International Journal of Approximate Reasoning 57 (2015) 19–39

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Natural language of uncertainty: numeric hedge words

Scott Ferson [a,*], Jason O'Rawe [a,b], Andrei Antonenko [a,c], Jack Siegrist [a], James Mickley [a,d], Christian C. Luhmann [e], Kari Sentz [f], Adam M. Finkel [g]

[a] Applied Biomathematics, 100 North Country Road, Setauket, NY 11733, USA
[b] Genetics Program, Stony Brook University, 100 Nicolls Road, Stony Brook, NY 11794, USA
[c] Linguistics Department, Stony Brook University, Stony Brook, NY 11794, USA
[d] Ecology & Evolutionary Biology, The University of Connecticut, 75 N. Eagleville Road, Unit 3043, Storrs, CT 06269-3043, USA
[e] Psychology Department, Stony Brook University, Stony Brook, NY 11794, USA
[f] Los Alamos National Laboratory, Post Office Box 1663, MS F609, Los Alamos, NM 87545, USA
[g] The University of Pennsylvania Law School, 3501 Sansom Street, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 1 January 2014
Received in revised form 2 August 2014
Accepted 10 November 2014
Available online 14 November 2014

Keywords:
Approximator
Linguistic expression of uncertainty
Hedge
Amazon Mechanical Turk
Elicitation
Uncertainty communication

ABSTRACT

An important part of processing elicited numerical inputs is an ability to quantitatively decode natural-language words that are commonly used to express or modify numerical values. These include 'about', 'around', 'almost', 'exactly', 'nearly', 'below', 'at least', 'order of', etc., which are collectively known as approximators or numerical hedges. Figuring out the quantitative implications of these expressions for the uncertainty of numerical quantities is important for being able to understand, for example, what is actually being reported by a patient who says a headache has lasted for "about 7 days", and how we should translate the patient's report into uncertainty about the duration. We used Amazon Mechanical Turk to empirically identify the implications of various approximators common in English. To evaluate the numerical range implied by each approximator, we analyzed paired statements differing only in the approximator used in numerical expressions. Despite often considerable diversity, there were several statistically significant findings, but far less quantitative variation implied by the approximators than might have been expected. The numerical implication of different approximators interacts with the magnitude and roundness of the nominal quantity. This investigation strategy generalizes easily to languages other than English.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The innate numerical acuity of humans is remarkably poor [16]. Although recognizing and thinking with numbers may involve multiple cognitive systems that are not yet fully understood (cf. [11]), laboratory and field observations show that without tutoring people typically have a number sense that can distinguish only up to about four items [29]. Human societies that have not developed number systems or at least finger counting have difficulty discerning the quantity four [26], and humans seem to innately distinguish only the quantities one, two and many, which represents any quantity more than two.

* Corresponding author. Tel.: +1 631 751 4350.
E-mail address: sandp8@gmail.com (S. Ferson).

http://dx.doi.org/10.1016/j.ijar.2014.11.003
0888-613X/© 2014 Elsevier Inc. All rights reserved.

Number systems were invented repeatedly in human history [30], apparently to facilitate commerce, to bring clarity to ideas otherwise expressed by words like 'some', 'many', 'more', 'less' and 'fewer'. These systems permit the description of quantities with expressions consisting of a numeral and units. The units specify the scale of measurement, which is either a count noun (e.g., dollars, days, chairs, bushels, people, acres), or what linguists call a measure word used with a mass noun (e.g., kernels of corn, bushels of wheat, liters of water, tanks of gas, rooms of furniture). The numeral represents an integral count or real-valued measurement revealing the multiplicity or fractionality of the unit equivalent to the quantity being described. The numeral expresses a magnitude, possibly spelled out in words ('one', 'two', 'sixty-eight', 'three quarters', etc.) or expressed with numerical digits ('1', '256', '0.5', etc.).

The clarity of number systems often implies greater precision than is practically achievable in many situations. This fact requires some scheme to relax or discount this precision. In linguistics, a *hedge* is a word or phrase that modifies the force or precision of ideas or statements [37]. Hedges serve several purposes in language, including expressing uncertainty or transience, stipulation, responsibility focusing, and obfuscation. Prince et al. [50] recognized two kinds of hedges: shields and approximators. Shields, such as 'I think' or 'probably' modify propositions, whereas approximators modify numerals to alter the magnitude or precision implied by the expression. The latter function of approximators is to convey that the quantity is either less precise or more precise than the meaning of the corresponding numerical quantity without the hedge word. For instance, the sentence 'About 105 people came to the party' may mean that any number of people between 100 and 110 came to the party. In contrast, the sentence '105 people came to the party' has a smaller range of possible values for the implied number of people attending the party. In this case, the approximator 'about' introduces more uncertainty into the interpretation of the sentence.

Sometimes uncertainty is implicit in a numerical expression because of the roundness of the number even though no explicit hedge words may be present at all. For example, a reasonable interpretation of the phrase '1000 people came to the protest' would infer that the number of people who attended the protest is somewhere in the neighborhood of 1000, but not necessarily exactly 1000. Comparing that example to a phrase 'Exactly 1000 people came to the protest', one can see that the hedge 'exactly' reduces the uncertainty of the statement: the latter example means that there were exactly 1000 people at the protest, no more, no less.

In English, quantities are described with expressions generally involving three elements: *unit*, *numeral*, and *approximator*. Grammatically, the approximator is an adverb that modifies the numeral which is an adjective which in turn modifies the unit which is a noun. The order in which the three elements appear is not fixed in English. For example, the written phrase '$100 or so' is unit–numeral–approximator, but 'nearly 5 pounds' is approximator–numeral–unit, and '35 years or more' is numeral–unit–approximator. Sometimes elements may be elided when context or convention allows. The phrase 'three coffees' omits the unit (measure word) 'cup'. Mathematicians discuss abstract quantities which are pure, dimensionless numbers without units. The idea is not so much that there are no units, but that the numbers represent quantities with *any* units. When the numeral is omitted, it is usually understood to be one, unless context forces another value. Omitting the approximator element—using what we might call the null hedge or *null approximator*—does not usually mean there is no imprecision whatever about the quantity. Instead, the value is understood to have a precision implicitly encoded in the roundness of the number, the discourse environment (e.g., bank statements versus barroom braggadocio), and measurability of the quantity.

