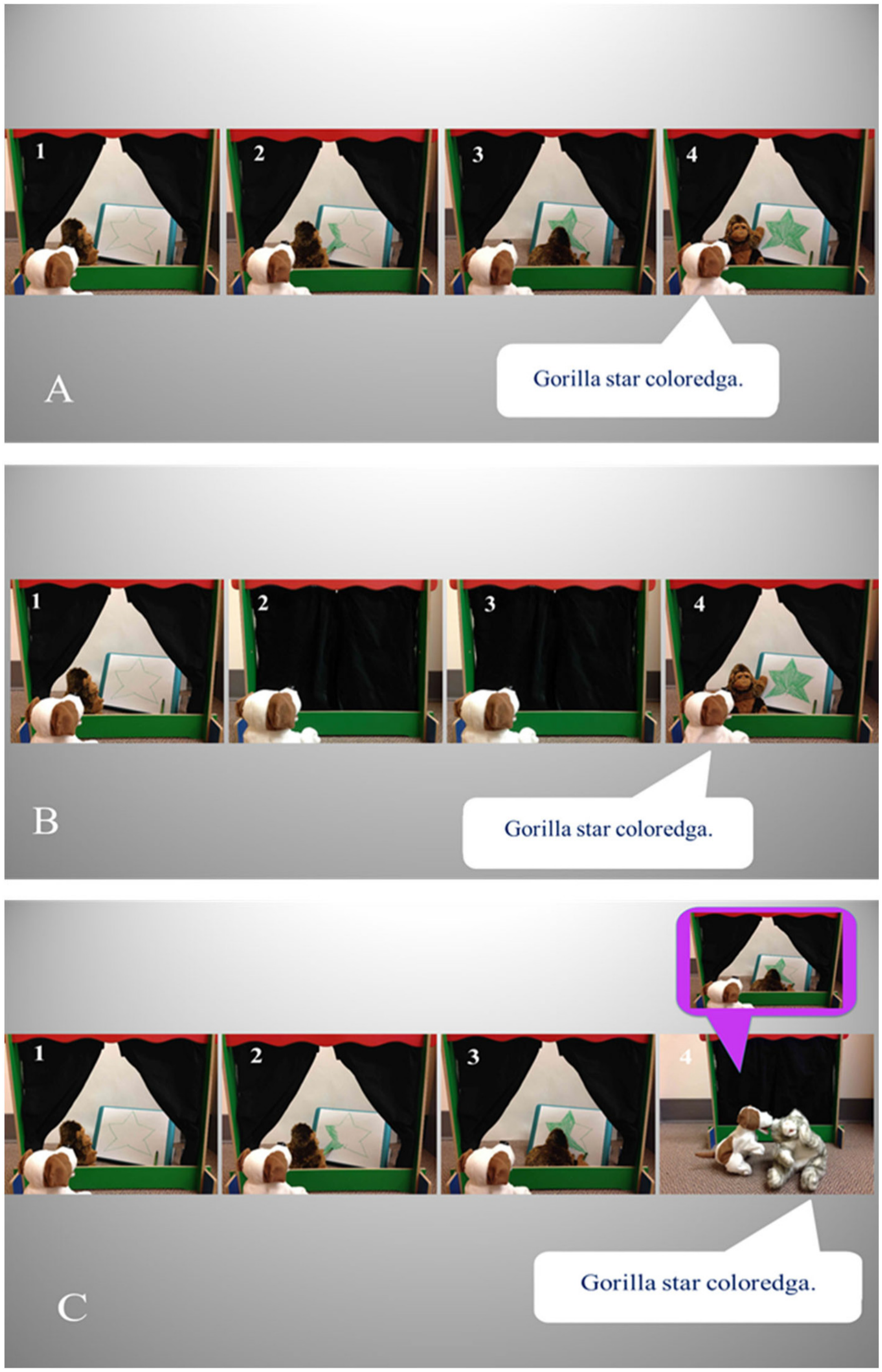


**Fig. 1.** Sample stimuli for the Artificial Language Learning task of Experiment 1. Versions of the same scenario are shown in which a character gained access to an event in one of three ways: (A) Visual, (B) Inferential, and (C) Reportative. At the end, the main character offered a description of the event with or without an evidential (e.g., "Gorilla star coloredga.").

the pictures within a scenario were displayed for 8 s and remained onscreen together with the pictures for 10 s. The event description was constructed in an artificial language (see Fig. 1). As mentioned already, this language shared the English lexicon (but lacked determiners) and had a Subject-Object-Verb word order. We created two versions of each description: one with a novel verb-final morpheme, *ga*, that served as an evidentiality morpheme (*Gorilla star coloredga*) and one without it (*Gorilla star colored*).

Of the 39 scenarios, 21 were used during the Training Phase and 18 during the Testing Phase of the experiment. Since each scenario had 3 versions depending on the Speaker's Access to the event, we created 3 different lists for each Phase by selecting an equal number of scenarios from each Access type and rotating the scenarios across lists so that no list contained the same scenario in more than one version. For each list, we created 3 randomized presentation orders, thus ending up with 9 presentation lists in total, each with 21 items for the Training Phase (7 from each Access type), and 18 for the Testing Phase (6 from each Access type). For the 18 scenarios in the Testing Phase, within each Access type, in half of the scenarios, the Speaker described the event erroneously used the morpheme *ga*, either failing to use the morpheme when it was needed or using it for the wrong type of Access. For the remaining half of the videos, the use of the morpheme was correct.

To instantiate different evidential systems during the Training Phase, only scenarios of a single Access type (Visual, Inferential, Reportative) per list were described with the morpheme *ga* to reflect one of three evidential systems (Visual, Inferential, or Reportative). Scenarios of the other two Access types were described by plain (unmarked) sentences during the Training Phase. We assigned participants randomly to one of 3 between-subjects conditions depending on the evidential system they were exposed to during training (such that 37 participants were assigned to each evidentiality system).

### 2.1.3. Procedure

We tested participants in small groups in a dimly lit, quiet room. Stimuli were displayed using PowerPoint, Microsoft Office on a Windows laptop computer and projected on a large screen. In the beginning of the experiment, participants were given an individual response sheet and told that the experiment was comprised of two phases: a learning phase during which they would not have to write anything down but would only have to pay close attention to the pictures shown and try to figure out what was happening. In the second part, they would have to fill out their answers on the response forms following the instructions they would be given before the actual part began.

In the beginning of the Training Phase, participants were shown the following instructions:

> "In this experiment you will watch a series of events in a puppet theater. For each event, there will be a character outside the puppet theater that will learn about the event. Here is what the puppet theater will look like:"

At this point, a single picture of the puppet theater was shown with two puppets, one on each side of the puppet theater in an effort to familiarize participants with the setting. Subsequently, participants were told that "each event would be depicted through four static clips shown next to each other from left to right in the order in which they unfold" and they were shown a sample event. This sample event was always the same and was structured in the same way as the main scenarios (a bear spilled candies while a donkey watched the complete event as the curtains of the puppet theater were open).

On the next instruction screen, participants received the following information:

> "At the end of each event, this outside character will offer a description of the event. The description will appear in a speech bubble next to the character. The language that the characters will be using is an alien language: it shares some words with English but is different in several ways. One difference is that the language includes a special word, "ga".

Subsequently, participants were told that they would have to pay attention "to when 'ga' is and is not used in the language to figure out what it means" and that they would be given several events to go through in order to figure this out. Each instruction screen was displayed for approximately 20s. Then the display continued to the main part of the Training Phase of the experiment during which participants were exposed to events in which the morpheme *ga* was used to mark only a single type of Access (Visual, Inferential or Reportative) depending on the Evidentiality System that each group of participants was randomly assigned to. When the Training Phase was completed, participants were informed that they would continue to the second part of the experiment.

The Testing Phase began with the following instructions to participants:

> "We will now show you a new set of events in the same puppet theater. As before, for each event, an outside character will offer a description of the event in the alien language. But this time some of these descriptions will contain errors: they will contain "ga" when it is not appropriate or omit it when it is needed. For each event, please mark "Yes" or "No" on your answer sheet to indicate whether the character described the event using the alien language correctly."

Following the instruction screen, participants watched one scenario at a time and they were given 30 s at the end of each scenario to fill in their answer (the question of whether the character used *ga* correctly or not was repeated on this response form and participants had to indicate "Yes" or "No"). In the Testing Phase, half of the scenarios within each Access type were marked with *ga* and the rest were unmarked. This meant that half of the time the puppet used the alien language correctly (by including or omitting *ga* as appropriate) and the other half the puppet was incorrect by using the morpheme for the wrong type of Access or failing to use the morpheme for the target Access type. When the experiment was over, participants were asked to write down what they believed *ga* meant and when it was or was not used.

## 2.2. Results

Participants' responses were coded for Accuracy, a binary outcome variable coded as 1 (Correct) or 0 (Incorrect). A summary of our data can be seen in Fig. 2. The data were analyzed using logistic mixed-effects modeling (Baayen, 2008; Baayen, Davidson, & Bates, 2008). Since our data include categorical variables, the generalized binomial linear mixed effects modeling (glmer) function of the lme4 package was used for our analysis (Bates, Mächler, Bolker, & Walker, 2015) in the R Project for Statistical Computing (R Core Team, 2018).

Our mixed-effects model included Accuracy as the dependent variable and System (Visual, Inferential, Reportative) as our fixed predictor. To explore the possibility that non-theoretically driven predictors affected participants' performance, we included List (1, 2, 3) as a second fixed predictor along with its interaction with System. Our model also included random intercepts for Participants and Items. The fixed effect of System was assessed with two planned comparisons using contrast

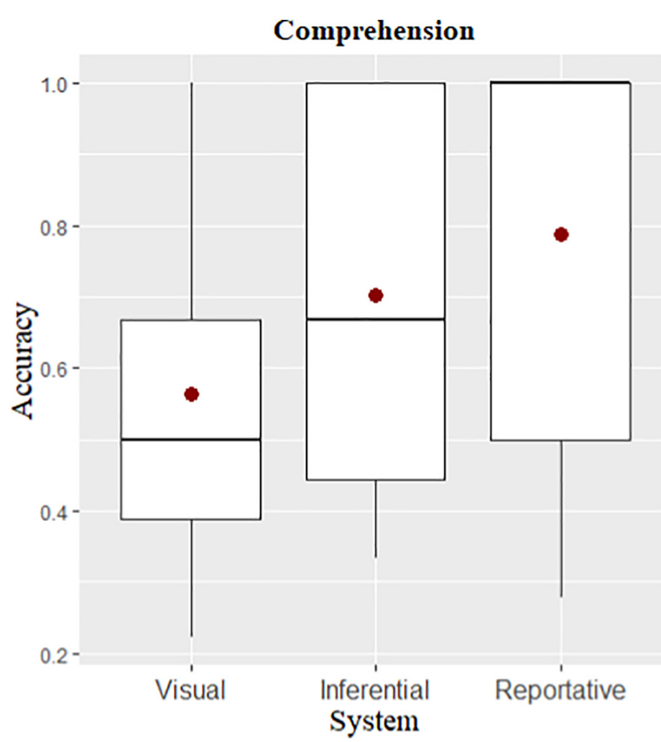


**Fig. 2.** Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in the Artificial Language Learning task of Experiment 1.

