TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 14 (2022) 258–281

ISSN: 1756-8765 online
DOI: 10.1111/tops.12564

This article is part of the topic "Everyday Activities," Holger Schultheis and Richard P. Cooper (Topic Editors).

Developmental Trajectories in the Understanding of Everyday Uncertainty Terms

Björn Meder,[a,b] Ralf Mayrhofer,[c] Azzurra Ruggeri[b,d]

[a]*Department of Health, Health and Medical University Potsdam*
[b]*MPRG iSearch, Max Planck Institute for Human Development*
[c]*Department of Psychology, University of Göttingen*
[d]*School of Education, Technical University Munich*

Received 28 August 2020; received in revised form 23 June 2021; accepted 25 June 2021

Abstract

Dealing with uncertainty and different degrees of frequency and probability is critical in many everyday activities. However, relevant information does not always come in the form of numerical estimates or direct experiences, but is instead obtained through qualitative, rather vague verbal terms (e.g., "the virus *often* causes coughing" or "the train is *likely* to be delayed"). Investigating how people interpret and utilize different natural language expressions of frequency and probability is therefore crucial to understand reasoning and behavior in real-world situations. While there is considerable work exploring how adults understand everyday uncertainty phrases, very little is known about how children interpret them and how their understanding develops with age. We take a developmental and computational perspective to address this issue and examine how 4- to 14-year-old children and adults interpret different terms. Each participant provided numerical estimates for 14 expressions, comprising both frequency and probability phrases. In total we obtained 2856 quantitative judgments, including 2240 judgments from children. Our findings demonstrate that adult-like intuitions about the interpretation of everyday uncertainty terms emerge fairly early in development, with the quantitative estimates of children

Correspondence should be sent to Björn Meder, Health and Medical University Potsdam, Olympischer Weg 1, 14471 Potsdam, Germany. E-mail: bjoern.meder@health-and-medical-university.de, meder@mpib-berlin.mpg.de

converging to those of adults from around 9 years on. We also demonstrate how the vagueness of verbal terms can be represented through probability distributions, which provides additional leverage for tracking developmental shifts through cognitive modeling techniques. Taken together, our findings provide key insights into the developmental trajectories underlying the understanding of everyday uncertainty terms, and open up novel methodological pathways to formally model the vagueness of probability and frequency phrases, which are abundant in our everyday life and activities.

Keywords: Everyday uncertainty terms; Verbal uncertainty terms; Frequency phrases; Probability phrases; Development; Everyday activities; Computational modeling

1. Introduction

Natural language expressions of frequency and probability are ubiquitous in our everyday lives and activities. Imagine you are planning and preparing your birthday dinner party. Some of your prospective guests tell you that they are *most likely* or *almost certain* to come, whereas others tell you that they are *unlikely* to make it. Despite the uncertainty as to who will eventually show up, you need to go ahead with planning. You know that some of your friends *frequently* travel by car and therefore will *probably* not drink alcoholic beverages, whereas others *often* come by public transport and are more *likely* to enjoy a drink. You know that about half of your friends are vegetarians, and it is *possible* that their partners are as well. Also, you would love to prepare a fancy dish, but in your experience complicated meals *rarely* work out exactly the way they are supposed to, and *sometimes* do not taste as great as promised.

As this example illustrates, verbal frequency and probability terms are ubiquitous and play a crucial role in many aspects of our daily lives—in thinking, communication, and many of our activities. And, of course, even scientists trained in formal methods and statistics often [sic!] use these rather vague and imprecise terms in sentences such as "similar findings have *frequently* been reported in the literature" or "it is *unlikely* to obtain such a result merely by chance." Nevertheless, how exactly we (learn to) understand, represent, and utilize everyday uncertainty terms is still poorly understood. This stands in stark contrast to the eminent role they play in our lives: without a shared understanding of such terms communication is prone to misunderstandings, and planning, executing, and coordinating actions with others would be difficult and unreliable.

While there are several studies assessing how adults interpret different frequency and probability phrases, there is only little work exploring from a developmental perspective how we learn to make sense of and attribute meaning to these expressions. The present paper investigates how children aged 4–14 interpret everyday uncertainty expressions, how closely their quantitative judgments resemble those of adults, and at what point in development adult-like intuitions emerge. Moreover, we demonstrate how the vagueness of linguistic terms can be represented within a probabilistic framework, which provides new pathways for building computational models of everyday reasoning and action.

2. Mapping words to numbers: How do people understand everyday expressions of uncertainty?

Research on how people understand expressions of frequency (e.g., *often*) and probability (e.g., *likely*) extends back to the middle of the 20th century (Cliff, 1959; Lichtenstein & Newman, 1967; Simpson, 1944, 1963), with several studies investigating what numerical equivalents adults assign to different terms (for reviews, see Clark, 1990; Mosteller & Youtz, 1990; Teigen & Brun, 2003; Wallsten & Budescu, 1995). In these studies, participants are typically asked to map words to numbers by providing quantitative judgments (e.g., percentages or frequencies) for different terms.

Two key findings have emerged from this line of research. On the one hand, the perceived meaning of verbal phrases can vary depending on context (Brun & Teigen, 1988; Weber & Hilton, 1990), base rates (Wallsten, Fillenbaum, & Cox, 1986b), kinds of events (Harris & Corner, 2011; Weber & Hilton, 1990), conversational rules and pragmatics (Bonnefon & Villejoubert, 2006; Honda & Yamagishi, 2017; Teigen & Brun, 1999), as well as the employed elicitation method (Hamm, 1991; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986a; Wallsten, Budescu, & Zwick, 1993). Generally, within-subject variation tends to be lower than between-subjects variability, indicating a fairly stable understanding of everyday uncertainty terms at the individual level, although different people may vary in their judgments (Budescu & Wallsten, 1985). On the other hand, between-subjects variation in the interpretation of particular terms notwithstanding, people can also be quite consistent in their quantitative judgments (Simpson, 1963). For instance, Mosteller and Youtz (1990) evaluated 52 expressions across 20 different studies and found that "the studies give similar, though not identical, results for the same expression when sampling and other sources of variability are considered" (p. 3). These analyses indicate that in many cases people have a shared understanding of linguistic uncertainty terms, in that they assign similar quantitative estimates to them.

To what extent people have a common understanding of uncertainty phrases is also critical from an applied perspective, as such expressions are used in many fields to communicate quantitative information obtained from experts (who are typically trained in statistical methods) to lay people, who lack this expertise but need to integrate this information in their reasoning and decision-making processes. In such cases, it is important to use calibrated language to ensure that the relevant information is appropriately understood, as mismatches between intended and perceived meaning loom large when verbal information forms the basis for decision making on the individual and policy level. For instance, reports of the Intergovernmental Panel on Climate Change (IPCC) use a codification scheme for mapping probability information (e.g., model predictions regarding temperature increase) to verbal expressions, to communicate the current state of scientific knowledge and the associated uncertainties to policy makers and the public (Mastrandrea et al., 2011). However, the numerical equivalents that people intuitively assign to these phrases do not always match the intended meaning, raising concerns about the appropriateness and effectiveness of the used codification schemes (Budescu, Broomell, & Por, 2009; Budescu, Por, Broomell, & Smithson, 2014; Harris, Corner, Xu, & Du, 2013b). Similar results have been obtained in other fields, such as medical risk communication (Berry, Knapp, & Raynor, 2002) and forensic science (Martire, Kemp, Sayle, & Newell, 2014; Thompson & Newman, 2015). These findings highlight the importance of carefully investigating how people understand linguistic expressions of uncertainty in different circumstances, and the importance of taking into account behavioral research when devising a codification scheme that explicates the relation between words and numbers.

3. (How) Do children understand everyday uncertainty terms?

While there is a rich literature on how adults interpret verbal expressions of uncertainty, little is known about how children interpret such terms and how their understanding develops with age.