There are many approximators in English, including generic hedges such as 'around' and 'nearly', archaic hedges such as 'well-nigh', and idiomatic constructions such as 'in the ballpark of'. Some hedges generally appear before the numeral like 'around' and 'as many as', and some generally appear after the numeral like 'or so' and 'and change'. Some approximators can appear either before or after the numeral like 'approximately', 'almost' and 'at least'. Table 1 lists many approximators in wide use which are distinguished into four categories. Channell [14] asserted that all of the approximators imply a range of possible values for the quantity being described. Sometimes this interval is explicitly indicated with ranging constructions like '5 or 6' and '15 or 20' and 'between 86 and 94', but many hedged numerical expressions refer to a single exemplar number, about which the interval of imprecision is understood to be symmetrically or asymmetrically positioned around this value. For example, 'around 5' is symmetric, whereas 'more than 5' is asymmetric. The null approximator is in a category by itself.

Sadock [55] argued that approximators but also many other factors affect the implied imprecision about a quantity. From introspective linguistic analysis of pairs of natural-language expressions such as

1 million	990 000
about 1 million	1 million
about 990 000	990 000
about 1 million	about 990 000
about a dozen	about 12
about two and a half	about 2.5
six-foot insect	six-foot man

where the value on the left is understood to be less precise than the value on the right, he concluded that the roundness of the number mentioned, its display format, the possible range of the quantity, the relevant standards of precision, and the units themselves including whether they refer to discrete entities or mass nouns all affect the implied imprecision about

Table 1
Linguistic approximators of various kinds.

Null	Symmetric	Asymmetric	Ranging
[]	about	above	at least x but no larger than y
	above	almost	between x and y
	almost	all but	between x and y, inclusive
	approximately	and change	x or $x+1$
	around	and some	ax or $a(x+1)$
	as good as	as high as	x to y
	a total of	as low as	within x (units) of y
	ballpark	as many as	etc.
	bordering on	at least	
	close to	at most	
	essentially	below	
	exactly	bordering on	
	for all practical purposes	close to	
	for the greatest part	down to	
	give or take	fewer than	
	in effect	less than	
	in the ballpark of	smaller than	
	in the neighborhood of	more than	
	in the vicinity of	no(t) fewer than	
	just about	no greater than	
	more or less	no larger than	
	near	no less than	
	near to	no more than	
	nearly	no smaller than	
	nigh	not quite	
	not far from	on the brink of	
	on the order of	on the edge of	
	or so	on the point of	
	or thereabout(s)	on the verge of	
	order	over	
	practically	up to	
	precisely	virtually	
	pretty much	etc.	
	pretty near(ly)		
	roughly		
	round about		
	virtually		
	well-nigh		
	within sight of		
	etc.		

a quantity. Sadock [55] was so vexed by the effects of context that he concluded that linguistic vagueness could not be effectively decoded into rules governing the meanings of approximators.

Speakers and writers use approximators frequently, especially when they are being careful and trying to be truthful [50]. They arise commonly in expert elicitations, medical case histories and patient complaints recorded by health care professionals, text analyses, and legal or linguistic interpretations of discourse and testimony. It seems intuitively clear that using an inappropriate approximator can convert a truthful or correct statement into an untruthful or incorrect one. Thus, to wring from natural-language inputs the full import that they convey without misreading those inputs, it is essential that we understand these approximators quantitatively so we can properly interpret elicited numerical values. Decoding approximators provides a key insight into the uncertainty expressed through natural language and its quantitative analysis benefits elicitation, communication, reasoning, and inferences based on hedged numerical values. Analysts of information expressed in natural language need to be able to interpret approximators quantitatively to appreciate the uncertainties expressed in speech and text. In this paper, we empirically quantify the numerical meanings of several common approximators, and describe their variations among individuals.

It would be advantageous to use approximators to encode the uncertainties estimated by risk analyses and similar numerical calculations into natural language for use in risk communication to humans. Scientists and engineers have long been taught to use only significant digits in reporting their numerical results so as to not inadvertently imply more precision about a conclusion than is warranted [57,72], and they may assume that decision makers implicitly understand this convention. Yet we know that the significant-digits convention, in which for example '1.23' is interpreted as the interval range $[1.225, 1.235]$, is inadequate to express large uncertainties. There is no single number that can express the uncertainty of $[10, 16]$, nor even the narrower uncertainty of $[14.8, 15.6]$. Fig. 1 depicts a part of the half-plane of all possible real intervals [*Left*, *Right*]. Every possible interval range corresponds to a point on this plane above the forty-five-degree line. But only the corners of the triangles depicted in the figure can be exactly represented by a scalar number under the significant-digits convention, and only interval ranges corresponding to points inside these triangles can even be enclosed

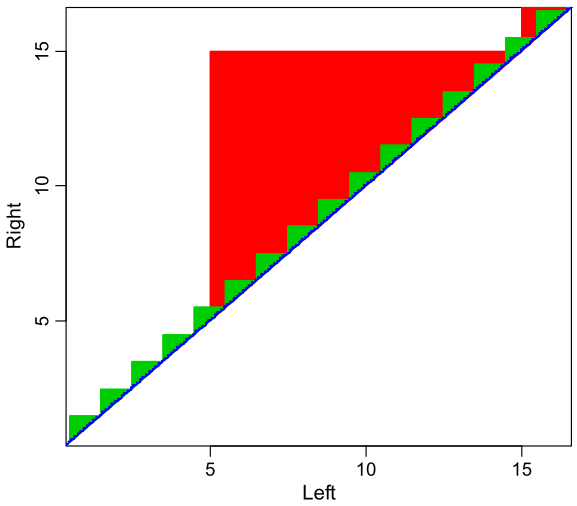

Fig. 1. Intervals of the form [*Left*, *Right*] that can be represented by a scalar number under the significant-digit convention (corresponding to the top-left corners of each triangle) and intervals that can be enclosed under that convention (corresponding to the triangles) on the half-plane of all real intervals. Different sized triangles represent different numbers of significant digits of the scalar number.

by the intervals implied by scalar numbers under the significant-digits convention. Because the results of risk analyses can often result in uncertainties larger than can be expressed under this convention, scientists and engineers perhaps should also use approximators to form verbal characterizations that express the uncertainty of their numerical estimations if these can better or more naturally be understood by their audience, or at least by linguistically competent speakers of English in that audience. Of course risk communicators, and expositors of scientific and technical information generally, already do use approximators in their explanations, but it is not clear that they use them in the best or most robust way. Quantitative study of approximators and their implications may help to fashion guidance for more effective risk communication that is less prone to misunderstanding.