**Table 2**
Parameter estimates for Accuracy in the Comprehension task of Experiment 1. Significance levels: **p < .01, ***p < .001.

| Effects | Estimate | SE | z |
|---|---|---|---|
| Intercept | 1.64 | 0.2615 | 6.30*** |
| System (Reportative vs. Visual/Inferential) | 1.80 | 0.47 | 3.81*** |
| System (Visual vs. Inferential) | −1.15 | 0.48 | −2.37*** |
| List (List 1 vs. Mean) | 0.14 | 0.29 | 0.50 |
| List (List 2 vs. Mean) | 0.38 | 0.29 | 1.29 |
| System (Reportative vs. Visual/Inferential): List 1 | 0.80 | 0.66 | 1.22 |
| System (Visual vs. Inferential): List 1 | 1.08 | 0.67 | 1.59 |
| System (Reportative vs. Visual/Inferential): List 2 | 0.86 | 0.67 | 1.28 |
| System (Visual vs. Inferential): List 2 | −0.37 | 0.68 | −0.54 |

coding: the first contrast compared the Reportative system against the Visual and Inferential Systems and the second contrast compared the Visual and Inferential Systems to each other (contrast 1: 0.66, −0.33, −0.33, contrast 2: 0, −0.50, 0.50). The same contrast setup was used in all subsequent analyses. For the exploratory predictor List, we set up our contrasts as deviations from the overall mean, including no specific reference group since there was no theoretical justification to assume one (see Table 2).

The inclusion of System significantly improved the model fit based on a chi- square test of the change in −2 restricted log likelihood ($\chi^2 = 18.05$, $p = .0001$) but no such improvement was found for List ($\chi^2 = 3.10$, $p = .21$). Participants performed better in the Reportative System compared to the Visual and the Inferential Systems ($M_{Rep}$, = 0.74, $M_{Vis/Inf} = 0.63$). There was also a significant difference in performance between the Visual and Inferential Systems, with higher accuracy for the Inferential System ($M_{Vis} = 0.56$, $M_{Inf} = 0.70$). The inclusion of the interaction between System and List also improved model fit ($\chi^2 = 9.79$, $p = .04$) but the source of this interaction could not be determined by follow-up comparisons (see Table 2). When we reran the analysis reordering the levels for List (3, 2, 1), results did not change but there was an interaction such that, in List 3, accuracy in the Reportative System was not better than in the Visual/Inferential Systems ($\beta = -1.6776$, SE = 0.6267, z = −2.677, $p < .01$).

Participants' answers to the post-experiment question about the meaning and use of *ga* confirmed this picture. Of the 37 participants exposed to the Visual System, only 5 correctly associated *ga* with complete visual access to the event and/or the verb *see*. Two participants incorrectly mentioned a different type of access. The majority ($n = 30$) associated *ga* with an irrelevant, i.e., non-evidential meaning (event completion, or other morphemes such as *the*, −*ed* or the plural -*s*). Of the 37 participants exposed to the Inferential System, 9 correctly mentioned that the speaker only saw the beginning and the end of the action; an additional participant mentioned doubt. The majority of the participants ($n = 27$) associated *ga* with irrelevant function morphemes (*the*, −*ed*, *it*, *them*).

Lastly, of the 37 participants exposed to the Reportative System, 17 participants correctly mentioned that the morpheme meant telling the other puppet what happened, and an additional participant linked its meaning to a secret. Four additional participants associated the morpheme with the act of one puppet talking to another (these answers were appropriate but did not unambiguously point to a mental-state meaning). The remaining 11 participants either associated *ga* with an irrelevant grammatical morpheme (*the*, *-ing*, *-ed*) or did not provide an answer. In sum, participants' overt conjectures about *ga* often revealed sensitivity to evidential interpretations (as opposed to other possible mental-state meanings such as uncertainty, or possibility, or other non-mental state interpretations). Furthermore, the distribution of correct evidential interpretations confirmed the advantage of the Reportative system. Finally, error patterns for incorrect evidential interpretations were consistent with the hypothesized aversion to mark visual access with a dedicated morpheme: the only cases where people adopted an evidential interpretation but chose the wrong source ($n = 2$) were observed for the Visual System.

### 2.3. Discussion

According to the TPH (Gentner & Bowerman, 2009), the most typologically common linguistic distinctions should be the easiest to learn. In an Artificial Language Learning experiment, we set out to explore whether this prediction is borne out within the semantic domain of evidentiality. Our results are in line with the TPH prediction: the most typologically common evidentiality system - the Reportative system - was learned more easily by our participants compared to the less common Visual system. The Inferential system was also learned better than the Visual system. This pattern was confirmed in participants' post-experiment responses: participants exposed to the Reportative System consistently associated the morpheme with the speaker's information access when explicitly reporting what the morpheme meant.

Even though the present experiment offers initial support for the TPH, it has a number of limitations. First, the use of static displays to depict events did not properly capture the way people access events in everyday life, since the sequence of pictures was composed of multiple ‘snapshots’ and always afforded a somewhat indirect path to the event. Second, and relatedly, properties of the static displays themselves might have influenced the results. For instance, instances of Reportative access were marked by a salient bubble containing a depiction of the event (see Fig. 1) and were the only scenarios that contained three (as opposed to two) characters. These stimulus features might have made Reportative access more salient and might have facilitated participants' learning of the Reportative system. Third, Experiment 1 only included a comprehension test. It is possible that production of evidentials behaves differently from comprehension (see Ünal & Papafragou, 2016, among others). In Experiment 2 we addressed these limitations.

## 3. Experiment 2

Experiment 2 had the same logic and learnability targets (i.e., evidential systems) as Experiment 1 but used videos of unfolding events in which a character gained access to events in different ways instead of static scenarios. These videos were closer to the situations in which people find out about events in the world and were designed to offer a clearer separation of various information sources. Moreover, the visual characteristics of the videos were consistent across systems (e.g., all versions of the same video contained three people). This step was meant to ensure that any advantage found for one evidential system over the other could not be attributed to properties of the stimuli. Lastly, unlike Experiment 1, Experiment 2 tested both production and comprehension of the novel evidential morphemes.

### 3.1. Methods

#### 3.1.1. Participants

We recruited 101 adult participants between the ages of 18 and 22 (73 female and 28 male). Participants were undergraduate students at the University of Delaware, enrolled in an Introductory Psychology course and earned course credit for their participation. None of them had participated in Experiment 1. All of them were native speakers of English. To the extent that participants were familiar with additional languages (mostly Spanish and French), these languages did not include grammaticalized evidentials.

#### 3.1.2. Stimuli and procedure

The experiment had a Training Phase and a Testing Phase (itself consisting of two tasks, a Production task and a Comprehension task). We filmed 69 videos in three versions each, with each version corresponding to a type of information access (Visual, Inferential, Reportative). Of these, 15 had some overlap with events from Experiment 1 and the rest were completely new. A complete list of the events in the videos for Experiment 2 can be found in Appendix B.

Each video displayed a different event with three characters. The roles of these characters were performed by the same female undergraduate research assistants and remained consistent across the videos. One of the characters performed an action using some materials and then put these materials away (the "Agent"). A second character gained access to the event in one of three ways (Visual Perception, Inference, Report), and at the end of the event, described what happened (the "Speaker"). The third character manipulated the Speaker's access to the event by either allowing her to watch the event or by blocking her visual access for the complete duration or part of the event. Moreover, the setting in which the event took place was identical for all the videos: the Agent and the Speaker were sitting at opposite sides of a table while the third character stood behind the Speaker and had full view of the table. Each video was approximately 12 s long. At the end, the Speaker turned to the camera so as to describe what happened; at that point, the video stopped and the speech bubble with the artificial language sentence appeared, and stayed visible for 8 s before the next video began. The artificial language sentences were similar to those in Experiment 1.

Consider the sample event in Fig. 3, in which the Agent copied a drawing on the whiteboard (all events in videos had a similar structure). In the Visual Access version in (A), the Speaker had continuous direct visual access to the event (we did include a brief initial phase in which the Speaker's eyes were blocked, see panel A1, for visual consistency with the other versions). In the Inferential Access version in (B), the Speaker had visual access only for the beginning and aftermath of the event (panels B1 and B4), but not the main phase (panels B2 and B3); therefore, she had to infer what happened based on visual clues. In the Reportative Access version in (C), the Speaker's visual access was blocked while the event was unfolding (panels C1-C3); later (panel C4), the Speaker heard the third character's report about what happened. At the end of the video (identical across versions, see panel 5), the Speaker's description of what happened appeared in a speech bubble.

As in Experiment 1, we designed 3 evidential systems to be acquired (Visual, Inferential, Reportative) by having the Speaker describe only one type of Access with an evidentially marked sentence and include no evidential morpheme for the other two Access types. Participants were randomly assigned to one of these three evidential Systems (33 participants were assigned to the Visual System and 34 to each of the other two). We tested participants in small groups using the same set-up and procedure as in Experiment 1.

For the Training Phase, for each evidential system, we created three basic lists. Each list had 21 videos, 7 for each Access type. Across lists, each video rotated through each Access type. For each of these basic lists, we created three randomized presentation orders, resulting in nine lists in total. Each participant was exposed to one of these nine lists. The instructions participants received in the beginning of the Training Phase were similar to those in Experiment 1:

> "You will watch a series of short clips involving three characters: in each clip, one character will perform an action on some object(s) and then put the objects away. A second character will learn about this action in different ways depending on what a third character does. After this, the second character will describe what happened. This will be depicted in a speech bubble onscreen."

Participants were then shown a still image of the three characters in the visual scene shown in the videos so as to further familiarize them with the content of the stimuli before the Training Phase would begin. In the next instructions screen, participants were told:

> "The characters will be speaking an alien language: it shares some words with English but it is different in several ways. One difference is that the language includes a special morpheme: "ga". You will have to pay attention to when "ga" appears in order to try and figure out what it means. You will go through several events to try to learn the language and figure this out."