Kuczaj (1975) investigated whether preschoolers aged 4–5 comprehended deterministic frequency terms like *always* and *never*, and found that most—but not all—children judged sentences containing these words as understandable. However, sentences containing frequency terms such as *sometimes*, *usually*, and *seldom* were judged by many children as not understandable, indicating that their meaning may be acquired at a later age. Hoffner, Cantor, and Badzinski (1990) analyzed 5- to 11-year-olds' understanding of the three probability terms *possibly*, *probably*, and *definitely*. Younger children showed a limited understanding of these terms and were not able to appropriately distinguish between them, but fourth graders were able to do so. Along these lines, Mullet and Rivet (1991) found that, with a sample of 9-, 12-, and 15-year-olds, older children were better than younger children at discriminating among 12 probability phrases (e.g., *likely*, *low chance*), and that judgment variability decreased with age. Biehl and Halpern-Felsher (2001) studied fifth, seventh, and ninth graders (age 10–14), as well as adults, asking them to assign numerical percentages to 30 linguistic expressions. They observed a fairly broad agreement across age groups in terms of the mean estimates, but also some significant differences. Similar to other studies, judgment variation was higher for children and adolescents than for adults.

4. Goals and scope

How does the understanding of everyday uncertainty terms develop with age and when do adult-like intuitions regarding their interpretation emerge? This paper makes two main contributions to address this question. First, we report empirical data on quantitative estimates for several linguistic expressions of frequency and probability, and for a broader age range than previously reported. For instance, Biehl and Halpern-Felsher (2001) investigated several terms in children 10 years and older, but their sample was limited to early adolescence, as participants were required to assign numerical percentages to the terms. Thus, a primary goal of the present study was to map developmental trajectories in the understanding of verbal uncertainty terms across childhood, rather than testing specific factors that could influence children's understanding of verbal expressions. To do so, we developed a nonnumerical experimental paradigm suitable for younger children, enabling us to investigate the understanding of uncertainty phrases within the same paradigm across a broad age range. Our data trace a full developmental trajectory of the interpretation of the different verbal expressions under consideration, showing children's quantitative estimates approximate those of adults' judgments around age 9.

The second contribution is methodological. Earlier studies with adult subjects used the framework of *fuzzy set theory* (Zadeh, 1965) to formally represent the inherent vagueness of linguistic terms (Bocklisch, Bocklisch, & Krems, 2012; Reagan, Mosteller, & Youtz, 1989; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986a; Zadeh, 1975). We introduce a conceptually distinct approach using a probabilistic modeling framework, where probability distributions quantify to what extent different numerical values belong to a term (cf. Meder & Mayrhofer, 2017). This method provides additional means to trace developmental trends and enables researchers to harness the framework of probabilistic inference for data analysis and cognitive modeling. We used this approach to formally assess the similarity of different age groups' distributions, providing an additional window into the development of understanding everyday uncertainty terms.

5. Method

5.1. Participants

Participants were recruited from public museums in Berlin, Germany, and tested individually using a tablet-based experiment. Our sample includes 44 adult participants (mean age 34.6, $SD = 11.8$) and 160 children: $N = 22$ 4- to 5-year-olds (mean age 4.7, $SD = 0.46$); $N = 22$ 6-year-olds; $N = 26$ 7-year-olds; $N = 23$ 8-year-olds; $N = 27$ 9-year-olds; $N = 18$ 10-year-olds; and $N = 22$ 11- to 14-year-olds (mean age = 11.7, $SD = 1.03$). An additional 24 children were excluded from the analyses for the following reasons: not native German speakers ($N = 3$), parents intervened during the experiment ($N = 3$), did not understand the instructions ($N = 3$), failed the practice test ($N = 10$), were too young ($N = 2$), missing consent ($N = 1$), or always answered with 0 or 100 throughout the experiment ($N = 2$). The study was approved by the Ethics Committee of the Max Planck Institute for Human Development; written informed consent was obtained from all participants/from their legal guardian. Children received stickers for participating.

5.2. Design

We elicited quantitative judgments for 14 different expressions, including seven frequency terms and seven probability terms (Table 1). Because we elicited judgments from children as young as 4 years old, it was important to keep the experiment and used uncertainty expressions as simple as possible. We therefore decided to use a variety of single-word terms covering different degrees of probability and frequency, rather than constructing composite expressions based on a limited number of root terms combined with different modifiers

Table 1
Frequency and probability terms used in the present study

Expression	Original German	Type
rarely	selten	frequency
sometimes	manchmal	frequency
occasionally	gelegentlich	frequency
half of the cases	Hälfte der Fälle	frequency
often	oft	frequency
frequently	häufig	frequency
most of the time	meistens	frequency
unlikely	unwahrscheinlich	probability
uncertain	unsicher	probability
perhaps	eventuell	probability
maybe	vielleicht	probability
possibly	möglich	probability
equiprobable	gleichwahrscheinlich	probability
likely	wahrscheinlich	probability

Note. Sometimes an English term could correspond to multiple German translations, and vice versa. For instance, the German term *wahrscheinlich* could be translated as *probably* or *likely*; for the purpose of this paper we chose the latter. The German word *selten* could be translated as *seldom*, *infrequently*, or *rarely*; we here use the latter.

(e.g., "very likely," "highly likely" etc.). The set of frequency expressions included all six single-word frequency expressions and the compound expression *half of the cases* considered by Bocklisch et al. (2012) who elicited numerical equivalents from a German-speaking sample of adult participants. The set of probability expressions included the German equivalents of the three single-word probability expressions *likely*, *possibly*, and *unlikely* from the meta-analysis by Mosteller and Youtz (1990). In addition, we included the probability terms *perhaps*, *equiprobable*, *uncertain*, and *maybe*.

5.3. *Materials and procedure*

We developed a child-friendly paradigm where quantitative estimates were provided using a slider on a tablet. Underlying the slider was a scale from 0 to 100 in steps of 1, but no numbers were shown throughout the experiment. To account for age-related differences in reading abilities, instructions on the screen were always read out aloud by the experimenter throughout the whole study. This was done for all participants regardless of age, in order to keep the experimental procedure identical across subjects. During the introduction and practice trials, participants were repeatedly asked if they understood the instructions and mechanics of the task and given the opportunity to ask clarification questions, which were answered verbally. During the actual judgment phase, the instructions for each term were read out aloud in the order in which they appeared on the screen (see next), with the experimenter pointing to the relevant part of the screen. All judgments were recorded by the software; we did not audio- or video-record the sessions.

Fig. 1. Example trials. The upper panel shows example trials for the frequency terms *rarely* and *often*. The lower panel shows example trials for the probability terms *unlikely* and *likely*.

Participants were first introduced to the cover story. They were told that they would visit different planets, each home to different (friendly) monsters. Throughout their journey they would be accompanied by Robbie the robot (Fig. 1), who had already been to all planets and knew everything about them. Next, participants were familiarized with the slider, which was introduced as a "control panel" they could use to answer questions in the game. The unlabeled slider was shown on a blank screen and participants were instructed to move the slider to different positions, including the middle, the rightmost, and the leftmost point.

The structure of each trial in the subsequent experimental phase was as follows. First, participants were shown a planet with two monsters differing in a single binary feature (e.g., dotted or not dotted, paws or no paws; Fig. 1). Next participants were presented with the term-evaluation question they should answer using the slider. Participants were first presented with four practice trials (in randomized order), with the two frequency terms *always* and *never*, and the two probability terms *impossible* and *certain*. We chose these terms, assumed to correspond to the extreme values of the slider (i.e., values of 0 or 100), to ensure that children understood the meaning of these relatively simple and clearly defined terms, and to further check they understood how to use the slider (the labels of the end points were identical to the subsequent judgment phase). After the practice trials, participants were asked to provide quantitative estimates for each of the terms shown in Table 1.

We used slightly different instructions for the frequency and probability terms. The frequency phrases referred to the population of monsters living on the planet. The probability phrases pertained to a single monster living on that planet. This was done to avoid that participants translated the probability terms into a frequency format or the other way around, thereby blurring the conceptual distinction. For the frequency terms participants were asked: "What do you think: How many monsters on the planet have [feature]?" Then Robbie appeared on the screen with a speech bubble saying "Monsters on this planet [term] have [feature]." For instance, during the practice trials Robbie might state that "Monsters on this planet *never* [*always*] have an antenna." In the actual trials Robbie might say "Monsters on this planet *rarely* have stripes" or "Monsters on this planet *often* have paws" (Fig. 1).