One way to develop a system for decoding and encoding uncertainty from and into approximators is an Aristotelian prescriptive approach in which scientists adopt technical meanings for some hedge words and turn them into jargon. This approach would simply assign quantitative meanings to various approximators based on ideas and rules conceived by experts or conventions. Such a prescriptive approach could be made to be internally consistent and to have properties that make the resulting system most useful in practice. For instance, a designed system could ensure that all uncertainty ranges can be conveniently represented in the system. Such completeness cannot be guaranteed if we restrict ourselves to the available approximators in natural language. In fact, various schemes applying technical meanings to English uncertainty words have been proposed by scientists several times in the past. For example, the Intergovernmental Panel on Climate Change recently defined ‘very likely’ to mean having probability between 90% and 99% [61].

Astronomers have likewise developed a scale converting between quantitative probability for impacts of near-Earth orbiting asteroids and English expressions [7] in which, interestingly, risks with 100-year cumulative probability lower than 10^{-8} have a likelihood of collision optimistically characterized as ‘none’. Weiss [66] suggested extending legal definitions of standards of proof to characterize scientific statements according to eleven levels of certainty. Kent [33] suggested formalizing the meanings of hedges used in intelligence briefings to prevent misunderstanding of “poetic” language such as ‘serious possibility of an invasion’. His scheme identified ranges of probabilities with English expressions:

100% certain

$(93 \pm \sim 6)\%$ almost certain

$(75 \pm \sim 12)\%$ probable

$(50 \pm \sim 10)\%$ chances about even

$(30 \pm \sim 10)\%$ probably not

$(7 \pm \sim 5)\%$ almost certainly not

0% impossible

This scheme was not complete in that, for example, a probability of 15% was not covered by any phrase, and it was criticized by his contemporaries as an imposition of bogus precision onto language. Wallsten et al. [65] revisited the quantification of vague probability terms using a fuzzy-sets approach.

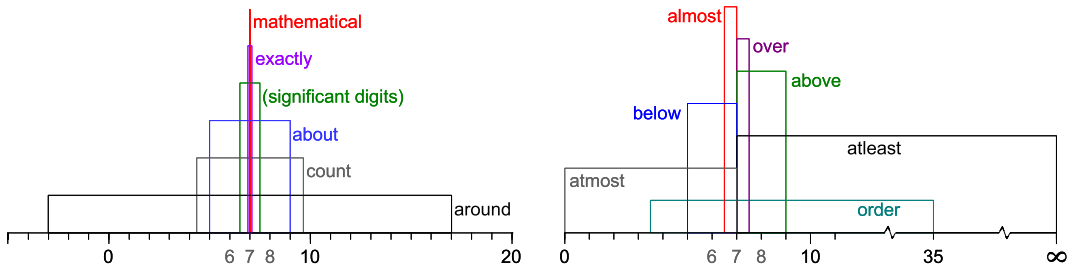

Fig. 2. Symmetric (left) and asymmetric (right) approximators of the number 7.

In fact, there have been innumerable other Aristotelian schemes to quantify the uncertainty expressions of natural language (e.g., [8,17]), but these definition systems are almost never broadly accepted even within a discipline, much less beyond a narrow technical field. There are two notable exceptions. The World Conservation Union specifies a "critically endangered" species to be one with a probability of extinction of at least 50% within 10 years, and a "vulnerable" species to have a 10% probability of extinction within 100 years [41]. The most famous exception was Ronald Fisher's suggestion that statistical findings in hypothesis testing with probability of 5% or less be called "statistically significant". Such schemes are usually created for the sake of convenience, but they can be especially useful in legal and regulatory settings, because they translate quantitative results into qualitative categories. These systems have disadvantages, including susceptibility to the sorites paradox, and suboptimal decisions in contexts where gradation or vagueness is misrepresented by bright lines [69,24,60].

As far as we know, no one has proposed an articulated system to quantitatively define approximators, but it is easy to imagine one. Indeed, there are many ways to construct one based on magnitude of the exemplar number or its number of significant digits. For instance, consider the scheme illustrated in Fig. 2 (different heights are used only to disambiguate the several ranges but are otherwise meaningless). This system of decodings reflects a belief that all numerical quantities in the real world have uncertainties. The default interpretation of an explicit scalar number x with the null approximator is the interval implied by the decimal place of its least significant digit. So the expression '7' is interpreted as $[6.5, 7.5]$. The 'about' approximator quadruples the width of this interval, and the 'around' qualifier increases the width by a factor of twenty. Even the phrase 'exactly 7' is interpreted as $[6.9, 7.1]$, which is something less precise than '7.0' with no approximator, which would be interpreted as $[6.95, 7.05]$ under the significant-digit convention. This system requires the new approximator 'mathematical' to describe numbers that are precise to infinitely many decimal places, as would be needed for the 2 in a square function. The definitions of these approximators are given in the table below, where d is the decimal place of the last significant digits of x. So, for instance, if x is 7, d is zero. If x is 7.0, d is 1, and if x is 700, d is -2. Nonsignificant digits are stripped away from x before the approximator function is applied.

Hedged numerical expression	**Interpretation**
mathematical x	x
exactly x	$x \pm 10^{-(d+1)}$
x	$x \pm 0.5 \times 10^{-d}$
about x	$x \pm 2 \times 10^{-d}$
around x	$x \pm 10 \times 10^{-d}$
count x	$x \pm \sqrt{x}$
almost x	$[x - 0.5 \times 10^{-d}, x]$
over x	$[x, x + 0.5 \times 10^{-d}]$
below x	$[x - 2 \times 10^{-d}, x]$
above x	$[x, x + 2 \times 10^{-d}]$
at most x	$[0, x]$, $or [-\infty, x]$
at least x	$[x, \infty]$
order x	$[x/2, 5x]$
between x and y	$[x, y]$

Note that some hedges are based on the number and position of significant digits used to express the quantity, while some are based solely on the magnitude of the quantity. The system is complete in the sense of being able to represent any given interval range because it supports the 'between' approximator. The characterizations can be compounded so that, for instance, 'at most about 300' would be the interval $[0, 350]$. The system can be extended in other obvious ways to handle points in time and temporal spans with approximators like 'since', 'until', '-ish', and 'on or about', although the use of the sexagesimal number system and the modulo-twelve clock scale also makes time rather different from quantities expressed in a decimal system on a linear scale with respect to how uncertainty is typically encoded.