Before the videos started, participants were also informed that a second part would follow that would test their understanding of where *ga* appeared based on its meaning.

The Testing Phase began with a transition screen ("Let's continue")

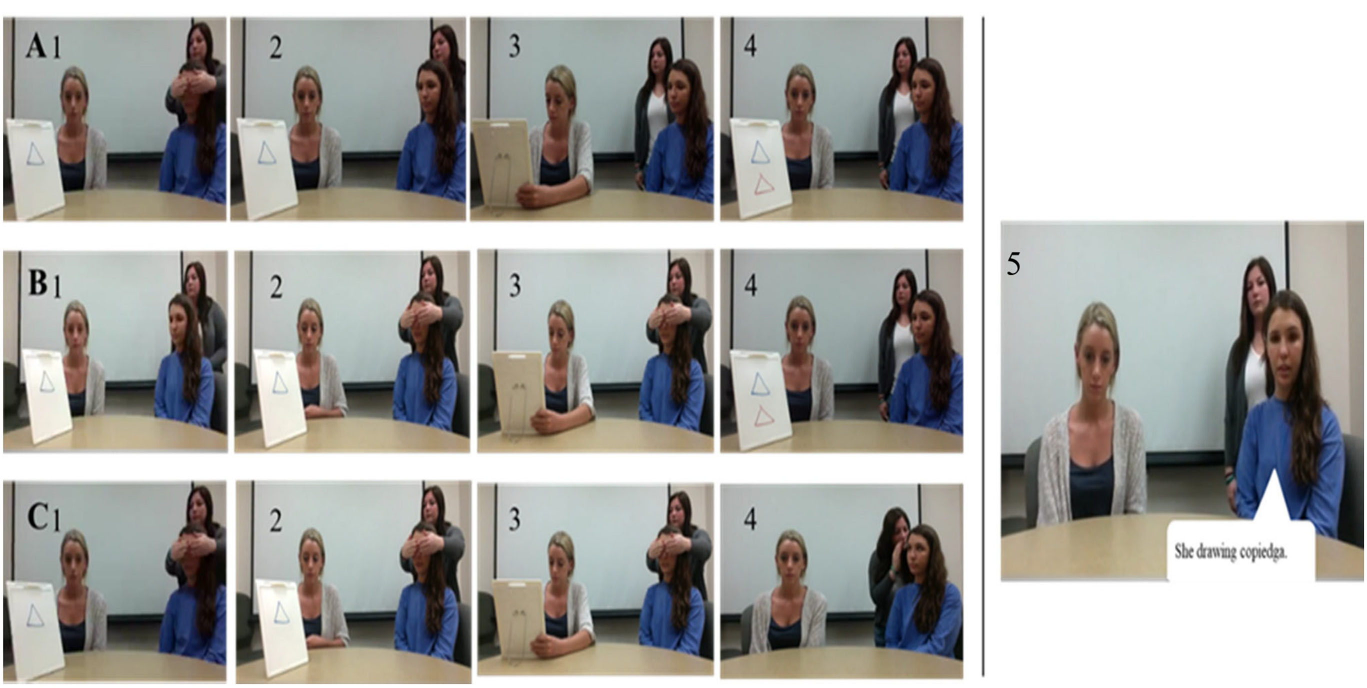


**Fig. 3.** Sample stimuli for the Artificial Language Learning task in Experiment 2. We include sample screenshots from versions of a single dynamic video in which a character gained access to an event in one of three ways: (A) Visual, (B) Inferential, (C) Reportative. The three versions of the video unfolded somewhat differently (Panels 1–4) but had the same ending (Panel 5, enlarged here for readability). At that point, the main character offered a description of the event with or without an evidential (e.g., "She drawing copiedga.").

and proceeded to the Production and later the Comprehension Task. For the Production Task, we used 12 new videos, each filmed in each of the three Access types. We created three basic lists, each containing 12 videos, 4 per Access Type. As in the Training Phase, the Access version shown for each video was rotated across these three lists. For each basic list, three randomized presentation orders were created, resulting in 9 presentation lists in total. Each participant was exposed to one of the nine lists.

The structure of the scenarios displayed was identical to those shown in the Training Phase with the difference that when the speech bubble appeared at the end, the evidential morpheme was missing. At its place, attached to the verb, there was a gap and participants had to decide whether the morpheme should be used or whether the gap should remain empty. The specific instructions participants received were as follows:

> "We will now show you a new series of clips. Here you will be asked to help describe some events using the morpheme "ga" in the alien language. Almost everything will be the same as before, but this time no morpheme is included in the speech bubble. For each clip, in the blank Response box on your Response Sheet 1, please write in the verb with the morpheme that is needed or write only the verb if no morpheme is needed in order to complete the speech bubble."

Following these instructions, participants were given an example of how their answer should be using the sentence "She piano played_": they were told that they should either write "played" or "playedga" depending on whether they thought the morpheme was needed based on what they figured out from the first part of the experiment.

For the Comprehension task we created 36 new videos, each filmed in three versions to correspond to the three Access Types. As before, three basic lists were created by rotating the Access Type shown by each video. Subsequently, three unique randomized presentation orders were created for each list (9 presenation lists in total). Each list contained 36 scenarios, 18 per each Access type. In half of these 36 scenarios, the Speaker erroneously used the morpheme *ga*: she either failed to use the morpheme when she should or used it for the incorrect type of access. In the remaining half of the scenarios, the use of the morpheme was correct. Participants' task would be to write "yes" or "no" in their response sheet depending on whether they thought the character was correctly used the morpheme or not. The instructions they received were the following:

> "We will now show you a third series of clips. Almost everything will be the same as when you paid attention to when "ga" is used, but this time some of character's descriptions will contain errors: they will omit 'ga' or include it when it is not needed. For each clip, please mark "Yes" or "No" on your Response Sheet 2 to indicate whether the character's description included or omitted "ga" correctly."

Similarly to Experiment 1, at the end of the session we asked participants to write down what they thought *ga* meant and when it was used or not used.

### 3.2. Results

Similarly to Experiment 1, participants' responses were coded for Accuracy for both Production and Comprehension. The results can be seen in Fig. 4.

For each of the tasks, we ran a mixed-effects model, using Accuracy as our dependent variable and System (Visual, Inferential, Reportative), List and their interaction as our fixed predictors. Our model also included random intercepts for Participants and Items. As in our previous analysis, we set up two planned comparisons using contrast coding such that we could compare the accuracy scores of the Reportative system to those of the other two systems, as well as the accuracy of the Visual and Inferential System (contrast 1: 0.66, −0.33, −0.33, contrast 2: 0, 0.50, 0.50, respectively; see Table 3).

For Production, our analysis revealed that including the fixed predictor System significantly improved the model fit based on a chi-square test of the change in −2 restricted log likelihood ($\chi^2 = 10.40$, $p = .005$). Participants' performance in the Reportative system was significantly better than their performance in the other two Systems ($M_{Rep} = 0.75$, $M_{Vis}/_{Inf} = 0.59$); there was no difference in performance between the Visual and Inferential systems ($M_{Vis} = 0.57$, $M_{Inf} = 0.61$). Model fit was not improved by the inclusion of either List ($\chi^2 = 3.67$, $p = .15$) or the interaction between System and List ($\chi^2 = 3.96$, $p = .41$).

For Comprehension, including System in our model significantly improved model fit ($\chi^2 = 16.16$, $p = .0003$), but no improvement of the model was observed when List ($\chi^2 = 5.33$, $p = .07$) or its interaction with System ($\chi^2 = 6.53$, $p = .16$) were included. Following up on the effect of System, we found a significant difference between Reportative and the other two Systems ($M_{Rep} = 0.77$, $M_{Vis/Inf} = 0.58$) but no significant difference between the Visual and Inferential Systems ($M_{Vis} = 0.57$, $M_{Inf} = 0.59$).

Participants' answers when asked about the meaning/use of *ga* mirror this image. Of the 33 participants exposed to the Visual System, 12 correctly mentioned that *ga* was used when the speaker saw the event. Eight participants incorrectly associated *ga* with an alternative type of access (they mentioned that the girl's eyes were covered, the girl was told what happened, or something similar). The remaining 13 participants gave incorrect non-evidential conjectures (singular/plural forms, past or completed actions, articles such as *the/a*). Of the 34 learners of the Inferential System, 9 correctly associated *ga* with the character not seeing the complete action, or inferring the action, and an additional 2 mentioned doubt. Two participants had evidential but incorrect (reportative) meanings. The remaining 21 participants associated the morpheme with irrelevant grammatical morphemes (articles, *them*, singular/plural). Lastly, of the 34 participants exposed to the Reportative System, 21 correctly associated *ga* with reportative access (importantly, these participants did not simply track the whispering or telling act from the perspective of the person committing the act but specifically alluded to the *speaker's* mental state by mentioning that she "was told" about the event). Three participants associated *ga* with the wrong source. The remaining 10 participants associated the morpheme with an irrelevant grammatical distinction. As in Experiment 1, these responses taken together confirmed that participants often chose evidential (as opposed to other possible) meanings for *ga*, did so more consistently in the Reportative System, and were more likely to reverse source meanings if asked to encode direct (visual) evidential meanings morphologically.

### 3.3. Discussion

In Experiment 2, we tested the learnability of evidential systems using videos displaying real-life, dynamic events. We also tested both receptive and productive knowledge of novel evidential morphemes. Our results reaffirm the TPH predictions: overall, the Reportative System was learned better compared to the cross-linguistically less-attested Visual and Inferential Systems. This learnability advantage is also supported by the large number of participants in the Reportative System condition that accurately associated the morpheme with the speaker's information access in a post-experiment probe.