For the probability terms the following instruction was given: "Imagine you land on the planet and meet one of the monsters. What do you think: Does the monster rather have an antenna or rather not?" Then Robbie appeared on the screen with a speech bubble saying "It is [term] that the monster has an [feature]." For instance, during the practice trials Robbie might state that "It is *impossible* [*certain*] that the monster has an antenna." In the actual trials Robbie might say "It is *unlikely* that the monster is dotted" or "It is *likely* that the monster has a tail" (Fig. 1).

After Robbie provided his information, participants were asked to answer the question using the slider, which appeared on the screen with the initial position set to the midpoint of the scale (Fig. 1). For the frequency terms, the endpoints of the slider were labeled "None of them have [feature]" and "All of them have [feature]." For the probability terms, the endpoints were labeled "Certainly does not have [feature]"and "Certainly does have [feature]."

After providing a judgment using the slider, the next trial started, presenting a new planet with different monsters and a novel feature. All terms were presented in random order, with the assignment of monsters and features to terms being randomized. Upon providing quantitative judgments for all 14 terms, the experiment ended.

6. Results

In total, we obtained $204 \times 14 = 2856$ quantitative judgments, including 2240 judgments from children aged 4–14 (17 estimates were not recorded due to technical error and not included in the analyses). We conducted both group-level and individual-level analyses to assess developmental trends in how children interpreted the different expressions. In addition, 10 participants provided a wrong judgment in more than two practice questions (i.e., did not assign maximal or minimal values by moving the slider to the leftmost or rightmost position) and were excluded from the analyses.

We compared adults' mean judgments in our sample to the mean estimates reported in the literature. Six of the frequency terms we used were included in Bocklisch et al. (2012) who elicited them from German-speaking adults. The correlation with these estimates was $r = .99$, with a mean absolute deviation (MAD) of 5.8. Mosteller and Youtz (1990) reported the mean estimates for several expressions across 20 studies with English-speaking participants, of which nine corresponding German terms were used in the present study (six frequency and three probability terms; here we used the English term "infrequently" for the German term "selten"). The correlation of adults' mean judgments in our study with the estimates reported in Mosteller and Youtz (1990) was $r = .96$, MAD = 9.3. Thus, adults' mean judgments in our paradigm were comparable to existing findings obtained with different methods and subject samples, and there was a reasonable cross-language consistency between the German terms used in our study and the corresponding English terms.

6.1. Group-level analyses

Fig. 2 shows the distribution of children's and adults' judgments for the different expressions. These plots reveal two main findings. First, for most expressions the variability in the quantitative judgments is much higher for young children (e.g., 4- to 6-year-olds), compared to older children and adults, indicating that older children and adults were more consistent in how they interpreted the terms. Also note that the only two sharply defined phrases corresponding to a clear and objective numerical estimation, *in half of the cases* and *equiprobable*, almost always received judgments of 50 only from age 8, whereas younger children showed much more variation in their judgments of these phrases.

Second, children's mean judgments converged to adult-like response patterns relatively early in development, for both types of terms. From age 9 the correlation of children's estimates with those obtained from adults was very high (Pearson's $r > .9$). The correlation of children's and adults' mean judgments increased strongly with age, while at the same time the MAD strongly decreased (see the Supporting Information for the full correlation matrix across all age groups).

Note that older children continued to further approximate adults' estimates, both in terms of correlation and MAD. This development in early adolescence is also supported by data from Biehl and Halpern-Felsher (2001), who asked 10-, 12-, and 14-years-olds, as well as young adults, to assign percentages to 30 terms. A reanalysis of their data shows that the correlation of 10-year-olds' mean estimates with those of adults was .93, further increasing to .97 and .99, respectively, for 12- and 14-year-olds (see the Supporting Information for details). Similarly, the MAD of children's average estimates from those of adults further decreased in that age range, from 11 for 10-year-olds to 9.7 and 5.6, respectively, for 12- and 14-year-olds.

Taken together, these results suggest that adult-like intuitions about the meaning of everyday uncertainty terms emerge quite early in development, and are further refined in early adolescence.

6.2. Variability of judgments

The between-subjects variability in the quantitative estimates strongly varied with age, with younger children being less consistent in their judgments than older children and adults (Fig. 2). We computed for each age group and term the variance in the judgments. Fig. 3a shows that the between-subjects variability across the verbal expressions decreased strongly as children grow older; from about 9 years on children are as consistent as adults.

Large parts of the variance in young children's judgments can be traced to a tendency to assign extreme values of 0 or 100 to the different terms, thus giving judgments of certainty, rather than uncertainty (Fig. 3b). These analyses suggest that younger children's judgments were not merely noisier than those of adults, in which case one would expect more evenly distributed answers across the response scale. Rather, young children's judgments were

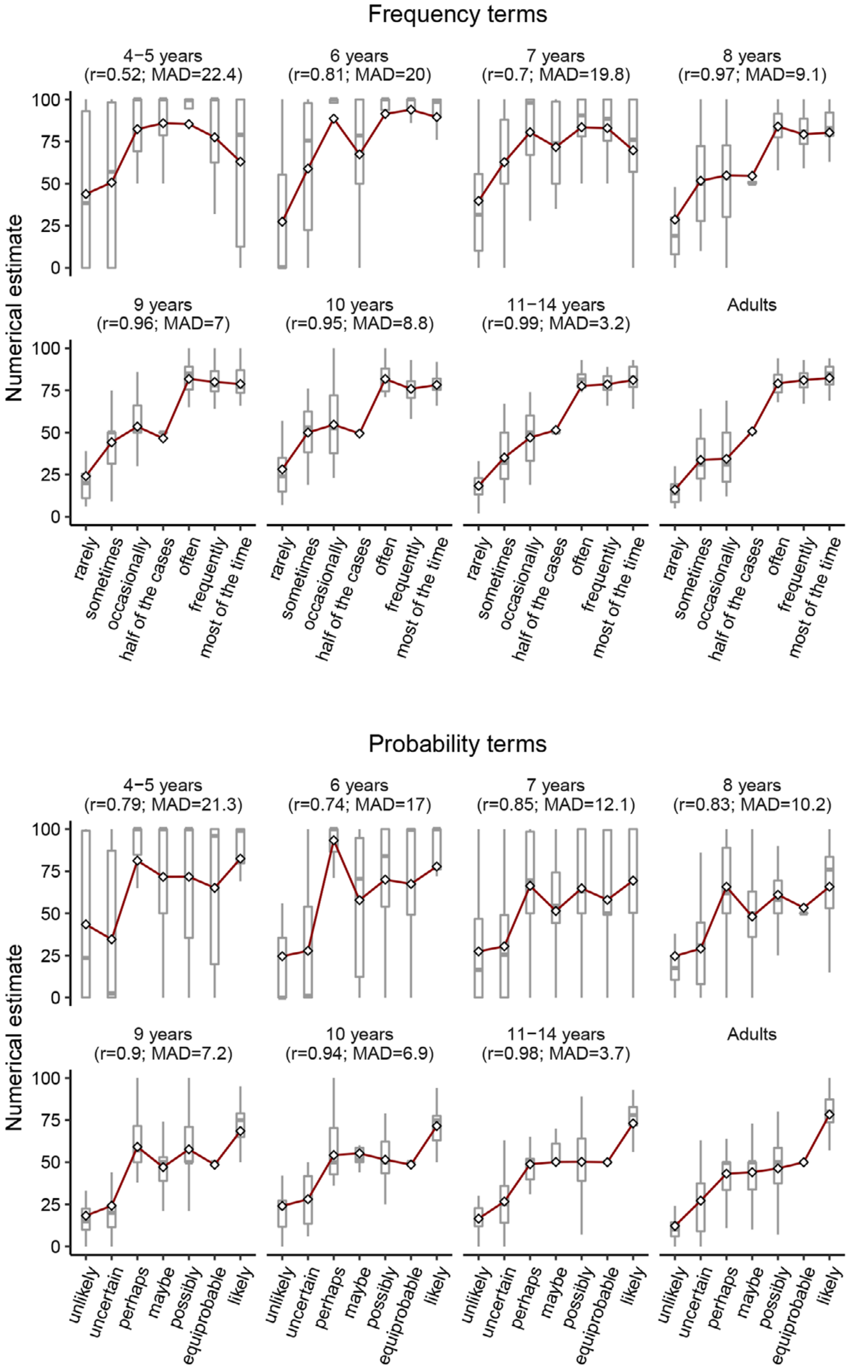

Fig. 2. Tukey box plots of the quantitative estimates assigned to different frequency and probability expressions. Diamonds indicate group means, the horizontal line in the box shows the group median. r = Pearson correlation with adults' mean estimates, MAD = mean absolute deviation from adults' average estimates.