Whatever the possible advantages of such an artificial system, because it would be imposed by fiat, it could only at best achieve the status of jargon. The wide adoption of this or any system originating in scientific convention or perhaps

government regulation would require buy-in and on-going broad education among would-be users to be successful. Of course, we have little reason to believe that such a prescriptive system such as that proposed in Fig. 2 is the best or most appropriate system that could be devised for interpreting and forming numerical expressions in English. There are alternative interpretations and various possible decoding schemes. Even among the authors of this paper, multiple competing systems were conceived and championed. Long-standing scientific convention about interpreting the uncertainty from a naked scalar number is not universally observed. For example, Schulte et al. [56] suggest that the unqualified expression '2.31' should be interpreted as [2.30, 2.32], and some authors argue compellingly against trying to use significant figures to express uncertainty at all (e.g., [18]).

In this paper, we adopt an alternative *descriptive* approach to developing a way to decode and encode approximators that characterize uncertainty in numerical expressions in English [12,65]. In this Galilean approach, the quantitative meanings of various approximators are empirically quantified to reflect how native language users interpret them. What does it actually mean when someone says a phrase like 'about 140'? Is it substantially different from what is implied by the similar phrase 'about 143'? Is it related to what is meant by the phrase 'about 143.26'? Does 'about' denote a tighter or broader range than various other approximators? When their uncertainty is a particular interval, what ballpark scalar number do humans typically select to represent it? Does the behavior change when the uncertainty represents actual variation rather than epistemic uncertainty or ignorance per se? Does the result change when the variability is expressed in time, across space or among individuals or components? We expect a lot of variation among individual respondents about these questions. We also expect, as Sadock [55] argued, that context will often be important. Different natural languages must be studied separately.

Channell [14,13,12] appears to have been the first to try to quantify the implications of approximators. She found that, for almost all her informants, the approximators were interpreted as denoting continuous intervals (or sequential integer ranges) of possible values, invariably including the exemplar number. She collated histograms of the breadths of these intervals for the approximators 'about', 'around', 'or so', 'not less than' and 'x or y'. Her quantitative results suggest there are differences between approximators but also substantial differences between informants, even for a small sample size of 26 informants. The minimum and maximum lengths of these intervals, expressed as percentages of the exemplar number, varied by a factor of about three among her small sample.

We collected some preliminary data from native and non-native speakers of English in the United States about the implication of approximators using traditional direct questionnaires distributed to students (which was also Channell's approach). Unfortunately, this method of interrogation is inefficient and tiresome, and even dizzying for the informant. It turns out to be quite difficult to ask enough people enough questions to nail down a distributional characterization of the quantitative meanings of the hedges. Fixed questions in static questionnaires are also highly susceptible to psychometric artifacts from cognitive biases such as anchoring. Obviously, the questions need to be randomized and the contextual numbers that are modified by the hedges should be varied to obtain general results.

2. Materials and methods: Amazon Mechanical Turk

We used Amazon Mechanical Turk (MTurk) to collect data about the meanings of approximators in English. Amazon Mechanical Turk is an Internet marketplace that allows *requesters* to crowdsource tasks to *workers* (or *turkers*) over the Internet. The tasks are posted by requesters on the Amazon Mechanical Turk Website, Sample tasks which can be commonly found on MTurk include answering questions, tagging images and videos, searching for relevant information on the Internet, etc. In MTurk terminology, tasks are known as human intelligence tasks, or HITs, as they usually require human intelligence, and cannot be accomplished by a computer.

Requesters pay to workers a set fee for performing a HIT upon approving the results of the task. If the result of the HIT is rejected by the requester, the worker receives no compensation, and the task remains incomplete, and can be performed by another worker. Requesters upload the list of questions/tasks to the MTurk website, and workers can choose from the list of various HITs which ones to perform. The requester can specify certain qualifications for workers such as native language, country of residence, minimal approval rate, i.e. the percentage of their completed tasks that were approved by the requester, or the total number of tasks approved.

Reasons for choosing MTurk as a platform for conducting the current experiment include the heterogeneity of the worker population, low cost, high response rate, and ethical and regulatory simplicity. While the population of MTurk is not known to be representative of the population of US or the group of people who speak English natively, it however is more heterogeneous than any group of people which can be recruited to participate in the experiment in academic settings (such as a group of students completing a questionnaire). The cost of recruiting workers on MTurk is low. In a pilot experiment, we paid one cent per question to workers. Thus, for a sample consisting of about 2500 questions, we paid about $25 to workers, plus half of this amount as overhead[1] to Amazon and received all answers in one day.

The size of the MTurk community results in a high response rate to HITs. In our case, answers to about 2500 questions were collected within 24 hours from making HITs available on MTurk. Because the identity of workers is effectively anonymous, and the investigators do not participate in answering the HITs, using MTurk to collect survey/interview data is

[1] The overhead to Amazon is $\max(10\% \times (\text{per-HIT fee}), \$0.005)$ times the number of completed and approved HITs.

exempt from the requirement of prior approval by an institutional review board under human subjects research regulation in the United States (45 CFR §46.101(b)(2) and §46.401(b); 45 CFR §690.101(b)(2); see HHS [28]; NSF [48]).

2.1. Experimental statements

A cache of experimental statements was constructed to control for possible effects of magnitude of the nominal value, its number of significant digits, its units, and other aspects of the context in which an approximator may be used. Over 800 statements containing numerical values most of which were sourced from the Internet website http://facts.randomhistory.com/archives.html which lists sundry historical or popular scientific facts by category. Examples of the original statements are given below, with the numerical expressions highlighted in bold:

Haiti's highest peak is the Pic la Selle at **2680 meters**.
A 14th-century book of Thai poems describes **23 types** of Siamese cats.
Soldiers (hoplites) in ancient Greece wore **up to 70 pounds** of bronze armor.
In 2007, a dog named Rocco discovered a truffle in Tuscany that weighed **3.3 pounds**.
Greece enjoys **more than 250 days** of sunshine a year.
Bats make up **about 20%** of all classified mammal species globally.

The statements to be used in the HITs put to MTurk workers were created by altering these original statements by randomizing both the approximator and the nominal value used in the numerical expression. The approximator, if present, was removed from each statement. In the examples above, the approximators 'up to', 'more than', and 'about' were omitted. A new approximator was then randomly selected from the test set of approximators and inserted into the statement.