Not all aspects of our data are compatible with the predictions of the TPH, however. Specifically, even though exclusive encoding of visual evidentials is rare, and there is a broad preference to mark non-visual/indirect over visual/direct sources cross-linguistically (Aikhenvald, 2018), the Inferential and Visual systems were equally learnable in our data. This pattern is different from Experiment 1. A possible explanation for this outcome lies with the fact that our Inference videos contained strong visual clues to what happened, bringing this type of information access closer to a direct perceptual experience than to an

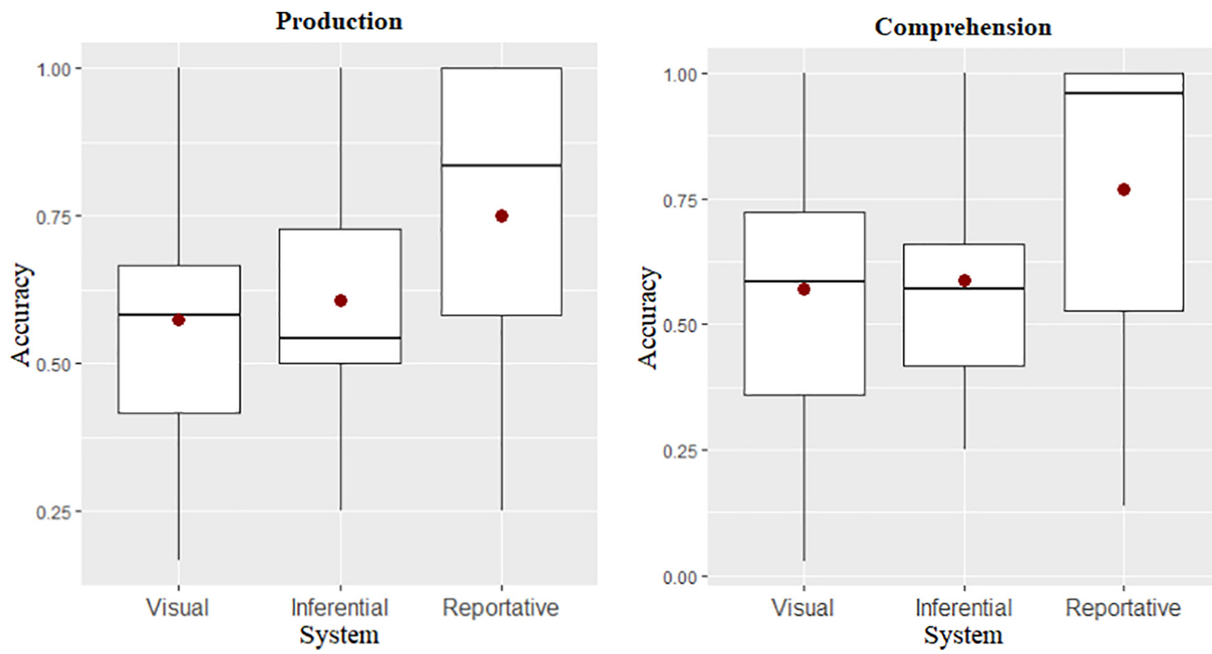


**Fig. 4.** Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in the Artificial Language Learning tasks of Experiment 2.

**Table 3**
Parameter estimates for Accuracy in the Production and Comprehension task of Experiment 2. Significance levels: $^{**}p < .01$, $^{***}p < .001$.

| Effects | Estimate | SE | z |
|---|---|---|---|
| Production task | | | |
| Intercept | 0.90 | 0.16 | 5.53*** |
| System (reportative vs. visual/inferential) | −1.10 | 0.34 | −3.28** |
| System (visual vs. inferential) | −0.16 | 0.37 | −0.44 |
| Comprehension task | | | |
| Intercept | 1.15 | 0.19 | 5.83*** |
| System (reportative vs. visual/inferential) | −1.73 | 0.42 | −4.09*** |
| System (visual vs. inferential) | −0.23 | 0.46 | −0.49 |

indirect inference on the speaker's part. This property might have blurred the distinction between Visual and Inferential access, a distinction that is subtle and somewhat language-dependent to begin with (de Haan, 2001, 2013), and might have led to a drop in performance with the Inferential system (cf. Figs. 2 and 4). By contrast, in Experiment 1, the Inferential scenarios were more distinct from the Visual ones because of the visual gap in the middle of the event (when the puppet theater curtains were drawn as the event was unfolding; see Fig. 1). If so, replacing Inferential scenarios in Experiment 2 with less direct cases of inference from visual cues (e.g., by removing the agent; see Rissman, Woodward, & Goldin-Meadow, 2019) should increase the distance between visual and inferential access and might lead to a learnability difference between the Visual and Inferential systems.

Overall, Experiments 1 and 2 taken together show that there is a learnability difference among evidential systems that mirrors cross-linguistic tendencies to mark indirect information sources (especially, reports) before direct ones. This pattern raises the question whether this asymmetry is an exclusively linguistic fact or could generalize to other kinds of communication. In the next experiment, we turn to this question.

## 4. Experiment 3

Experiment 3 adopted the structure and design of Experiment 2 but replaced the evidential morpheme with a pictorial symbol (i.e., it was an Artificial Symbol Learning task). If learnability patterns remained similar to those obtained in the Artificial Language Learning task of Experiment 2, we could conclude that the driving force for this pattern is not merely linguistic but might relate to the naturalness of forming or communicating certain source categories over others.

### 4.1. Methods

#### 4.1.1. Participants

We recruited 98 participants in total (age range: 18–22 years of age, 39 male and 59 female) from the same population as our previous experiments. None of them had participated in our earlier studies. All participants were native speakers of English. Some participants reported some knowledge of an additional language, mostly Spanish, French or Italian (none of these languages have grammaticalized evidentials).

#### 4.1.2. Stimuli and procedure

The stimuli and procedure were the same as in Experiment 2, with two modifications: (a) The speech bubble containing the witness' description of the event was in English instead of an artificial language (e.g., *She copied the drawing*); (b) the linguistic morpheme “ga” was replaced by a symbol (a black filled circle) that appeared next to the bubble and at the same time as the bubble. Both of these modifications can be seen in the example in Fig. 5. The instructions were also adjusted to reflect these changes, as described below.

For the Training Phase, similarly to Experiment 2, we created three basic lists across which the Access version of each scenario was rotated through all three Access types (Perception, Inference and Report). For each basic list, three randomized presentation orders were created, resulting in nine lists in total, each of which contained 21 videos, 7 per Access Type.

The instructions in the Training Phase were the following (changes from Experiment 2 are in bold font):

> “You will watch a series of short clips involving three characters: in each clip, one character will perform an action on some object(s) and then put the objects away. A second character will learn about this action in different ways depending on what a third character does. After this, the second character will describe what happened. This will be depicted in a speech bubble onscreen. **Some clips will**


**Fig. 5.** Example stimulus for the Artificial Symbol Learning task in Experiment 4 (Reportative System).

**display a symbol on the screen in a separate box next to the speech bubble. This symbol is related to what is happening**. You will have to pay attention to **when this symbol appears** in order to try and figure out what it means. You will go through several events to try and figure this out."

Turning to the Testing Phase, for each task (Production and Comprehension), we used the three basic lists we constructed for Experiment 2, across which the Access version of the scenarios was rotated. For each of these lists, we created three new randomized presentation orders, resulting in 9 new presentations lists for each Task.

For the Production Task, each participant saw 12 scenarios, 4 per Access type. The instructions were modified (changes from Experiment 2 are indicated in bold font), such that participants had to produce a circle when appropriate (either draw a circle or, if desired, write the word "circle"):

"We will now show you a new series of clips. Here you will be asked to fill in when appropriate **the same symbol** you have already seen. Almost everything will be the same as before BUT this time **no symbol is included** next to the speech bubble. For each clip, in the blank Response box on your Response Sheet 1, **please write in any symbol that is needed to complete the screen** or **cross out the blank** if no symbol is needed."

For the Comprehension Task, each participant saw 36 scenarios. In half of those scenarios, the symbol marked the correct Access type and in the remaining half, the symbol either appeared with the wrong Access type or was missing from the correct Access type. Participants had to write in the response form 'Yes' or 'No', depending on whether the symbol was included or omitted correctly. The instructions that participants received were as follows (once again modifications from Experiment 2 are in bold font):

"We will now show you a third series of clips. Almost everything will be the same as when you paid attention to when **the symbol** [circle inserted here] appeared. BUT **some of the screens will contain errors:** they will omit it or include it when it is not needed. For each clip, please mark "Yes" or "No" on your Response Sheet 2 to indicate whether **the screen included or omitted the symbol** correctly."

As in Experiments 1 and 2, at the end of the session we asked participants to write down what they thought the symbol meant and when it was or was not used.

### *4.2. Results*

Participants' responses were again coded for accuracy. The results can be seen in Fig. 6. As for previous experiments, a mixed-effects logistic regression model was used, with Accuracy as the dependent, binomial variable, and System and List as the fixed predictors along with their interaction. The model also included random intercepts for Participants and Items (see Table 4).

Starting with the Production task, our analysis showed that only the fixed predictor System significantly improved the model fit based on a chi- square test of the change in −2 restricted log likelihood ($\chi^2 = 14.99$, $p = .0005$). Neither List ($\chi^2 = 3.92$, $p = .14$) nor the interaction between System and List improved model fit ($\chi^2 = 5.34$, $p = .25$). Following up on the role of System, there was a significant difference in performance between the Reportative and the other two Systems ($M_{Rep} = 0.93$, $M_{Vis/Inf} = 0.77$) but there was no difference between the Visual and Inferential Systems ($M_{Vis} = 0.79$, $M_{Inf} = 0.76$).

For the Comprehension task, our model similarly showed that only System ($\chi^2 = 9.78$, $p = .007$) but not List ($\chi^2 = 3.02$, $p = .22$) or the interaction between System and List ($\chi^2 = 3.35$, $p = .50$) improved model fit. Looking into the role of System, we found a significant difference between accuracy for Reportative when compared to the Visual and Inferential Systems ($M_{Rep} = 0.86$, $M_{Vis/Inf} = 0.71$) and no difference in accuracy between the Visual and Inferential System ($M_{Vis} = 0.71$, $M_{Inf} = 0.72$).

The debriefing data confirmed this picture. Of the 33 participants exposed to the Visual System, 17 gave correct conjectures (15 reported that the character saw the complete action and 2 said that character did not have her eyes covered). Three participants erroneously associated the symbol with reportative access. The remaining 11 participants gave incorrect answers (change in appearance of the objects, kind of motion). Of the 33 participants learning the Inferential System, 13 gave correct answers (11 mentioned that the character could only see the beginning and the end of the action and 2 that the character was blindfolded). Seven participants reported an incorrect evidential meaning (either saying that the symbol was used when the character could see "most or all the action" or associating the symbol with a report). The remaining 13 participants offered either incorrect answers (related with the type of action) or no answers. Finally, of the 32 participants exposed to the Reportative System, 21 correctly associated the symbol with the speaker getting informed about what happened by another person (all of them referred to the speaker's mental state and not only to the telling/whispering act). One additional participant associated the symbol with the character's lack of visual access. The remaining 10 participants gave irrelevant or no answers.

Lastly, we compared the results obtained from Experiments 2 and 3 that tested the learnability of evidential language and symbols respectively. Using a mixed-effect logistic regression model, as for the previous analyses, we included Accuracy as our dependent variable and System (Visual, Inferential, Reportative) and Experiment (Exp.2 - Language, Exp.3 - Symbol) as our fixed predictors. For this analysis, we excluded the predictor List since it did not significantly improve the model fit in any of the two Experiments. We performed a separate analysis for each task across the two experiments (see Table 5).