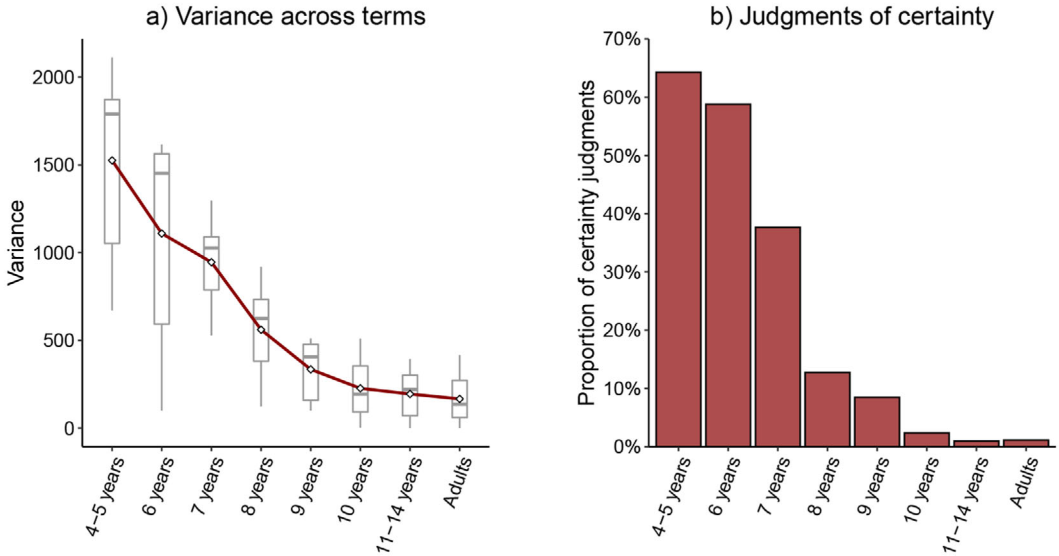

Fig. 3. Variance in quantitative estimates and proportion of certainty judgments.

characterized by a strongly dichotomous response pattern. This finding echoes other developmental findings, which consistently indicated that young children tend to make extreme judgments across different response scales and formats (Chambers, 2002; Light, Zax, & Gardiner, 1965; O'Dowd, 1980; Zaman, Abeele, & De Grooff, 2013), in line with the Piagetian theory positing that young children characteristically engage in dichotomous thinking (Gelman & Baillargeon, 1983). This tendency to give extreme judgments quickly tapered off with age though, indicating that children's interpretation of the different terms became more nuanced and consistent as they got older (Fig. 3b).

Note that while we observed a large proportion of extreme judgments in children, especially for ages 4–7, only 15 of 160 children always assigned 0 or 100 to the different terms (four 4- to 5-year-olds, six 6-year-olds, four 7-year-olds, and one 8-year-old). Thus, even young children showed some sensitivity to the gradedness of the terms.

6.3. *Individual-level analyses*

To account for the variability in children's judgments we computed the Pearson correlation of each child's individual judgments with adults' mean estimates (Fig. 4, upper panel). Consistent with the group-level analyses, we obtained a strong developmental trend, with children's judgments increasingly converging to the average estimates of adults. This is also reflected in the deviation of children's individual judgments from adults' mean estimates, which strongly declined with age (Fig. 4, lower panel). Also note that the relationship between children's and adults' judgments becomes less variable as they get older, both in terms of correlation with and deviation from adults' mean estimates.

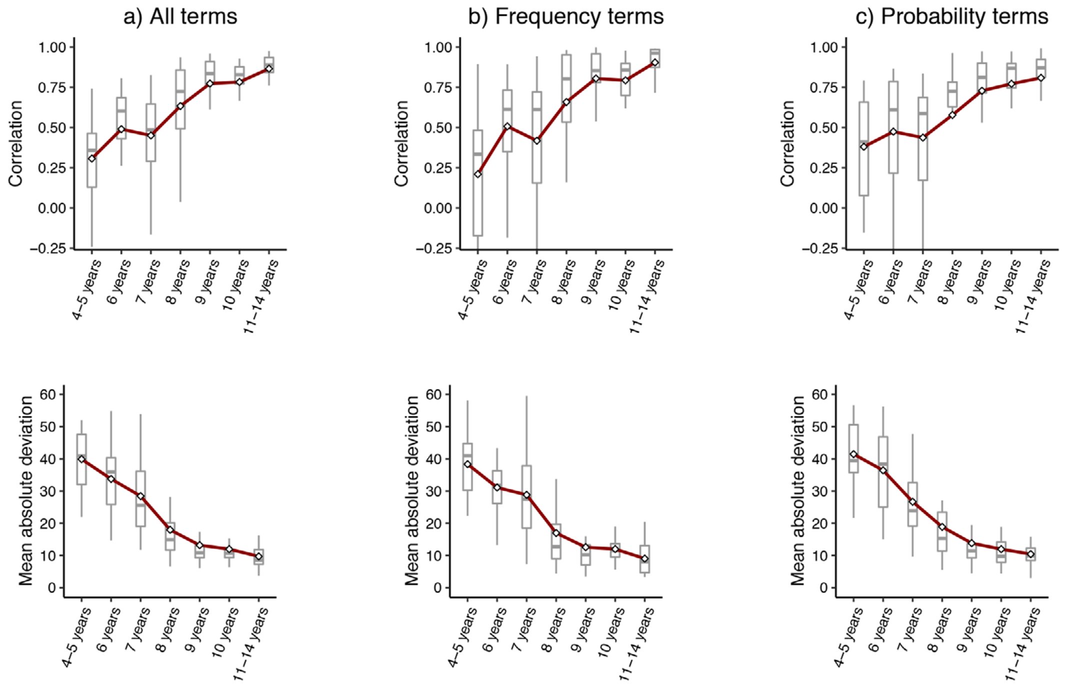

Fig. 4. Tukey box plots of children's individual (Pearson) correlations with adults' mean estimates (upper panel) and their mean absolute deviation from these estimates (lower panel). The horizontal line in the box shows the group median and the diamonds indicate the group means, connected by the line.

Taken together, our results map strong and consistent age-related trends in the understanding of everyday language expressions of uncertainty. While young children's judgments were highly variable and characterized by a tendency to assign extreme values to many terms, this tendency quickly diminished with age, with children's judgments approaching those of adults from about age 9. These findings demonstrate that adult-like intuitions about the meaning and nuances of verbal uncertainty terms emerge quite early in development.