The original magnitude specified in each statement was also replaced by a substitute number, randomly selected from a set of numbers comparable in size to the original magnitude, but varying in its apparent precision. The set of possible substitutes was generated using a function written in Python given in Appendix A. This function produces, for each explicit numeral input, a finite list of values that we might consider to be reasonable substitutes for use in the experimental statements so that they vary widely in terms of numbers of significant digits and roundness of the nominal value. The function creates the list of substitute values by, for each digit in the input numeral, replacing that digit with a random digit, a '0', and a '5', and replacing each digit to the right of that one with zeros, to yield three possible alternative values. Numerals with a '5' as the last significant digit are held by some [35] to be rounder than similar numerals with other nonzero digits in this spot. Example results from applying this function to the inputs on the left below generated the random variation and varying number of significant digits in substitute magnitudes on the right.

Input	**Substitute magnitudes**
246	200, 240, 244, 245, 246, 250, 290, 500
1300	1000, 1300, 1305, 1309, 1330, 1350, 1400, 1500, 4000
0.74	0.3, 0.5, 0.7, 0.74, 0.742, 0.745, 0.75, 0.76
23.13	23.0, 23.1, 23.13, 23.1305, 23.1307, 23.135, 23.137, 23.15, 23.17, 23.2, 23.5

A number was chosen randomly from the set generated by the function and this number replaced the original magnitude in the numerical expression in each statement.

Statements were then manually reviewed to catch and remove nonsensical constructions such as '500 days a year' and '120% of people'. But we made no attempt to remove or edit statements with constructions that were merely linguistically implausible. For instance, although natural language speakers may only rarely or perhaps never use a phrase such as 'roughly 10 023' such a phrase would be submitted for interpretation to MTurk workers. For the MTurk experiment, we generated 866 experimental statements containing numerical expressions involving an approximator. To facilitate later statistical comparisons, we strove for a balanced design using the same number of experimental statements for each approximator, and the same number of statements with units from among the following groups:

discrete ('people', 'dogs', 'species', etc.),
money ('$', 'dollars', 'cents'),
length ('meters', 'miles', 'inches', 'feet', 'inches', 'km', 'cm', 'mm', 'm'),
weight ('lbs', 'kg', 'tons', 'ounces', 'pounds', 'oz', 'milligrams'),
time ('years', 'minutes', 'seconds', 'weeks', 'days', 'hours'),
percent ('%'), and
speed ('kph', 'mph').

Several approximators were considered, including the null hedge, 'about', 'above', 'almost', 'approximately', 'around', 'at least', 'at most', 'below', 'exactly', 'nearly', 'no more than', 'no less than', 'over', 'precisely', and 'roughly'. The complete list of experimental statements is available at https://sites.google.com/site/numericalhedging/amazon-mechanical-turk.

Thousands of samples were collected using multiple interrogation formats. In the primary format, workers were asked to evaluate the ranges implied by hedged numerical expressions in paired statements differing only in their approximators. For

each statement, they were asked to provide minimal and maximal possible values for a quantity described by a phrase extracted from the statement. Each phrase consisted of an approximator, numeral and unit. For example, workers were shown

Statement: Roughly 25% of Canadians are Protestant.
Phrase: Roughly 25%

Statement: No more than 25% of Canadians are Protestant.
Phrase: No more than 25%

and asked to provide the minimum and the maximum possible values for the percentage of Protestant Canadians. Each HIT was a bundle of four statement pairs preceded by the following thirteen lines of instructions which did not change from task to task:

What do you think the following statements means? What are the possible values for the phrases below?

- Read the following **statements**.
- **There are in total 4 pairs of statements, 8 statements all together.**
- Please tell us what you think the ***minimal possible values*** for the **phrases** below are.
- Please tell us what you think the ***maximal possible values*** for the **phrases** below are.
- Type in these values in the same format as they are in the phrase, if possible.
- For example (the values given below are just random examples, according to you *the minimum and the maximum could be very different*!)
  - **Statement**: Roughly 100 people came to the meeting.
  - **Phrase**: **roughly 100 people**
  - If you think that "roughly 100 people" means somewhere between 90 and 100, you enter:
  - **Minimum**: 90
  - **Maximum**: 110

Thank you!

Each HIT was answered by 2 different workers before it was retired. Each statement was used in two HITs so as to make it possible to compare the effect of changes to the quantity's magnitude, significant digits, and roundness. There were 422 HITS $\times$ 8 phrases $\times$ 2 workers $=$ 6752 minimum-maximum ranges. The experimental results were obtained within 5 days after uploading the HITs to MTurk.

3. Results

The data downloaded from MTurk were reviewed manually. When a worker gave a maximum that was smaller than the minimum, which happened in only 16 cases, the values were swapped in the data set. We were prepared to interpret missing maximal or minimal values as plus or minus infinity or zero, but almost none of the missing values could reasonably be interpreted this way, and therefore they were omitted from the analysis. Gross numerical outliers representing potential mistakes or misunderstandings were identified, and we conducted parallel analyses with and without these outliers. A result was identified as an outlier when the maximum given by the worker was greater than nine times the exemplar value, or when the minimum was less than one ninth of it. We chose nine as the factor in order to remove cases where the worker may have inadvertently typed an extra or omitted a necessary digit. Such a slip of the fingers would increase or decrease the value by at least ten fold, as might have occurred, for instance, in characterizing the maximum of the interval for 'about 1000' as 11 000 rather than 1100. There were 95 such cases. After data filtering and removing unanswered questions, there were a total of 6038 minimum–maximum ranges.

Statistical and graphical analysis of the results was conducted using the R Environment for Statistical Computing [51]. Analysis of variance (anova) methods were used to discern the effects of approximators on the implied imprecision of the numerical quantity. Regressions were used to uncover the relationship between the uncertainty implied by a hedged numerical expression and various explanatory variables, including the approximator used, and the roundness, magnitude and unit group of the exemplar number mentioned in the numerical expression. The magnitude was characterized in two ways, first as the *actual* value of the exemplar number, and second as its *order* of magnitude taken to be the base-ten logarithm of the exemplar number. We computed both the *number* of significant digits in the exemplar number, and the *order* of significance, which is defined as 10^{-d} where d is the decimal place of the last significant digit in the exemplar number. The number of significant digits in the expression '1.3' is 2, and order of significance is 0.1. Roundness was characterized as the base-ten logarithm of the order of significance. The unit group and approximator were both categorical variables taking values from the sets above of 7 and 16 items respectively. We also included as explanatory variables in the analyses two Boolean variables: the discreteness of the unit (whether the unit group was the discrete category or one of the other 6 measurable categories), and whether the last significant digit was a '5'. Ancillary variables returned by MTurk such as the amount of time in seconds which the worker spent on a given HIT or details about the worker were included in exploratory regression analyses, but were not found to be statistically interesting. The dependent variables in these analyses were measures of the breadth between the maximum and minimum values given by the workers in each case. We considered five output variables, including the range (difference between the maximum and minimum), the base-ten logarithm of the range, the relative range (ratio of the range to the absolute value exemplar number), and the minimum and maximum in units relative to the exemplar number.