For the Production task, both System ($\chi^2 = 21.37$, $p = .00002$) and Experiment ($\chi^2 = 27.93$, $p = .0000001$) but not their interaction ($\chi^2 = 2.07$, $p = .35$) significantly improved the model fit based on a chi- square test of the change in −2 restricted log likelihood. There was

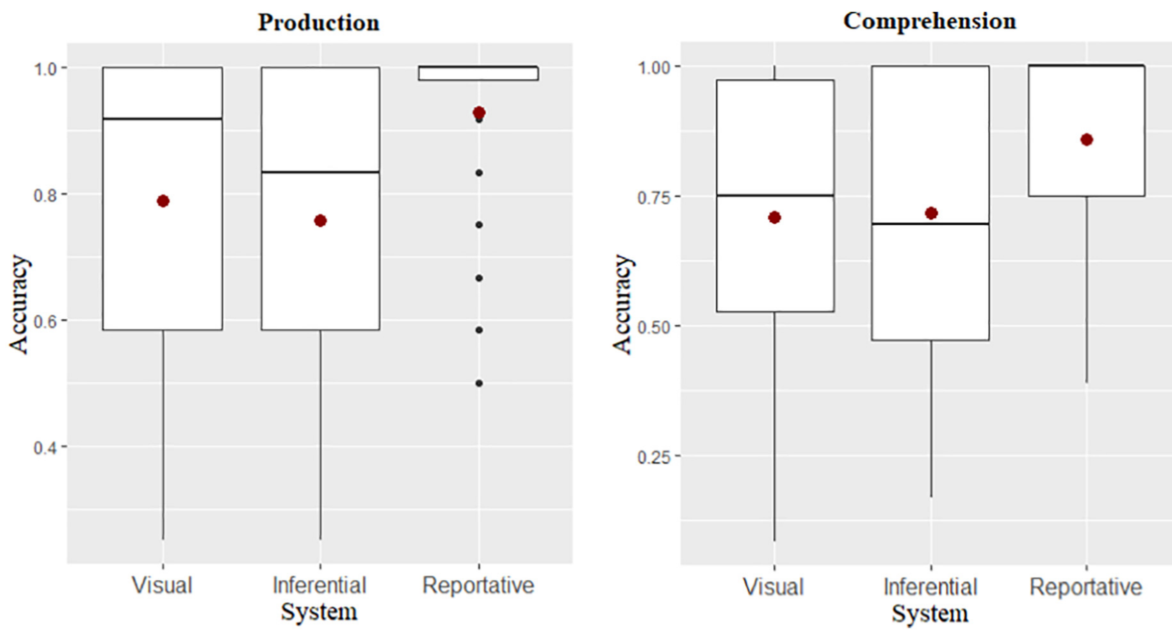


**Fig. 6.** Accuracy score distribution, median (horizontal bar) and mean (dot) for each Evidential System in the Artificial Symbol Learning tasks of Experiment 3.

**Table 4**
Parameter estimates for Accuracy in the Production and Comprehension task of Experiment 3. Significance levels: $^{**}p < .01$, $^{***}p < .001$.

| Effects | Estimate | SE | z |
|---|---|---|---|
| Production task | | | |
| Intercept | 2.94 | 0.35 | 8.25*** |
| System (reportative vs. visual/inferential) | −2.33 | 0.63 | −3.66*** |
| System (visual vs. inferential) | −0.39 | 0.61 | 0.63 |
| Comprehension task | | | |
| Intercept | 5.57 | 0.31 | 8.15*** |
| System (reportative vs. visual/inferential) | −2.00 | 0.66 | −3.03*** |
| System (visual vs. inferential) | −0.17 | 0.66 | −0.26 |

**Table 5**
Parameter estimates for Accuracy in the Production and Comprehension tasks of Experiments 2 and 3. Significance levels: $^{**}p < .01$, $^{***}p < .001$.

| Effects | Estimate | SE | z |
|---|---|---|---|
| Production task | | | |
| Intercept | 1.00 | 0.19 | 5.20*** |
| System (reportative vs. visual/inferential) | −1.52 | 0.31 | −4.82*** |
| System (visual vs. inferential) | −0.06 | 0.32 | 0.19 |
| Experiment (Exp.2 vs. 3) | 1.49 | 0.28 | 5.26*** |
| Comprehension task | | | |
| Intercept | 1.22 | 0.22 | 5.35*** |
| System (reportative vs. visual/inferential) | −1.81 | 0.37 | −4.91*** |
| System (visual vs. inferential) | −0.19 | 0.39 | −0.51 |
| Experiment (Exp.2 vs. 3) | 1.13 | 0.33 | 3.43*** |

a significant difference in performance between the Reportative System and the other two Systems ($M_{Rep} = 0.83$, $M_{Vis/Inf} = 0.68$) but no difference between the Visual and Inferential System ($M_{Vis} = 0.68$, $M_{Inf} = 0.68$). Participants were overall more accurate in Experiment 3 that involved Symbol learning compared to Experiment 2 that involved Language learning ($M_{Exp1} = 0.64$, $M_{Exp2} = 0.82$).

For the Comprehension task, we obtained similar results: System ($\chi^2 = 22.95$, $p = .00001$) and Experiment ($\chi^2 = 11.75$, $p = .0006$) significantly improved the model fit but their interaction did not ($\chi^2 = 0.02$, $p = .98$). Follow-up analyses showed that participants performed better in the Reportative System compared to the other two Systems ($M_{Rep} = 0.81$, $M_{Vis/Inf} = 0.64$) but no difference was found between the Visual and Inferential Systems ($M_{Vis} = 0.64$, $M_{Inf} = 0.65$). Participants were more accurate in Experiment 3 compared to Experiment 2 ($M_{Exp2} = 0.64$, $M_{Exp3} = 0.76$).

### *4.3. Discussion*

Performance on the Artificial Symbol Learning task of Experiment 3 was better compared to the corresponding Language Learning task of Experiment 2, presumably because of the costs of acquiring an unfamiliar language. Despite this difference, Experiment 3 replicated the basic finding from Experiment 2: the evidential system that marked only another speaker's report as source of information (Reportative) was learned better compared to both the evidential system that only marked inference (Inferential) and the system that only marked direct, perceptual experience (Visual). No significant difference was observed between the Inferential and the Visual systems. We also see the same pattern in participants' reports about what the symbol meant: the vast majority of those exposed to the Reportative System associated the symbol with the speaker's epistemic state rather than the mere act of whispering or telling. We conclude that the bias to mark certain sources of information over others applies beyond language to other forms of ostensive communication. We discuss the nature of this bias and its effects on both typology and learnability in the following section.

## 5. General discussion

### *5.1. Evidentiality and the learnability of semantic distinctions*

A widespread assumption within cognitive science is that highly frequent cross-linguistic patterns should be easier to learn than less frequent ones because they reflect biases in human cognitive mechanisms. Within the domain of semantics, this idea has been captured by Gentner and Bowerman's (2009) TPH but has not received much empirical scrutiny. To our knowledge, our study was among the first (and very few) research studies that used an Artificial Language Learning methodology to explore learnability within the semantic domain (cf. also Xu et al., 2013; Carstensen et al., 2015; Maldonado & Culbertson, 2019). We compared the ease of acquisition of three evidential systems marking a single information source (Visual, Inferential, Reportative) by adult native speakers of English whose language lacks evidential morphology. Based on the relative frequency of these systems cross-linguistically, the TPH predicts that the Reportative should be the most learnable and the Visual system the least learnable of the three systems.

Overall, this prediction was partially supported. Across two different experiments with different stimuli (Experiments 1 and 2), the Reportative system was learned more easily compared to the Visual System. The learnability profile for the Inferential system was less consistent: in Experiment 1, the Inferential system was learned more easily than the Visual System but in Experiment 2, the Visual and the Inferential systems produced comparable learnability outcomes (and were both harder to learn than the Reportative system).

Together our data offer evidence for the conclusion that highly frequent semantic distinctions are more learnable than less frequent ones, as predicted by the TPH (Gentner & Bowerman, 2009). Furthermore, they add to previous studies that have studied learnability with the same methodological paradigm as ours mostly within the domains of syntax, phonology and morphology (Christiansen, 2000; Newport & Aslin, 2004; Hudson Kam & Newport, 2005, 2009; Seidl & Buckley, 2005; Wilson, 2006; Thompson & Newport, 2007; Wonnacott et al., 2008; Finley & Badecker, 2009; Tily et al., 2011; Merkx et al., 2011; Culbertson, 2012; Culbertson & Smolensky, 2012; Culbertson et al., 2012; Fedzechkina et al., 2012; Tabullo et al., 2012).

### *5.2. Learnability and 'natural categories'*

A striking aspect of our data was the finding that the preference to mark reportatives persisted when the participants' task was to acquire the evidential meaning of a pictorial symbol and not a linguistic morpheme (Experiment 3). This finding coheres with evidence from other non-linguistic communicative systems: many European and American communities that we are familiar with use a two-finger co-speech gesture for reportative attributions (e.g., *He wrote* "*the best essay in the whole school*") but have no such gesture for visual or inferential attributions. Together, these results support the position that both the typological and the learnability patterns that characterize evidentiality are motivated by factors that go beyond language to characterize non-linguistic systems.

What could such factors be? Recall that, on Gentner and Bowerman's (2009) original proposal, the roots of the TPH lie in the cognitive naturalness of the semantic classes that the learner acquires: "All else being equal, within a given domain, the more frequently a given way of categorizing is found in the languages of the world, the more natural it is for human cognizers, hence the easier it will be for children to learn" (p. 467). One possibility, then, is that the present advantage of the Reportative system is the result of the naturalness of the underlying concepts. Upon closer inspection of our data, this possibility seems surprising, even paradoxical. Within the class of evidential distinctions, it is visual perception – not reports - that could plausibly be considered as a primary, highly salient and valued source of information for humans and hence a primary candidate for both cross-linguistically consistent encoding and easy acquisition. For instance, perception as a knowledge source is represented in areas related to mental-state reasoning in both seeing and blind individuals (Koster-Hale, Bedny, & Saxe, 2014). Some understanding that seeing leads to knowing emerges early in children (Pillow, 1989; Pratt & Bryant, 1990; Ozturk & Papafragou, 2016) and is even found in other primates (Hare, Call, & Tomasello, 2001). In memory, a common mistake is to think one has seen things with one's own eyes that were only indirectly experienced (e.g., imagined, inferred or read about; Johnson, Hashtroudi, & Lindsay, 1993; Hannigan & Reinitz, 2001; Ünal, Pinto, Bunger, & Papafragou, 2016); the opposite error rarely occurs. Even within the domain of language, aside from evidential systems, the domain of vision is richly represented cross-linguistically with a variety of verbs (e.g., *see*, *perceive*, *look*; Viberg, 1984; San Roque et al., 2015; but see Majid et al., 2018), and the broad information-access meanings of these verbs seem to be accessible even in congenitally blind individuals (Landau & Gleitman, 1985). In sum, contrary to our findings, visual perception as an information source appears to be a 'natural' candidate for languages to mark and for learners to encode.