7. Modeling the vagueness of verbal uncertainty terms

A characteristic feature of everyday uncertainty expressions is that they are not sharply defined, but are inherently imprecise or vague. There are some boundary cases which have a clear-cut interpretation that is also reflected in the numerical estimates assigned to them, such as *always* or *never* (although, never say never). Generally though, when linguistic expressions are mapped to quantitative estimates, there typically is not a single value that people consider to exactly correspond to the term, but multiple values (or a range of values) are assumed to represent the term to a varying extent. In this sense, everyday uncertainty phrases resemble

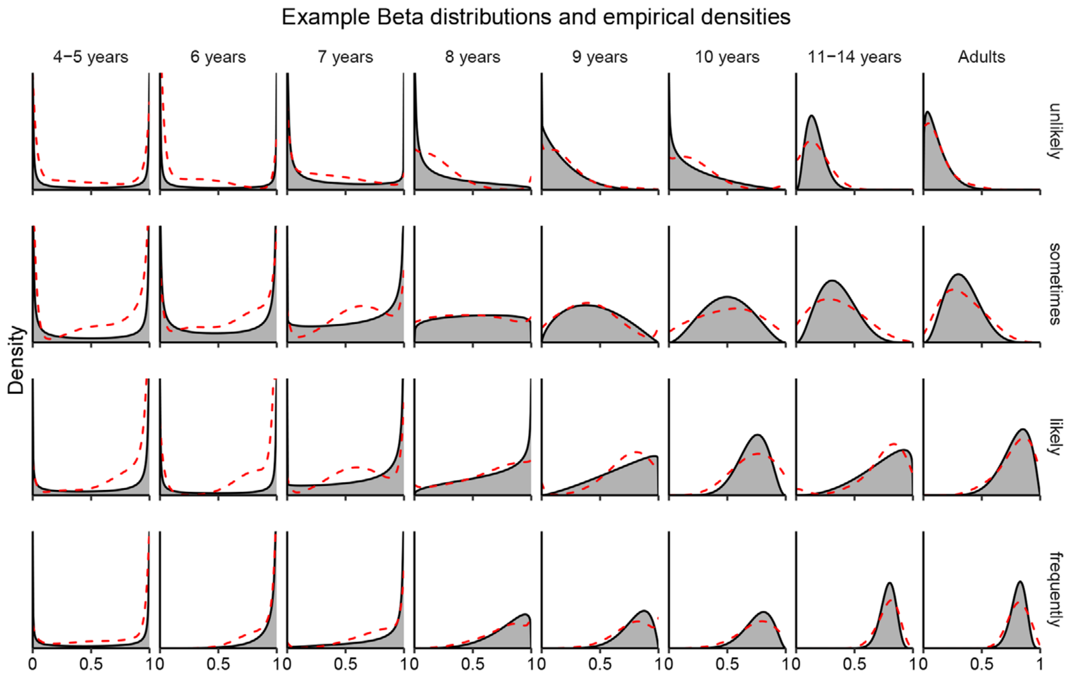

Fig. 5. Fitted beta distributions and empirical densities (dashed line) for the terms *unlikely*, *sometimes*, *likely*, and *frequently*. *Note*. We used the $\hat{f}_1$ beta-kernel estimator recommended by Chen (1999) to estimate the empirical densities. Because the beta distribution is not defined on the bounds of the unit interval, we shifted data points located exactly on the bounds (i.e., judgments of 0 and 1, respectively) by 0.001.

fuzzy sets (Zadeh, 1975) or prototype representations (Rosch, 1973; Wittgenstein, 1953) of concepts—there might be typical values, and more or less likely numerical values that could be subsumed under a term, but the boundaries are not sharply defined.

Previous work has often used *membership functions* (Bocklisch et al., 2012; Rapoport, Wallsten, Erev, & Cohen, 1990; Reagan et al., 1989; Wallsten et al., 1986a; Zadeh, 1975) to represent the vagueness of verbal expressions and to account for the variability in people's quantitative judgments. Formally, this approach is well defined, with fuzzy set theory (Zadeh, 1965) providing the mathematical framework. However, membership functions are not density functions, therefore it is difficult to analyze them using common statistical methods or to integrate them with probabilistic models of cognition. A conceptually distinct approach is to represent the vagueness of verbal terms using probability distributions, where a density function on the interval [0,1] is used to denote the likelihood of different numerical values belonging to the concept (Meder & Mayrhofer, 2017, also see Dhami & Wallsten, 2005). Represented this way, each numerical value has a certain likelihood of belonging to a particular term, and the dispersion of the distribution encodes the phrases' inherent vagueness.

A natural choice is the family of beta distributions, which are defined on the interval [0,1] and are parameterized by shape parameters α and β. Consider Fig. 5, where the densities represent the frequency terms *sometimes* and *frequently* and the probability terms *unlikely*

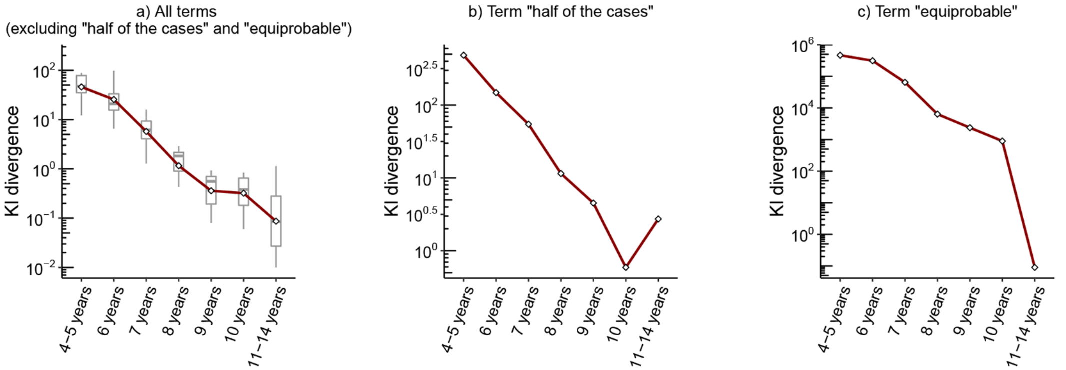

Fig. 6. Kullback–Leibler (KL) divergence between children's and adult's beta distributions representing linguistic expressions of frequency and probability.

and *likely*. To obtain these distributions we mapped the obtained quantitative estimates in the range [0,100] to the interval [0,1] by dividing each value by 100. We then calculated for each term and age group the beta distributions' shape parameters α and β based on the sample mean and variance, using the method of moments (see Appendix A.1 and Supporting Information for details and overview of all terms used in the present study).

The fitted beta distributions have the same mean and variance as the empirical estimates. However, if there is large variation in the quantitative estimates, similarities in the distributions' central tendencies (e.g., mean or median judgments) are not particularly informative. The beta distributions, by contrast, have more expressive power to capture key aspects of the behavioral data. One example is the tendency of young children to assign extreme values of 0 or 100 to the terms, which renders the mean and median of the empirical distribution largely uninformative. By contrast, the beta distribution appropriately represents this tendency, as indicated by the U-shaped distribution for children until about age 6, where most of the density is located toward the boundary values 0 and 1. As children grow older, this tendency tapers off, with the variance decreasing and the shape of the distribution approaching that of adults.

To further analyze the observed developmental trends and to assess the similarity of the distributions across age groups we computed the Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951) between children's distributions and the corresponding distributions derived from adults' judgments (Appendix A.1). Fig. 6a shows the KL divergence across all terms, excluding the terms "half of the cases" and "equiprobable." The plot shows how the KL divergence strongly decreases across childhood, indicating that the distributions representing the terms become more similar to those of adult participants with age.

Fig. 6b and c plots the KL divergence for the frequency term *half of the cases* and *equiprobable*, respectively. The reason for plotting these expressions separately is that 11- to 14-year-olds and adults almost always (correctly) gave judgments of 50, rendering the variance minimal (for term *half of the cases*) or zero (for term *equiprobable*). In these cases, the shape parameters of the corresponding beta distributions are very high or not defined, respectively,