Fig. 3 shows the relative breadths of all 6038 reported intervals distributed over 16 approximators. The red curves are empirical distribution functions for the left endpoints of each interval minus and then divided by the magnitude of the

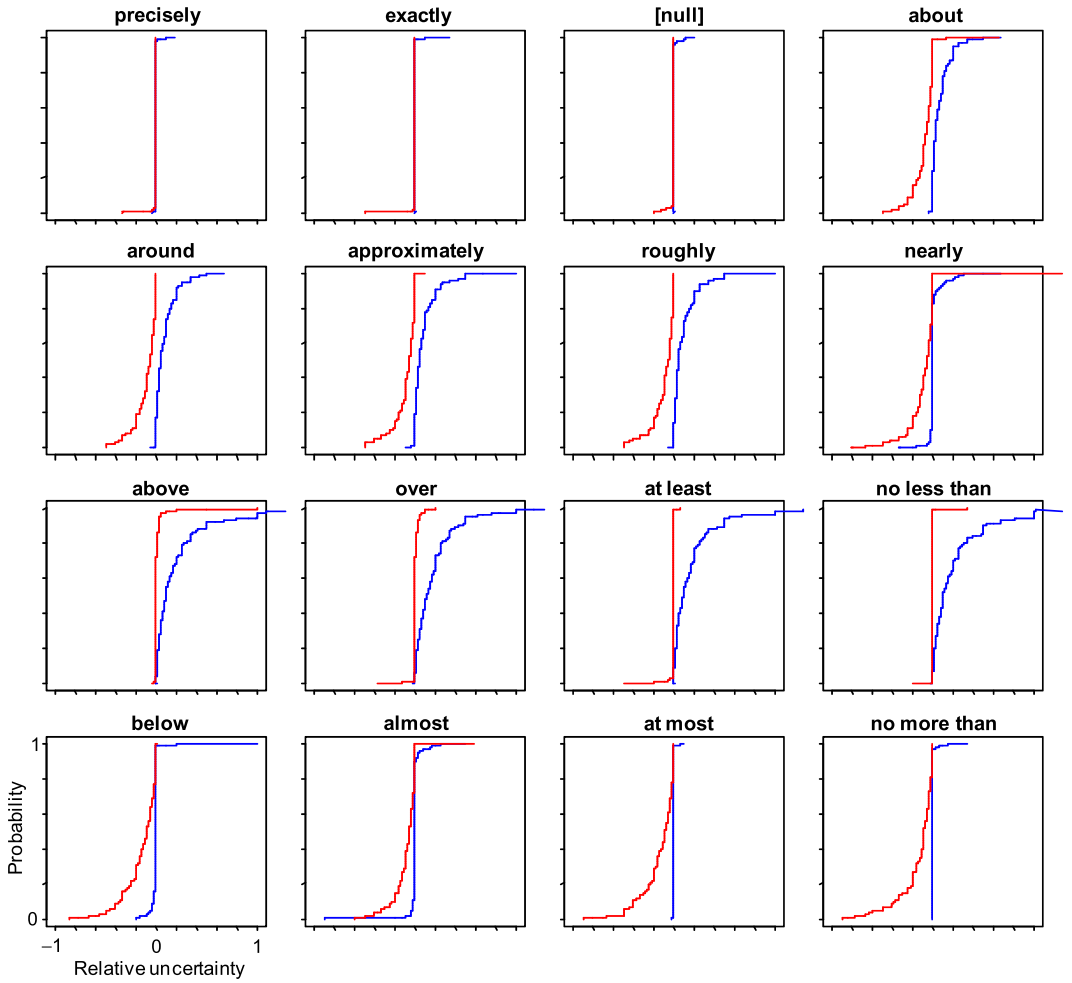

Fig. 3. Relative uncertainty implied by various hedges characterized as empirical distribution functions of $(m-e)/e$, shown as the left distribution in each graph, and $(M-e)/e$, shown as the right distribution in each graph, where m and M are the left and right endpoints of the reported intervals, and e is the magnitude of the exemplar value in each numerical expression.

respective exemplar value. Likewise, the blue curves are the same for the right endpoints. These distribution functions form interval-type bounds on cumulative distribution functions called probability boxes that depict the relative uncertainties associated with the several approximators. These uncertainties could—but only very rarely do—exceed absolute values of one. Of the approximators studied, the tightest intervals are associated with the hedge 'precisely', but those for 'exactly' are nearly as tight. Those for the null hedge are a surprisingly close third. The breadths of 'about', 'around', 'approximately' and 'roughly' are substantially wider, and all quite similar to each other. The remaining approximators are understood to represent asymmetric uncertainties. Note that for several approximators the upper tails of the left bounds in red are actually *above* the exemplar value. In fact, for both 'above' and 'over', fully 42% of the left bounds were above their exemplar values. This means, for example, that the phrase 'above 6.3' for some people implies a range that includes the value 6.3, but for others implies a range that does not include 6.3. This rarely occurs for the other approximators for which in all cases less than 1% of the left endpoints exceed their respective exemplar values. The patterns for 'at least' and 'no less than' seem to be very similar to each other. The graphs for 'above' and 'over' are also largely similar to one another, but differ from the graphs of 'at least' and 'no less than' because the latter pair almost always includes the exemplar value. Analogous statements can be made about 'at most', 'no more than', 'below' and 'almost', which are effectively transpositions of the previous four approximators.

Fig. 4 shows the same relative uncertainties in detail for four of the approximators. The intervals consisting of left and right endpoints, again scaled as relative displacements from the exemplar value of the numerical expression, are plotted as horizontal line segments. The intervals are sorted by their breadths, with the narrowest intervals at the bottom and the widest intervals at the top of the graphs. The graphs in Fig. 4 reveal which left endpoint is associated with which right endpoint. As a consequence, we can observe the breadth and pattern of specific responses to specific hedges across respondents. The graph for 'about' and the null approximator are symmetric, with a few idiosyncrasies by some workers.