In what follows, we propose two distinct but not mutually exclusive possibilities to reconcile the prevalence of visual perception as a source of information for humans with the fact that such a prevalent source is neither easily acquired by learners acquiring a single evidential morpheme (or symbol) in our experiments nor preferentially marked in grammatical evidential systems cross-linguistically. Each of the two possibilities looks beyond the notion of naturalness as originally proposed in the context of the TPH. Together these two possibilities represent different, and probably complementary, principles that affect the nature and learnability of meaning distinctions within and beyond language.

### *5.3. Learnability and categories in the meaning hypothesis space*

A first possible explanation for the learnability asymmetries we observed in both linguistic and non-linguistic (symbol) tasks attributes such asymmetries to the relative ease of splitting events into the right evidential categories (the 'category partition' explanation): events that were marked by the Reportative system were more clearly distinct as a class compared to events that were marked by the Visual and Inferential systems. This is because, as we have already mentioned in the discussion of Experiment 2, some instances of inference from perceptibles can be hard to distinguish from instances of perception proper. As a result, when learning the Visual or Inferential system, participants were required to learn to mark half of a set of events that were often indistinguishable from each other, which was likely to be difficult; when learning the Reportative system, all and only the events within a distinct class were marked, so learning to mark that category of events was easier.

The category partition explanation is in line with the observation in the linguistics literature that inferential evidentials have a mixed status, sometimes behaving like visuals and at other times behaving like reportatives (de Haan, 2001). It is also consistent with several findings from a recent study by Ünal et al. (2016) suggesting that inference from visual experience is not a homogeneous category in cognition (see also Ünal & Papafragou, 2018). In one experiment of that study, English speakers were exposed to an event and had to describe it and state how they had found out about it. When participants had experienced an event directly (e.g., a woman drink milk), they stated having seen the event. However, when they had only seen visual cues from the aftermath of an event and had to "fill in" the rest, their statements varied. Closer inspection suggested that, when the visual cues were indeterminate (e.g., a woman next to a birthday cake, leading to the inference that she blew out the candles), participants consistently stated that they had inferred the event. By contrast, when the visual cues were more determinate and highly constrained the inference (e.g., a woman next to eggshells, leading to the inference that she cracked the egg), participants were equally likely to say that they had seen or inferred the event. The authors proposed that there are several varieties of inference, and that stronger, more constrained (and thus more secure) inferences from visual cues might be difficult to distinguish from purely perceptual experience. In another experiment, Ünal et al. (2016) found that these varieties of inference had implications for the use of evidential morphology in a language that grammaticalizes evidentiality: speakers of Turkish used the indirect evidential in their language more consistently for the very same subset of events that English speakers had judged to be unequivocally inferred compared to those events that elicited more mixed source judgments. Furthermore, inference types had effects on memory: building on classic studies showing that people often have a false memory of having actually experienced events that they have only inferred (Johnson et al., 1993; Hannigan & Reinitz, 2001; cf. Strickland & Keil, 2011), Ünal et al. (2016) found that, across

English and Turkish speakers, such misattributions to perception were more common when inferences were strongly constrained by visual cues and thus harder to distinguish from pure perception.

The category partition explanation captures the typological prevalence of reportative systems, since such systems mark a category of source meanings that is easily separated off from the others (Cornillie, Arrese, & Wiemer, 2015; Plungian, 2001). Furthermore, this explanation is reminiscent of the idea that observed patterns of semantic typology can be understood in terms of connectedness of categories in a discrete semantic map (Haspelmath, 1997; Jurafsky, 1996; van der Auwera & Plungian, 1998), or the relative positions of specific meaning distinctions within a continuous semantic space (Gärdenfors, 2014; Levinson & Meira, 2003; Majid, Boster, & Bowerman, 2008).

### 5.4. Learnability and pragmatic marking

A second possibility is that the learnability biases (and the broad typological facts) in the semantic domain of evidentiality are related to the pragmatic implications carried by different sources of information, specifically the fact that direct or visual information sources are generally (even though not always) more reliable compared to non-visual or less direct sources (Aikhenvald, 2018; Dancy, 1985; Matsui & Fitneva, 2009; Papafragou, Cassidy, & Gleitman, 2007; Wiemer, 2018), with reports being the least direct sources of all. We know that even young children track a speaker's trustworthiness and choose to learn things from reliable over unreliable speakers (Fusaro, Corriveau, & Harris, 2011; Harris, 2012; Jaswal, 2010; Jaswal & Neely, 2006; Koenig, 2012; Koenig & Harris, 2005; Mascaro & Sperber, 2009; Sabbagh & Baldwin, 2001). Both typology and learnability then tend to prioritize marking indirect, potentially less reliable information sources (especially verbal reports). This perspective coheres with proposals according to which human cognition is equipped with epistemic vigilance so to avoid unreliable sources and the risk of being misinformed (Sperber et al., 2010).

According to this 'pragmatic bias' hypothesis, non-perceptual sources in our data are selectively marked in systems that have a single evidential morpheme or symbol because they are *informative*, i.e., they represent a departure from the primacy of perception as an information source (Bannard, Rosner, & Matthews, 2017; Grice, 1989). Furthermore, given that reportative access is the least direct type of information source (since the speaker need have experienced no part of an event), it is the most likely to be encoded in a system that has a single evidential marker. This perspective makes more substantive the conjecture that "the tendency to mark direct, or visual, or sensory evidentials less than others may reflect the primacy of vision as an information source" (Aikhenvald, 2018, p.16). From a theoretical viewpoint, this proposal is consistent with evidence that, early on in life, humans selectively comment on whatever they find unexpected (Greenfield, 1979; Greenfield & Smith, 1976), and even young children calculate informativeness as they learn (Frank & Goodman, 2014) and produce (Bannard et al., 2017) language. From an empirical viewpoint, this proposal is reminiscent of work showing that both young children and adults are more likely to interpret novel verbs in a story as mental verbs such as *think* when the verbs are used to describe *false* as opposed to true beliefs (Papafragou, Cassidy, & Gleitman, 2007) – presumably because such beliefs have an unusual, and thus pragmatically notable, property (cf. also Hacquard & Lidz, 2019; Pyers & Senghas, 2009).

The pragmatic bias hypothesis predicts that visual evidentials should be learnable as long as they are part of an evidential system that also encodes some form of non-visual or indirect evidential (such two-way systems are encountered frequently in the world's languages; Aikhenvald, 2018). Existing developmental evidence supports this prediction (see de Villiers et al., 2009; Ünal & Papafragou, 2016, 2019). On this approach, then, what learners can and cannot learn in our data and beyond the lab is determined not only by what is easy or hard to conceptualize or categorize but also by what sorts of meaning distinctions are more or less informative and thus likely to be chosen for encoding by a communicator during an interactive exchange. This approach is compatible with other recent perspectives that have linked the organization of the lexicon cross-linguistically to communicative pressures related to efficiency (Gibson et al., 2017; Kemp, Xu, & Regier, 2018; cf. also Haspelmath, 1997; Fedzechkina et al., 2012), and connects with pragmatically-inspired work in both adult psycholinguistics (Sperber, Cara, & Girotto, 1995; Van der Henst, Bujakowska, Ciceron, & Noveck, 2006) and semantic development (Clark, 1973; Grigoroglou, Johanson, & Papafragou, 2019).

### 5.5. Towards a theoretical synthesis

Each of the two hypotheses presented above goes beyond a simple mapping between 'natural' concepts and learnability outcomes as originally proposed by the TPH and instead attempts to explain learnability patterns in the domain of evidentiality in terms of which divisions of the meaning space are internally coherent to entertain (category partitions hypothesis) or informationally appropriate to convey (pragmatic bias hypothesis). Each of these hypotheses can explain cross-linguistically robust preferences to single out and encode specific classes of meaning over others, and can accommodate the fact that such preferences can also extend to meanings communicated by non-linguistic symbols. Furthermore, the two hypotheses are not mutually exclusive: for instance, even if the advantage of the Reportative system is explained pragmatically, one would need to invoke the category partitions hypothesis to explain why the Visual and Inferential systems behave similarly across our studies.

At present, our data leave open the precise contributions of these two hypotheses for the learnability of evidential distinctions but future experiments can throw light on their relative roles in a systematic way. For instance, the semantic partitions hypothesis predicts that learners should be equally successful with the current Reportative system and a mirror image of the system that uses a single morpheme to encode Inferential and Visual events and leaves Reportative events unmarked: the idea is that such a novel system would maintain the conceptual coherence of the classes of meaning to be learned. By contrast, the pragmatic bias hypothesis maintains that learning should asymmetrically favor the current Reportative system over its mirror image. We are currently pursuing this and related predictions in ongoing work.

### 5.6. Final thoughts

Taken together, our findings lend support to the hypothesis that the learnability of linguistic distinctions is related to their typological frequency across languages: when adults were given equal opportunities to acquire one of three miniature evidential systems, the most frequent grammatical system yielded the highest learnability rates compared to less frequent systems. We have proposed that these facts reflect either major, internally coherent categories of meaning distinctions for human cognizers, or pragmatic pressures on which types of linguistic meanings are informative to encode (or possibly a combination of these two factors). Depending on the relative contribution of such competing factors, frequently expressed and easy-to-acquire linguistic meanings can be a window onto conceptual categories but can alternatively or additionally reveal the role of pragmatic mechanisms within cognitive architecture.

## CRediT authorship contribution statement

**Dionysia Saratsli:** Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing - original draft, Visualization. **Stefan Bartell:** Conceptualization, Methodology, Investigation. **Anna Papafragou:** Conceptualization, Methodology, Validation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

## Acknowledgements

This work was supported by a Leventis Foundation Graduate Educational Grant (first author) and NSF Grant #1632849 (last author). We would like to thank the undergraduate research assistants Casey Corallo and Juliette Lassig for their help in collecting and coding data for this project. The work was completed while the last author was at the University of Delaware.