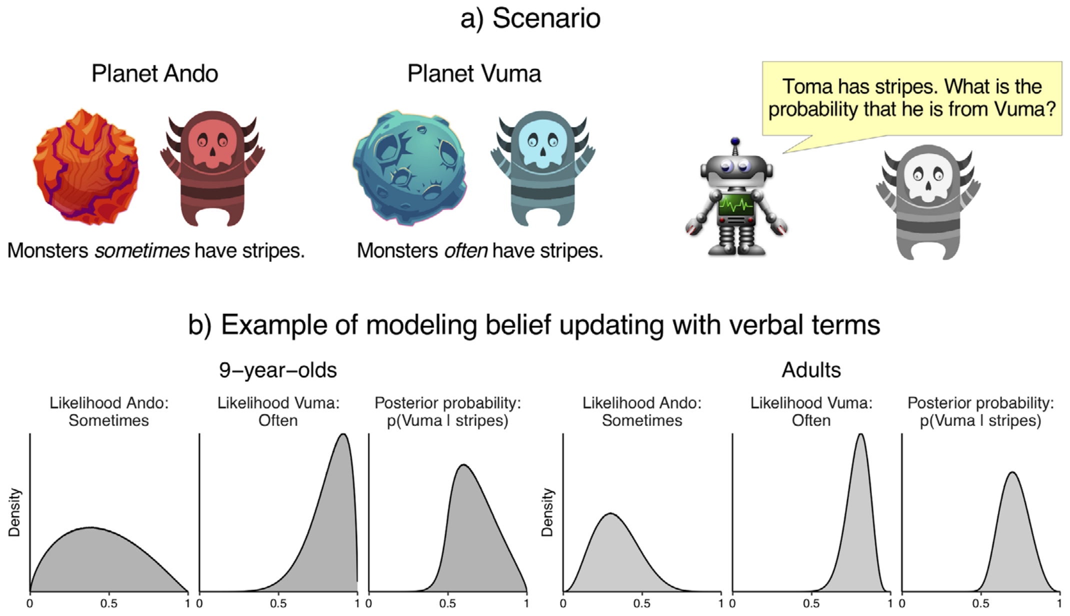

Fig. 7. Example of belief updating with everyday uncertainty terms. (a) Scenario where the goal is to derive the posterior probability that Toma comes from Planet Vuma, given that he has stripes. (b) Modeling Bayesian reasoning, where the probability distributions representing the terms "sometimes" and "often" for 9-year-olds and adults, respectively, serve as likelihoods to derive the posterior distribution over $p(\text{Vuma}|\text{stripes})$, assuming equal prior probabilities (i.e., $p(\text{Vuma}) = p(\text{Ando}) = 0.5$).

as all or most of the density is located at 0.5. For the term *equiprobable* we therefore manually set the shape parameters to $\alpha = \beta = 10^5$, such that virtually all density is located at 0.5 and the variance is minimal (Appendix A.1). As can be seen from the plots, the developmental shift observed for all other expressions holds for these terms, too—the older children get, the closer their distributions resemble those of adult participants.

In sum, using probability distributions to represent the vagueness of verbal expressions provided additional traction for tracing developmental trajectories in the quantitative estimates children assign to them. Importantly, this approach offers new methodological pathways for investigating how children and adults reason with such vague and nonnumerical information (Meder & Mayrhofer, 2017). Probability distributions representing verbal terms seamlessly integrate with Bayesian inference mechanisms, making it possible to model different kinds of reasoning processes and belief updating.

Fig. 7a illustrates a belief updating task with verbal uncertainty terms. Robbie has visited two planets, Ando and Vuma. On both planets there live monsters with stripes, but in different proportions: On Vuma, monsters "often" have stripes, whereas on Ando monsters "sometimes" have stripes. Robbie made a new friend, Toma, who has stripes. What is the probability that Toma is from planet Vuma? Formally, this question corresponds to computing the posterior probability $p(\text{Vuma}|\text{stripes})$. Using probability distributions to represent the verbal terms enables a formal treatment of this inference that preserves the vagueness of the expressions, where the terms' distributions represent the likelihoods (probability of stripes on planets Vuma and Ando, respectively) and the posterior distribution is derived using Bayesian inference. The example in Fig. 7b illustrates this for adults and 9-year-olds, using the fitted beta distributions of the term "sometimes" and "often" of the age groups to derive the posterior distribution over $p(\text{Vuma}|\text{stripes})$. (For simplicity, we here assume equal priors, that is, $p(\text{Vuma}) = p(\text{Ando}) = .5$.) Deriving age-specific posterior distributions enables the derivation of precise and empirically testable predictions about expected mean (or median) judgments, variability of judgments, and shape of the distribution. Thus, whereas both membership functions and probability distributions enable a formal representation of verbal uncertainty terms, using probability distributions additionally provides traction for modeling more complex inferences based on nonnumerical, rather vague verbal information.

8. Concluding remarks and future directions

The present study mapped developmental trajectories in the understanding of everyday uncertainty terms. Our analyses show strong and systematic differences in the quantitative estimates that children aged 4–14 assign to different frequency and probability expressions, suggesting that adult-like intuitions of their meaning emerge around age 9. This finding was true both in terms of how strongly children's judgments correlated with adults' judgments at the group and individual level, and when evaluating the similarity of the probability distributions used to represent the verbal terms via KL divergence.

Our findings bear several relations to key issues in developmental research. One important question is how the understanding of everyday uncertainty expressions relates to the development of quantitative reasoning skills. Several studies have demonstrated that even infants are sensitive to core principles of probabilistic inference (Denison, Reed, & Xu, 2013; Xu & Garcia, 2008), and that both infants and preschoolers are able to use probability information in judgment and decision making (see Gweon, Tenenbaum, & Schulz, 2010; Kushnir, Xu, & Wellman, 2010). Notwithstanding these early competencies, even 10-year-olds perform substantially worse than adults' in simple proportional reasoning tasks, even when supported by children-friendly presentation formats (Ruggeri, Vagharchakian, & Xu, 2018). While this is consistent with developmental theories positing that the ability to represent and manipulate quantitative and numerical proportions depends on the acquisition of a verbally mediated system of numbers and inference strategies assumed to mature only in late adolescence (Carey, 2009; Inhelder & Piaget, 1958), such findings also raise important questions about the interplay between the ability to reason with numerical information and the understanding of verbal uncertainty terms. Is the ability to reason with numerical information a precursor for a shared understanding of verbal probabilities? Conversely, to what extent does sound proportional reasoning require a more mature understanding of verbal uncertainty terms? For instance, one could construct proportional reasoning tasks where participants receive the relevant information either numerically or verbally through matched linguistic terms (Meder & Mayrhofer, 2017), and compare performance on these tasks across development. Such studies would help bridging the literature on the development of quantitative reasoning skills and the understanding of verbal probabilities, thereby fostering theory integration and providing insights than cannot be gained from investigating these competencies in isolation.

How exactly do children develop their understanding of verbal uncertainty terms, and what factors do contribute to their learning process over the course of development and through their everyday activities? This question is particularly interesting because, compared to other competencies such as reading, writing, and doing basic math, this ability is not subject to any formal training—no one explicitly teaches children the meaning of words such as *likely* or *sometimes*. One approach to shed light on the underlying learning processes would be to conduct ecological analyses that relate children's learning processes to their everyday social and informational environments. For instance, researchers could study the prevalence and use of everyday uncertainty terms in conversational interactions between children and parents by analyzing corpus data such as the Child Language Data Exchange System (CHILDES). Another approach would be to examine when and how often children are presented with such terms in their daily lives and everyday activities. For instance, the childLEX database (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015) contains corpus data from children's books intended for different age groups. Researchers could track which terms the books targeting different age groups contain, and use this information as a proxy for evaluating the prevalence with which children of different ages encounter different phrases. Such analyses would also provide traction for relating the understanding of everyday uncertainty terms with language development more generally. For instance, the vocabulary size in a German sample has been estimated to increase from approximately 6000 lemmas in first grade to about 73,000 lemmas in young adulthood (Segbers & Schroeder, 2017). Given the plethora of natural language expressions denoting degrees of probability and frequency, it would be particularly interesting to assess whether the growth of the mental lexicon of such terms is characterized by similar trends and growth rates, thereby helping to gain a more comprehensive understanding of how language development more generally is related to developing adult-like intuitions about the meaning of everyday uncertainty terms.