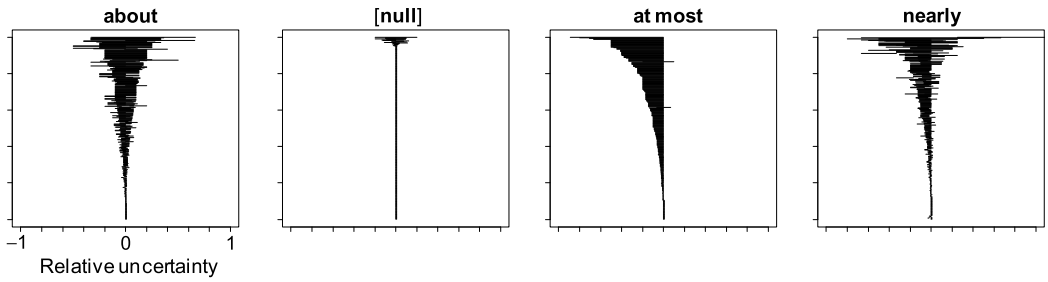

Fig. 4. Intervals of relative uncertainty understood by individual workers for several hedges.

The graph for 'at most' is strongly asymmetric. Although there are a couple of unusual intervals for this approximator, the graph reveals strong conformity about the direction of, and fewer idiosyncratic interpretations of, the uncertainty implied by the approximator. The graph for 'nearly' on the other hand reveals both symmetric and asymmetric intervals, the admixture of which seems to correspond to semantic ambiguity in this approximator. This result suggests guidance for elicitors. For example, if an informant uses the approximator 'nearly', an elicitor would be advised to seek to disambiguate which sense of the word was intended. Similarly in automated text mining applications, alternative potential interpretations of a sentence including 'nearly' as a hedge term could be flagged as more ambiguous than other consistently symmetric or asymmetric qualifying terms.

3.1. Uncertainty is commonly understated

An important consideration is whether people seem to be underestimating the uncertainty implied by numerical expressions involving hedges. We may ask how often workers characterize a numerical expression as a zero-width interval, and how often the intervals are narrower than would be implied by the significant-digits convention. The results vary sharply among the different approximators, as shown in Table 2 below. In the table, the *Zero Width* column gives for each approximator the percentage of intervals reported by workers for which the left and right endpoint are the same value, indicating a point value as the only possible value, and a degenerate interval of width zero. The *Too Narrow* column gives the percentage of reported intervals which are narrower than would be implied by the significant-digits convention applied to the respective exemplar numbers. The corresponding sample sizes are shown in the last column of the table.

Zero-width intervals were surprisingly often reported by workers. This response was nearly universal for the 'precisely' and 'exactly' approximators, with only 3% of intervals interpreted to correspond to more than a single possible number, and it was nearly as common for the null approximator. Even more surprisingly, zero-width intervals were also occasionally interpreted for the other hedges. It might seem to defy all mathematical sensibility to say that the meaning of "approximately 2300" could not entail the value 2302 and must be identical to 2300.000... with infinitely many decimal places, but the workers responded this way about 2% of the time. We thought that such responses might arise from a conflict between the semantics of the hedge and the discreteness of the units of the value being characterized, such as when a test phrase is "approximately 2346 people" which might be interpreted as fatuous hedging because one cannot have a fractional person. This conflict may explain the zero-width interval reported for the expression "approximately 3005 coffee houses", but does not seem to be at play in the majority of zero-width intervals such as those reported for the expressions

"approximately 25 kg",
"approximately 70 countries",

and even

"approximately $124 000", and
"approximately 10 billion dollars",

which apparently are interpretations by particularly literal-minded workers.

Outside of mathematics, perfectly precise numbers are incredibly rare topics in human discourse. All numbers that are actually physically *measured*, which are essentially all numbers that have units other than integer counts or tallies of integer counts, inescapably have measurement error [52]. It is possible that workers understand this fact intuitively yet lack the mathematical background to express this in terms that could be captured in the Amazon Mechanical Turk experiments.

Even if we set aside the reports of zero-width intervals, the results are also striking because so many intervals are narrower than would be implied by the significant-digits convention. With exceptions perhaps for the 'precisely' and 'exactly' approximators, the interval implied by the significant-digits convention is arguably the logical *lower limit* on the width of an interval of possible values implied by a numerical expression, which arise from basic considerations about how precise numbers might have been rounded to form the exemplar value. We had expected that the intervals reported by workers

Table 2
Evidence that uncertainties implied by numerical expressions are underestimated.

Approximator	Zero width (%)	Too narrow (%)	Sample size
Precisely	97	98	450
Exactly	97	99	362
[Null]	94	98	407
About	0	48	366
Roughly	~0	40	367
Approximately	2	44	369
Around	0	42	397
Almost	2	62	411
Nearly	2	58	388
At most	0	46	359
No more than	1	46	321
Below	~0	51	340
Over	~0	49	371
Above	~0	47	352
No less than	1	48	376
At least	1	51	402

would rarely be narrower than the respective significant-digits intervals. Thus, apart from maybe the first two rows, the percentages in the *Too Narrow* column are quite surprising in that between 40 and 60 percent of intervals are narrower than this limit. Nearly half of the responses to a question asking, for instance, what "about 2300" means specify intervals that are narrower than 100, which is the width of the interval [2250, 2350] implied by the significant-digits convention. We had presumed that 100 would be a lower limit on the width, because this interval is implied by the structure of the exemplar alone [5,72,52], before considering the semantic implications of the hedge 'about' which should only broaden the interval. Apparently, this reasoning does not apply in the interpretations of many people.

3.2. Regression analyses

We undertook linear regression analyses to explain the variation observed for the base-ten logarithm of the unnormalized widths of the intervals reported by workers as a function of the magnitude and roundness of the exemplar values and other predictor variables. We used separate regressions for each approximator. We could have instead constructed a grand regression with the approximator itself as a predictor variable along with the magnitude and roundness of the exemplar value, but doing so would have entailed an assumption of homoscedasticity among residuals for the various approximators. Such an assumption would be untenable given the scatter evident in the graphs of Fig. 3.

Not all variables offered as predictors could be reasonably used together in regressions. For instance, the Boolean variable indicating the discreteness of the unit is entailed by the unit group categorical variable, so both should not be used together. Likewise, the magnitude, number of significant digits and roundness variables are essentially coplanar as one is a function of the other two. (Roundness is the integer part of the base-ten log of the exemplar value plus one and minus the number of significant digits.) Thus, a regression can only use two of these variables as predictor variables. We found that the log exemplar value and roundness yielded slightly higher regression fits as measured by the coefficient of determination. We excluded from the regression analyses any intervals of zero width because their log ranges would be negative infinity. Note that this included most of the intervals reported for the approximators 'precisely' and 'exactly' and [null].