## Appendix A. List of events used in Experiment 1

| Events | | | |
|---|---|---|---|
| 1 | The frog erased the board. | 21 | The gorilla mixed the salad. |
| 2 | The gorilla shelled the peanuts. | 22 | The giraffe grated the cheese. |
| 3 | The squirrel transferred the strawberries. | 23 | The mouse chopped the celery. |
| 4 | The gorilla prepared the s'more. | 24 | The giraffe labeled the cards. |
| 5 | The squirrel opened the drawer. | 25 | The frog drew the square. |
| 6 | The gorilla colored the star. | 26 | The squirrel gathered the flowers. |
| 7 | The frog decorated the cupcake. | 27 | The giraffe plucked the petals. |
| 8 | The frog folded the tortillas. | 28 | The mouse removed the coughdrops. |
| 9 | The frog sliced the pepper. | 29 | The mouse crumpled the papers. |
| 10 | The squirrel emptied the pouch. | 30 | The doggy replaced the eggs. |
| 11 | The squirrel sorted the crayons. | 31 | The giraffe aligned the figurines. |
| 12 | The gorilla drank the juice. | 32 | The doggy split the crackers. |
| 13 | The gorilla inflated the balloon. | 33 | The doggy peeled the banana. |
| 14 | The frog stacked the cups. | 34 | The doggy hammered the nails. |
| 15 | The frog picked the grapes. | 35 | The gorilla skinned the potato. |
| 16 | The gorilla husked the corn. | 36 | The squirrel connected the dots. |
| 17 | The squirrel poured the cereal. | 37 | The frog ate the apple. |
| 18 | The frog completed the crossword. | 38 | The gorilla strung the beads. |
| 19 | The squirrel built the steps. | 39 | The mouse solved the puzzle. |
| 20 | The squirrel sharpened the pencils. | | |

## Appendix B. List of events used in Experiments 2 and 3

| Events | | | | | |
|---|---|---|---|---|---|
| 1 | The girl divided the rings. | 24 | The girl dissolved the tablet. | 47 | The girl clipped the sheets. |
| 2 | The girl squeezed the toothpaste. | 25 | The girl piled the cards. | 48 | The girl folded the towel. |
| 3 | The girl unzipped the jacket. | 26 | The girl circled the word. | 49 | The girl colored the star. |
| 4 | The girl filled the bowl. | 27 | The girl packaged the foam. | 50 | The girl cracked the peanut. |
| 5 | The girl dumped the trash. | 28 | The girl closed the book. | 51 | The girl strung the beads. |
| 6 | The girl solved the puzzle. | 29 | The girl layered the stickers. | 52 | The girl shoveled the dirt. |
| 7 | The girl undressed the doll. | 30 | The girl pasted the photos. | 53 | The girl separated the plates. |
| 8 | The girl sorted the morphemes. | 31 | The girl unfurled the umbrella. | 54 | The girl aligned the animals. |
| 9 | The girl bit the apple. | 32 | The girl flattened the putty. | 55 | The girl drew the square. |
| 10 | The girl flipped the pans. | 33 | The girl suspended the curtains. | 56 | The girl deflated the balloon. |
| 11 | The girl laid the bricks. | 34 | The girl built the tower. | 57 | The girl popped the coughdrops. |
| 12 | The girl fanned the spoons. | 35 | The girl erased the board. | 58 | The girl stapled the packet. |
| 13 | The girl untied the shoe. | 36 | The girl copied the drawing. | 59 | The girl connected the dots. |
| 14 | The girl wound the yarn. | 37 | The girl formed the snowman. | 60 | The girl emptied the pouch. |
| 15 | The girl rotated the vase. | 38 | The girl unscrewed the lightbulb. | 61 | The girl arranged the blocks. |
| 16 | The girl decorated the cookie. | 39 | The girl unrolled the shirt. | 62 | The girl removed the letter. |
| 17 | The girl chopped the celery. | 40 | The girl poured the juice. | 63 | The girl braided the string. |
| 18 | The girl lit the lamp. | 41 | The girl unfolded the napkin. | 64 | The girl gathered the flowers. |
| 19 | The girl hung the hanger. | 42 | The girl transferred the candies. | 65 | The girl stacked the cups. |
| 20 | The girl pierced the page. | 43 | The girl cleared the table. | 66 | The girl sliced the bread. |
| 21 | The girl mixed the gumballs. | 44 | The girl peeled the banana. | 67 | The girl unpacked the bookbag. |
| 22 | The girl sealed the envelope. | 45 | The girl cut the triangle. | 68 | The girl unbuttoned the sweater. |
| 23 | The girl crumpled the papers. | 46 | The girl matched the socks. | 69 | The girl split the cracker. |

## References

Aikhenvald, A. (2004). *Evidentiality.* Oxford University Press.
Aikhenvald, A. (2014). The grammar of knowledge: A cross-linguistic view of evidentials, and the expression of information source. In A. Aikhenvald, & R. M. W. Dixon (Eds.). *The grammar of knowledge: A cross-linguistic typology* (pp. 1–51). Oxford: Oxford University Press.
Aikhenvald, A. (2018). Evidentiality: The framework. In A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality* (pp. 1–43). Oxford: Oxford University Press.
Aikhenvald, A., & Dixon, R. (Eds.). (2003). *Studies in evidentiality*. Amsterdam: Benjamin.
Aikhenvald, A., & Dixon, R. (Eds.). (2014). *The grammar of knowledge: A cross-linguistic typology*. Oxford: Oxford University Press.
Aksu-Koç, A. (1988). *The acquisition of aspect and modality: The case of past reference in Turkish.* Cambridge, UK: Cambridge University Press.
Aksu-Koç, A., & Alici, D. (2000). Understanding sources of beliefs and marking of uncertainty. In E. Clark (Ed.). *Proceedings of the 30th annual child language research forum* (pp. 123–130). Stanford, CA: CSLI.
Aksu-Koç, A., Ögel-Balaban, H., & Alp, E. I. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development, 125*, 13–28. https://doi.org/10.1002/cd.247.
AnderBois, S. (2014). On the exceptional status of reportative evidentials. *Proceedings of SALT, 24*, 234–254.
Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* New York, NY: Cambridge University Presshttps://doi.org/10.1017/CBO9780511801686.
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.
Bannard, C., Rosner, M., & Matthews, D. (2017). What's worth talking about? Information theory reveals how children balance informativeness and ease of production. *Psychological Science, 28*(7), 954–966. https://doi.org/10.1177/0956797617699848.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.
Bowerman, M. (1993). Typological perspectives on language acquisition: Do cross-linguistic patterns predict development? In E. V. Clark (Ed.). *Proceedings of the 25th annual child language research forum* (pp. 7–15). Stanford: Center for the Study of Language and Information.
Brosseau-Liard, P. E., & Birch, S. A. J. (2011). Epistemic states and traits: Preschoolers appreciate the differential informativeness of situation-specific and person-specific cues to knowledge. *Child Development, 82*, 1788–1796.
Carstensen, A., Xu, J., Smith, C., & Regier, T. (2015). Language evolution in the lab tends toward informative communication. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin: TX. Cognitive Science Society.
Chafe, W., & Nichols, J. (Eds.). (1986). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.
Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In J. L. Dessalles, & L. Ghadakpour (Eds.). *The evolution of language: 3rd international conference* (pp. 45–48). Paris: France.
Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective.* New York: Oxford University Press.
Clark, H. (1973). Space, time semantics, and the child. In T. E. Moore (Ed.). *Cognitive development and the acquisition of language* (pp. 27–63). New York: Academic Press.
Clark, E. (1976). Universal categories: On the semantics of classifiers and children's early word meanings. In A. Juilland (Vol. Ed.), *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*. *Vol. 3*. *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday* (pp. 449–462). Saratoga, CA: Anna Libri.
Cornillie, B., Arrese, M., J. & Wiemer, B. (2015). Evidentiality and the semantics-pragmatics interface. *Belgian Journal of Linguistics*, *29*, 1–17. doi 10.1075/bjl.29.001int
Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Linguistics and Language Compass, 6*(5), 310–329. https://doi.org/10.1002/lnc3.338.
Culbertson, J., & Smolensky, P. (2012). A Bayesian model of biases in artificial language learning: The case of a word-order universal. *Cognitive Science, 36*(8), 1468–1498. https://doi.org/10.1111/j.1551-6709.2012.01264.x.
Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition, 122*(3), 306–329. https://doi.org/10.1016/j.cognition.2011.10.017.
Dancy, J. (1985). *An introduction to contemporary epistemology.* Oxford: Blackwell.
Delancey, S. (2002). The mirative and evidentiality. *Journal of Pragmatics, 33*, 369–382.
Faller, M. (2012). Evidential scalar implicatures. *Linguistics and Philosophy, 35*, 285–312. https://doi.org/10.1007/s10988-012-9119-8.
Faller, M. (2014). Reportative evidentials and modal subordination. *Lingua, 186*, 55–67. https://doi.org/10.1016/j.lingua.2014.10.004.
Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences, 109*(44), 17897–17902. https://doi.org/10.1073/pnas.1215776109.
Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2016). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science, 41*(2), 416–446. https://doi.org/10.1111/cogs.12346.
Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language, 61*(3), 423–437. https://doi.org/10.1016/j.jml.2009.05.002.
Fitneva, S. A. (2008). The role of evidentiality in Bulgarian children's reliability judgments. *Journal of Child Language, 35*, 845–868. https://doi.org/10.1017/S0305000908008799.
Fitneva, S. (2018). The acquisition of evidentiality. In A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality* (pp. 185–201). Oxford, UK: Oxford University Press.
Fitneva, S. A., Lam, N. H. L., & Dunfield, K. A. (2013). The development of children's information gathering: To look or to ask? *Developmental Psychology, 49*, 533–542.
Folia, V., Uddén, J., de Vries, M. H., Forkstam, C., & Petersson, K. M. (2010). Artificial language learning in adults and children. Language Learning, 60(s2), 188–220.
Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology, 75*, 80–96. https://doi.org/10.1016/j.cogpsych.2014.08.002.
Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology, 110*(4), 561–574. https://doi.org/10.1016/j.jecp.2011.06.008.
Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces.* Cambridge, MA: MIT Press.
Garrett, E. (2000). *Evidentiality and assertion in Tibetan.* (UCLA doctoral diss).
Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Özçalıskan (Eds.). *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 465–480). Hillsdale, NJ: Lawrence Erlbaum Associates.
Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences, 114*(40), 10785–10790.
Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135–176.
Greenfield, P. M. (1979). Informativeness, presupposition, and semantic choice in single-word utterances. In E. Ochs, & B. Schieffelin (Eds.). *Developmental pragmatics* (pp. 159–166). New York, NY: Academic Press.
Greenfield, P. M., & Smith, J. H. (1976). *The structure of communication in early language development.* New York, NY: Academic Press.
Grice, P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.
Grigoroglou, M., Johanson, M., & Papafragou, A. (2019). Pragmatics and spatial language: The acquisition of *front* and *back*. *Developmental Psychology, 55*, 729–744.
de Haan, F. (2001). The place of inference within the evidential system. *International Journal of American Linguistics, 67*, 193–219.
de Haan, F. (2013). Coding of evidentiality. In M. Dryer, & M. Haspelmath (Eds.). *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Hacquard, V., & Lidz, J. (2019). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind and Language, 34*(1), 73–96. https://doi.org/10.1111/mila.12192.
Hannigan, S. L., & Reinitz, M. T. (2001). A demonstration and comparison of two types of inference-based memory errors. *Journal of Experimental Psychology: Learning Memory and Cognition, 27*(4), 931–940. https://doi.org/10.1037/0278-7393.27.4.931.
Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*, 139–151. https://doi.org/10.1006/anbe.2000.1518.
Harris, P. L. (2012). *Trusting what you are told.* Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
Haspelmath, M. (1997). *Indefinite pronouns.* London: Clarendon.
Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*(2), 151–195.
Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*(1), 30–66.
Ifantidou, E. (2001). *Evidentials and relevance.* Amsterdam: Benjamin.
Izvorski, R. (1998). The present perfect as an epistemic modal. In A. Lawson, & E. Cho (Eds.). *Proceedings of SALT 7*. Ithaca, NY: CLC Publications.
Jakobson, R. (1971). *Studies on child language and aphasia.* The Hague/Paris: Mouton.
Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology, 61*, 248–272.
Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science, 17*, 757–758.
Johanson, L., & Utas, B. (Eds.). (2000). *Evidentials: Turkic, Iranian and neighbouring languages.* Berlin: de Gruyter.
Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28. https://doi.org/10.1037/0033-2909.114.1.3.
Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language, 72*, 533–578.
Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics, 4*, 109–128.
Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development, 83*(3), 1051–1063. https://doi.org/10.1111/j.1467-8624.2012.01742.x.
Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.
Koring, L., & De Mulder, H. (2015). Understanding different sources of information: The acquisition of evidentiality. *Journal of Child Language, 42*, 947–968. https://doi.org/10.1017/S030500091400052X.
Koster-Hale, J., Bedny, M., & Saxe, R. (2014). Thinking about seeing: Perceptual sources of knowledge are encoded in the theory of mind brain regions of sighted and blind adults. *Cognition, 133*(1), 65–78. https://doi.org/10.1016/j.cognition.2014.04.006.
Kratzer, A. (1991). Modality. In A. von Stechow, & D. Wunderlich (Eds.). *Semantics: An international handbook of contemporary research* (pp. 639–650). Berlin: de Gruyter.
Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.
Lee, T., & Law, A. (2000). Evidential final particles in child Cantonese. In E. Clark (Ed.). *Proceedings of the 30th annual child language research forum* (pp. 131–138). Stanford, CA: CSLI.
Levinson, S. C., & Meira, S. (2003). Natural concepts' in the spatial topological domain—Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language, 79*, 485–516.
Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition, 109*, 235–250.
Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., ... Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences of the United States of America, 115*(45), 11369–11376.
Maldonado, M., & Culbertson, J. (2019). *Something about "us": Learning first person pronoun systems.* In Proceedings of the 41st Annual Meeting of the Cognitive Science SocietyAustin, TX: Cognitive Science Society.
Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112*(3), 367–380. https://doi.org/10.1016/j.cognition.2009.05.012.
Matsui, T., & Fitneva, S. A. (2009). Knowing how we know: Evidentiality and cognitive development. In S. A. Fitneva, & T. Matsui (Eds.). *Evidentiality: a window into language and cognitive development (new directions for child and adolescent development 125)* (pp. 1–11). San Francisco: Jossey-Bass.
Matsui, T., Miura, Y., & McCagg, P. (2006). Young children's implicit and explicit understanding of speaker knowledge. In R. Sun, & N. Miyake (Eds.). *Proceedings of the 28th annual cognitive science society* (pp. 1789–1794). Mahwah, NJ: Lawrence Erlbaum.
Matsui, T., Yamamoto, T., & McCagg, P. (2006). On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development, 21*, 158–173. https://doi.org/10.1016/j.cogdev.2005. 10.001.
Matthewson, L. (2012). When fieldwork meets theory: Evidence about evidentials. In B.