A related issue concerns conversational pragmatics and social function (Collins & Hahn, 2018; Harris, Corner, & Hahn, 2013a; Honda & Yamagishi, 2017; Juanchich, Teigen, & Villejoubert, 2010; Piercey, 2009). For instance, Bonnefon and Villejoubert (2006) found that the numerical estimates assigned to the word "possibly" in a statement such as "The doctor tells you, you will *possibly* suffer from insomnia soon" differed depending on the assumed social function. Listeners assigned lower estimates when they assumed the goal was to communicate a degree of likelihood, compared to when they assumed that the expression was used to communicate bad news in a tactful manner. Another example is that verbal uncertainty statements can be directional in that they draw attention to either the occurrence or nonoccurrence of an outcome (Teigen & Brun, 1995). For instance, when participants were told that there is "some possibility" that a treatment would be helpful, they were more likely to recommend it than when it was "quite uncertain" that the treatment would help (Teigen & Brun, 1999). Importantly, this asymmetry held even when similar numerical equivalents were assigned to the two phrases, showing how additional factors mediate people's understanding of and reasoning with verbal uncertainty terms. Interestingly, children around age 9 have been shown to be less influenced by the directionality of verbal expressions (Gourdon & Villejoubert, 2009). Given that the quantitative estimates of 9-year-olds in our study already seem to be fairly similar to those of adults, this might indicate that children's sensitivity to conversational rules and pragmatics matures later in development.

The present findings also provide leverage for evaluating and improving the communication of risk information to children and adolescents, which often relies on verbal phrases (e.g., "car drivers are *likely* to not see you because you're small"). Using verbal terms is a natural choice in educational campaigns targeting children, because their ability to understand and reason with numbers is typically less developed than that of adults. Careful investigation of how they understand such terms is critical to develop effective educational materials based on codified language that is tailored to the specific groups targeted. For instance, while in our study most children from around age 8 onward clearly understood the meaning of the term "half of the cases" and assigned a value of 50 to this term, younger children's judgments were much more variable, indicating that they had not yet developed a precise meaning of this term (which might be tied to having acquired basic math skills and understanding of proportions). In this context, it is also important to take into account possible cross-language differences in the understanding of verbal uncertainty expressions (Doupnik & Richter, 2003). For instance, codification schemes such as the one used by the International Panel on Climate Change are typically translated into different languages, but the numerical estimates that people intuitively assign to them have been shown to differ across languages and cultures (Harris et al., 2013b). Conducting comparisons on children's understanding of verbal probability phrases across different cultures, communities, and languages can provide additional insights and help to take into account these differences in applied settings. A related issue concerns children's understanding of quantifiers such as *very* and *highly*, which are frequently used in tandem with different uncertainty terms (Cliff, 1959; Mosteller & Youtz, 1990; Teigen, 1988). This issue is important as such modifiers are also used in many codification schemes (e.g., the guidelines of the International Panel on Climate Change use the modifier "very" for the root terms "likely" and "unlikely"). Investigating how such modifiers influence children's understanding in comparison to adults will provide additional insights into cognitive development and the understanding of everyday uncertainty terms.

Finally, an important venue for future research is to evaluate the understanding of everyday uncertainty terms over the whole life span, assessing whether and how the interpretation of verbal expressions changes in later adulthood and as a function of experience or changes in cognitive ability. This is of particular importance as many societies face fundamental demographic changes due to increased longevity and decreasing birth rates, resulting in substantial shifts in the population structure. Building artificial systems and digital personal assistants that interact with and aid the elderly in their everyday activities plays a critical role in keeping pace with these changes and the ever-increasing demand for support and care. Our and related findings can support these developments by helping to devise calibrated language and codification systems grounded in empirical research on how people intuitively understand such terms. This is of particular importance because a mismatch between the intended and perceived meaning of verbal probabilities can have serious implications. For instance, Berry et al. (2002) investigated how people understand different frequency terms recommended by the European Union to communicate the side effects of medical treatments (e.g., "common," "rare"). The numerical equivalents assigned to these terms were much higher than intended, leading to an increased judgments of the expected severity of the side effects and reduced intention to comply (compared to a control group that received numerical information). Such findings are critical for real-world applications designed to support everyday activities. For instance, if a person inquires about possible side effects of a medication, a digital assistant should not merely provide the verbal labels used in medicine information leaflets, but additionally provide the relevant numerical information to avoid a mismatch between the intended and perceived meaning (see Jenkins, Harris, & Lark, 2018, for a detailed analysis). Ideally, such systems and their natural language processing capacities could be tailored to individual users or target populations, based on empirical data on how the interpretation of everyday uncertainty terms develops over the life span and how particular age groups interpret them. Ultimately, this could also help to improve artificial systems' ability to use and represent common-sense knowledge.

In sum, verbal expressions of uncertainty form an integral part of everyday communication and activities. Thus, from both an applied and basic science view, understanding how people—children and adolescents, young and elderly adults—interpret verbal uncertainty terms is paramount to developing a comprehensive theory of everyday reasoning and activity.

Open Research Badges

This article has earned Open Data. Data are available at https://osf.io/g2c6x.

Acknowledgments

We thank Federico Meini for programming the experiment, and Nadine Fleischhut, Leon Gellrich, Lola Herrmann, and Angela Jones for helpful comments and feedback. We thank Rebecca Maria dos Santos and Cäcilie Hildebrandt for collecting the data, and we are grateful to the Museum für Naturkunde and FEZ in Berlin for supporting our research and providing space for running the study. Data and R code for reproducing the analyses and plots is available at https://osf.io/g2c6x/.

Open access funding enabled and organized by Projekt DEAL.

References

Berry, D. C., Knapp, P. R., & Raynor, T. (2002). Is 15 per cent very common? Informing people about the risks of medication side effects. *International Journal of Pharmacy Practice*, *10*(3), 145–151.

Biehl, M., & Halpern-Felsher, B. L. (2001). Adolescents' and adults' understanding of probability expressions. *Journal of Adolescent Health*, *28*(1), 30–35.

Bocklisch, F., Bocklisch, S. F., & Krems, J. F. (2012). Sometimes, often, and always: Exploring the vague meanings of frequency expressions. *Behavior Research Methods*, *44*(1), 144–157.

Bonnefon, J.-F.& Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, *17*(9), 747–751.

Brun, W., & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, *41*(3), 390–404.

Budescu, D. V., Broomell, S., & Por, H.-H. (2009). Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science*, *20*(3), 299–308.

Budescu, D. V., Por, H.-H., Broomell, S. B., & Smithson, M. (2014). The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change*, *4*(6), 508–512.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, *36*(3), 391–405.

Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

Chambers, C. T. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology*, *27*(1), 27–36.

Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, *31*(2), 131–145.

Clark, D. A. (1990). Verbal uncertainty expressions: A critical review of two decades of research. *Current Psychology*, *9*(3), 203–235.

Cliff, N. (1959). Adverbs as multipliers. *Psychological Review*, *66*(1), 27–44.

Collins, P. J., & Hahn, U. (2018). Communicating and reasoning with verbal probability expressions. In K. D. Federmeier (Ed.), *Psychology of Learning and Motivation* (Vol. *69*, pp. 67–105). Cambridge, MA: Academic Press.

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, *49*(2), 243.

Dhami, M. K., & Wallsten, T. S. (2005). Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities. *Memory & Cognition*, *33*(6), 1057–1068.

Doupnik, T. S., & Richter, M. (2003). Interpretation of uncertainty expressions: A cross-national study. *Accounting, Organizations and Society*, *28*(1), 15–35.

Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child development: cognitive development* (Vol. *1*, pp. 167–230). Hoboken, NJ: John Wiley & Sons.

Gourdon, A., & Villejoubert, G. (2009). A comparison of children and adults' judgements and decisions based on verbal uncertainty statements. In *Proceedings of the annual meeting of the Cognitive Science Society*, Amsterdam (pp. 1121–1126).

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*(20), 9066–9071.

Hamm, R. M. (1991). Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes*, *48*(2), 193–223.

Harris, A. J., & Corner, A. (2011). Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1571–1578.

Harris, A. J., Corner, A., & Hahn, U. (2013a). James is polite and punctual (and useless): A Bayesian formalisation of faint praise. *Thinking & Reasoning*, *19*(3-4), 414–429.

Harris, A. J., Corner, A., Xu, J., & Du, X. (2013b). Lost in translation? Interpretations of the probability phrases used by the Intergovernmental Panel on Climate Change in China and the UK. *Climatic Change*, *121*(2), 415–425.

Hoffner, C., Cantor, J., & Badzinski, D. M. (1990). Children's understanding of adverbs denoting degree of likelihood. *Journal of Child Language*, *17*(1), 217–231.