Stolterfoht, & S. Featherston (Eds.). *Empirical Approaches to Linguistic Theory: Studies of Meaning and Structure, (pp. 85–114). Berlin.* Germany: De Gruyter.
McCready, E. (2008). Semantic heterogeneity in evidentials. *New Frontiers in Artificial Intelligence, LNCS, 4914*.
McCready, E., & Ogata, N. (2007). Evidentiality, modality, and probability. *Linguistics and Philosophy, 30*(2), 147–206.
Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *Quarterly Journal of Experimental Psychology, 64*(6), 1200–1220. https://doi.org/10.1080/17470218.2010.538211.
Murray, E. S. (2011). A Hamblin semantics for evidentials. In C. S. Ito, & D. Lutz (Vol. Eds.), *Proceedings from Semantics and Linguistic Theory (SALT) XIX. 2009. Proceedings from Semantics and Linguistic Theory (SALT) XIX* (pp. 324–341). Ithaca, NY: CLC Publications.
Murray, E. S. (2014). Varieties of update. *Semantics and Pragmatics, 7*, 1–53.
Mushin, I. (2001). *Evidentiality and epistemological stance: Narrative retelling.* Amsterdam: Benjamin.
Newport, E., & Aslin, R. (2004). Learning at a distance in statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162.
Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development, 12*(2), 199–230. https://doi.org/10.1080/15475441.2015.1024834.
Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition, 105*, 125–165. https://doi.org/10.1016/j.cognition.2006.09.008.
Papafragou, A., Li, P., Choi, Y., & Han, C. H. (2007). Evidentiality in language and cognition. *Cognition, 103*, 253–299. https://doi.org/10.1016/j.cognition.2006.04.001.
Pillow, B. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*, 116–129.
Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.
Plungian, V. A. (2001). The place of evidentiality within the universal grammatical space. *Journal of Pragmatics, 33*, 349–357. https://doi.org/10.1016/S0378-2166(00)00006-0.
Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development, 61*(4), 973–982.
Pyers, J. E., & Senghas, A. (2009). Language promotes false- belief understanding: Evidence from learners of a new sign language. *Psychological Science, 20*, 805–812. https://doi.org/10.1111/j.1467-9280.2009. 02377.x.
R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.
Rissman, L., Woodward, A., & Goldin-Meadow, S. (2019). Occluding the face diminishes the conceptual accessibility of an animate agent. *Language, Cognition and Neuroscience, 34*(3), 273–288. https://doi.org/10.1080/23273798.2018.1525495.
Rosch, E. H. (1972). Universals in color naming and memory. *Journal of Experimental Psychology, 93*(1), 10–20.
Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070.
San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., ... Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics, 26*, 31–60. https://doi.org/10.1515/cog-2014-0089.
Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development, 1*, 289–316.
Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Vol. Ed.), *The crosslinguistic study of language acquisition. Vol. 2. The crosslinguistic study of language acquisition* (pp. 1157–1256). Hillsdale, NJ: Lawrence Erlbaum Associates Theoretical issues.
Speas, M. (2004). Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua, 114*, 255–276.
Speas, M. (2018). Evidentiality and formal semantic theories. In A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality* (pp. 286–311). Oxford: Oxford University Press.
Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31–95.
Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*, 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x.
Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition, 121*(3), 409–415. https://doi.org/10.1016/j.cognition.2011.04.007.
Tabullo, Á., Arismendi, M., Wainselboim, A., Primero, G., Vernis, S., Segura, E., & Yorio, A. (2012). On the learnability of frequent and infrequent word orders: An artificial language learning study. *Quarterly Journal of Experimental Psychology, 65*(9), 1848–1863. https://doi.org/10.1080/17470218.2012.677848.
Thompson, S., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3*, 1–42.
Tily, H., Frank, M., & Jaeger, F. (2011). The learnability of constructed languages reflects typological patterns. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.). *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1364–1369). Boston, MA: Cognitive Science Society.
Ünal, E., & Papafragou, A. (2016). Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language, 89*, 179–199. https://doi.org/10.1016/j.jml.2015.12.001.
Ünal, E., & Papafragou, A. (2018). Relations between language and cognition: Evidentiality and sources of knowledge. *Topics in Cognitive Science. 1–21*Advance Online: Publication. https://doi.org/10.1111/tops.12355.
Ünal, E., & Papafragou, A. (2019). How children identify events from visual experience. *Language Learning and Development, 15*, 138–156.
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. https://doi.org/10.1016/j.jml.2015. 10.009.
Uzundag, B. A., Taşçı, S. S., Küntay, A. C., & Aksu-Koç, A. (2018). Functions of Turkish evidentials in early child-caregiver interactions: A growth curve analysis of longitudinal data. *Journal of Child Language, 45*(4), 878–899. https://doi.org/10.1017/s0305000911000183.
Van der Auwera, J., & Plungian, V. (1998). Modality's semantic map. *Linguistic Typology, 2*, 79–124.
Van der Henst, J-B., Bujakowska K., Ciceron C., & Noveck I.A.. (2006). How to make a participant logical: The role of premise presentation in a conditional reasoning task. In D. Andler, Y. Ogawa, M. Okada & S. Watanabe (Eds.), *Reasoning and Cognition* (pp. 7–17). Keio University.
Viberg, Å. (1984). The verbs of perception: A typological study. In B. Butterworth, B. Comrie, & Ö. Dahl (Eds.). *Explanations for language universals* (pp. 123–162). Berlin: Mouton de Gruyter.
de Villiers, J.G., Garfield, J., Gernet Girard, H., Roeper, T. & Speas, P. (2009). Evidentials in Tibetan: Acquisition, semantics and cognitive development. *New Directions for Adolescent and Child Development* (Special Issue on evidentiality), *125*, 29–48.
Wiemer, B. (2018). Evidentials and epistemic modality. In A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality* (pp. 85–108). Oxford: Oxford University Press.
Willett, T. (1988). A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language, 12*(1), 51–97.
Wilson, C. (2006). An experimental and computational study of velar palatalization. *Cognitive Science, 30*, 945–982.
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*.
Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*, 165–209. https://doi.org/10.1016/j.cogpsych. 2007.04.002.
Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences, 280*(1758), 20123073.