Honda, H., & Yamagishi, K. (2017). Communicative functions of directional verbal probabilities: Speaker's choice, listener's inference, and reference points. *Quarterly Journal of Experimental Psychology*, *70*(10), 2141–2158.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures*. New York: Basic Books.

Jenkins, S. C., Harris, A. J., & Lark, R. (2018). Understanding "unlikely (20% likelihood)" or "20% likelihood (unlikely)" outcomes: The robustness of the extremity effect. *Journal of Behavioral Decision Making*, *31*(4), 572–586.

Juanchich, M., Teigen, K. H., & Villejoubert, G. (2010). Is guilt "likely" or "not certain"? Contrast with previous probabilities determines choice of verbal terms. *Acta Psychologica*, *135*(3), 267–277.

Kuczaj, S. A. (1975). On the acquisition of a semantic system. *Journal of Verbal Learning and Verbal Behavior*, *14*(4), 340–358.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*(1), 79–86.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*(8), 1134–1140.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, *9*(10), 563–564.

Light, C. S., Zax, M., & Gardiner, D. H. (1965). Relationship of age, sex, and intelligence level to extreme response style. *Journal of Personality and Social Psychology*, *2*(6), 907–909.

Martire, K. A., Kemp, R., Sayle, M., & Newell, B. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*, *240*, 61–68.

Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Edenhofer, O., Stocker, T. F., Field, C. B., Ebi, K. L., & Matschoss, P. R. (2011). The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Climatic Change*, *108*(4), 675–691.

Meder, B., & Mayrhofer, R. (2017). Diagnostic causal reasoning with verbal information. *Cognitive Psychology*, *96*, 54–84.

Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science*, *5*(1), 2–12.

Mullet, E., & Rivet, I. (1991). Comprehension of verbal probability expressions in children and adolescents. *Language & Communication*, *11*(3), 217–225.

O'Dowd, S. C. (1980). Semantic differential studies of children's language development: Some methodological problems. *Journal of Experimental Child Psychology*, *29*(1), 36–49.

Piercey, M. D. (2009). Motivated reasoning and verbal vs. numerical probability assessment: Evidence from an accounting context. *Organizational Behavior and Human Decision Processes*, *108*(2), 330–341.

Rapoport, A., Wallsten, T. S., Erev, I., & Cohen, B. L. (1990). Revision of opinion with verbally and numerically expressed uncertainties. *Acta Psychologica*, *74*(1), 61–79.

Reagan, R. T., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, *74*(3), 433–442.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.

Ruggeri, A., Vagharchakian, L., & Xu, F. (2018). Icon arrays help younger children's proportional reasoning. *British Journal of Developmental Psychology*, *36*(2), 313–333.

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childlex: A lexical database of German read by children. *Behavior Research Methods*, *47*(4), 1085–1094.

Segbers, J.& Schroeder, S. (2017). How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing*, *34*(3), 297–320.

Simpson, R. H. (1944). The specific meanings of certain terms indicating differing degrees of frequency. *Quarterly Journal of Speech*, pp. 328–330.

Simpson, R. H. (1963). Stability in meanings for quantitative terms: A comparison over 20 years. *Quarterly Journal of Speech*, *49*(2), 146–151.

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, *68*(1-3), 27–38.

Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica*, *88*(3), 233–258.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, *80*(2), 155–190.

Teigen, K. H., & Brun, W. (2003). Verbal expressions of uncertainty and probability. In D. Hardman (Ed.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 125–145). New York, NJ: Wiley.

Thompson, W. C., & Newman, E. J. (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law and Human Behavior*, *39*(4), 332–349.

Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *Knowledge Engineering Review*, *10*(1), 43–62.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986a). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, *115*(4), 348–365.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, *39*(2), 176–190.

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986b). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, *25*(5), 571–587.

Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 781–789.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

Xu, F.& Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences*, *8*(3), 199–249.

Zaman, B., Abeele, V. V., & De Grooff, D. (2013). Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction*, *1*(2), 61–70.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab of this article:

Developmental Trajectories in the Understanding of Everyday Uncertainty Terms – Supplementary Material

APPENDIX

A.1 Probabilistic modeling of everyday uncertainty terms

The beta distribution is defined on the interval [0,1] and is parameterized by two (positive) parameters, α and β, which determine its shape. We fitted individual beta distributions to participants' numerical estimates (mapped on the interval [0,1] by dividing each judgment by 100), using the method of moments to derive the shape parameters α and β separately for each verbal term from the age group's sample mean $\hat{x}$ and sample variance s^2. The estimate for shape parameter α, $\hat{\alpha}$, is given by

$$\hat{\alpha} = \hat{x}\left(\frac{\hat{x}(1-\hat{x})}{s^2} - 1\right) \tag{A1}$$

and the estimate for shape parameter β, $\hat{\beta}$, is given by

$$\hat{\beta} = (1-\hat{x})\left(\frac{\hat{x}(1-\hat{x})}{s^2} - 1\right). \tag{A2}$$

For each term, the corresponding beta distribution has the same mean and variance as the empirical distribution of judgments. Numerical values of all shape parameters can be found at https://osf.io/g2c6x/; visualizations of the fitted in distributions for each term across age groups can be found in the Supporting Information and OSF repository.

Note that the terms *half of the cases* and *equiprobable* are special in the sense that they have a clear-cut definition, such that both of them should correspond to the midpoint (i.e., 50 out of 100) of the used slider. As of age 8 almost all participants gave this judgment, such that the variance for *half of the cases* was very small, resulting in high values for the shape parameters (see Eqs. A1 and A2). A similar pattern was obtained for the term *equiprobable*. For 11- to 14-year-olds the variance was in fact zero, as all subjects gave a judgment of 50. Therefore, we set $\hat{\alpha} = \hat{\beta} = 10^5$, such that virtually all density of the beta distribution is located at 0.5 and its variance is minimal. Furthermore, note that adults' shape parameters for this term were also very high (>50,000), as almost all of subjects assigned a value of 50 to this term, rendering the variance minimal.

A.2Using Kullback–Leibler divergence to assess developmental trends

Given the fitted beta distributions, we can use the Kullback–Leibler (KL) divergence as a measure of the similarity of two probability distributions P and Q (Kullback & Leibler, 1951). Here, we used the KL divergence to assess how similar each of the children's distribution P is to the corresponding distribution Q derived from adult participants. For continuous probability distributions, KL divergence is defined as

$$D_{KL}(P\|Q) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right), \tag{A3}$$

where $p(x)$ and $q(x)$ denote the densities of distributions P and Q. It can be solved analytically when P and Q are beta distributions with parameters (α_p, β_p) and (α_q, β_q):

$$\begin{aligned} D_{KL}(P\|Q) = {} & \ln\left(\frac{\Gamma(\alpha_p+\beta_p)}{\Gamma(\alpha_p)+\Gamma(\beta_p)}\right) - \ln\left(\frac{\Gamma(\alpha_q+\beta_q)}{\Gamma(\alpha_q)+\Gamma(\beta_q)}\right) \\ & + \left(\alpha_p - \alpha_q\right)\left(\Psi(\alpha_p) - \Psi(\alpha_p+\beta_p)\right) \\ & + \left(\beta_p - \beta_q\right)\left(\Psi(\beta_p) - \Psi(\alpha_p+\beta_p)\right), \end{aligned} \tag{A4}$$

where Γ and Ψ denote the gamma and digamma functions, respectively.

For our analyses, adults' beta distributions for the different terms serve as reference distributions Q for each of the distributions P derived for the corresponding distribution obtained for the children. Thus, for each term and age group the obtained KL divergence quantifies how similar children's distribution is to adults' distribution. Fig. 6 shows the distribution of KL divergences across the terms and age groups, illustrating how children's distributions become more similar to those of adult subjects as they grow older. We evaluated the terms *half of the cases* and *equiprobable* separately, as the variances for these terms was minimal or zero (see above). Accordingly, the KL divergences for terms *half of the cases* and *equiprobable* are much higher than for the other terms; therefore we plot them separately (Fig. 6b and c).