Table 3 gives coefficients for an expression to predict the log width of the interval implied by a hedged numerical expression involving a given approximator and exemplar number. The log width is

$$L = A + Bz + Cr + Df + Ezr + Fzf + Grf + Hzrf$$

where z is the base-ten log of the magnitude of the exemplar number, and r is its roundness, computed as the base-ten log of its order of significance, and f is 1 if the exemplar numeral ends in a '5' and 0 otherwise. The interval is predicted to be 10^L units wide. The residual uncertainty about this width not accounted for in the regression is expressed as a lognormal distribution with mean equal to $10^{\sigma^2/2}$ and variance equal to $10^{2\sigma^2}-10^{\sigma^2}$, where σ is given in Table 3. The last column of the table gives the (multiple, unadjusted) coefficient of determination which characterizes the goodness of fit of each regression. Comparable coefficients from parallel regression analyses for the minima and maxima of the intervals directly (rather than their ranges) are available at the project website https://sites.google.com/site/numericalhedging/hedge-code.

Previous work quantifying the numerical implications of approximators by Channell [14] suggested that the *units* of the quantity make a difference in how wide the perceived interval would be. Because we employed 189 individual units in the test statements, we did not have sufficient sample sizes to explore this question for each unit separately. Instead, we grouped units into seven dimensional categories (money, length, weight, time, speed, percent, and discrete). These seven levels formed a factor in the anovas. The linear model included this factor as well as the magnitude and roundness of the exemplar number as predictor variables. In none of the analyses for the sixteen approximators was the unit group factor statistically significant. We also looked for an effect of whether the unit of the quantity was countable or not. The linear models were expanded to include a binary variable that was set to true if the quantity was discrete (countable) or false if

Table 3
Coefficients from regression analyses for each approximator to predict the width of an interval implied by a hedged numerical expression.

Approximator	A	B	C	D	E	F	G	H	σ	R^2
About	−0.2085	0.4285	0.2807	0.0940	0.0147	−0.0640	−0.0102	0.0404	0.5837	0.7412
Roughly	−0.103	0.3687	0.2559	−0.0303	0.0353	0.1051	0.1422	−0.0562	0.5966	0.7211
Approximately	−0.3171	0.4993	0.254	0.6410	0.0177	−0.2835	0.1025	0.0169	0.6192	0.7364
Around	−0.1018	0.3429	0.3169	0.0951	0.0381	−0.0005	0.0174	−0.0029	0.5261	0.8118
At most	−0.3076	0.4751	0.2477	0.1168	0.0088	−0.0619	0.1551	0.0052	0.5956	0.7432
At least	−0.1128	0.3624	0.3829	0.3188	0.0087	−0.1404	0.0069	0.0409	0.5927	0.7562
No more than	−0.2699	0.4187	0.2418	0.2216	0.0382	−0.0467	0.0689	0.0069	0.5916	0.7640
No less than	−0.0187	0.3412	0.2207	−0.1427	0.0341	0.1616	−0.1480	0.0083	0.6314	0.7475
Over	−0.0668	0.3793	0.2490	−0.1635	0.0344	0.1353	−0.0068	−0.0176	0.6666	0.7643
Above	−0.1736	0.4483	0.2625	0.0224	0.0112	0.0669	−0.1354	0.0156	0.6668	0.7079
Below	−0.3052	0.4275	0.2666	0.3141	0.0353	−0.1348	0.1678	−0.0168	0.6577	0.7500
Almost	−0.4539	0.4593	0.3567	0.4006	−0.0196	−0.2245	−0.0534	0.0882	0.6640	0.7140
Nearly	−0.2716	0.3420	0.2722	0.0923	0.0440	0.0196	0.0386	−0.0270	0.5969	0.7677
[Null]	0.2070	0.1374	−0.4265	−0.4267	0.2450	0.3341	2.1650	−0.6876	0.7869	0.6400
Precisely	−0.4989	0.5884	0.3812	1.3500	−0.0774	−0.8274	−0.7248	0.2464	0.3859	0.8566
Exactly	−0.8360	0.7434	0.6058	5.427	−0.2055	−0.7757	0.0000	0.0000	1.0370	0.7019

it was continuous (measurable). In only two of the sixteen analyses was this variable a statistically significant predictor of the log range ($P = 0.025$ for 'almost' and $P = 0.047$ for 'approximately'). Thus, although it may be true that the unit of a numerical expression makes some difference in the implied width of the interval of possible values, it does not appear that either the dimension or the countability or measurability of the unit makes such a difference as far as we detect.

3.3. Numbers without hedges

There is an old joke about a janitor at the American Museum of Natural History who was heard to tell museum visitors that some dinosaur skeleton was 65 million and 7 years old. When asked about the 7 years, he explained that, when he was hired 7 years ago, the museum curator had told him the skeleton was 65 million years old. Unfortunately for both risk communicators and automated text analysis systems alike, the premise of this joke is true to life in that numbers that are not modified with hedge words are commonly interpreted as being more precise than they likely are.

As shown in Table 2, we found that few workers interpret numbers expressed with the null approximator as implying any uncertainty at all. Out of a sample size of 407 intervals describing the possible values implied by numbers without any explicit approximator, 94% had *zero width*, as though those numbers were perfectly precise. These responses created the long spikes at zero in the '[null]' graphs in Figs. 3 and 4. Fully 98% of the intervals reported for numeric expressions with the null approximator understated uncertainty relative to the traditional scientific interpretation based on the significant-digits convention.

Preliminary research using questionnaires that was previously conducted with respondents recruited from among graduate students in a class on risk analysis supported the same conclusions. When asked to indicate the smallest and largest possible values that were consistent with a given phrase, all but one of the respondents gave zero-width intervals for the unhedged numerical expressions '7', '470', and '2.31'. Thus, even (neophyte) risk analysts seem to be prepared to accept numerical expressions with the null hedge as perfectly precise quantities. They also gave the same degenerate intervals when asked about the phrases 'exactly 7', 'exactly 470', 'exactly 2.31', 'precisely 7', 'precisely 470', and 'precisely 2.31', suggesting that the approximators 'exactly' and 'precisely' have no substantive quantitative implications on the numeric interpretation compared to the null hedge, although they may have some linguistic role.

Research suggests that effective risk communication requires clear and consistent messages, and risk communicators worry that adding uncertainty information to forecasts may confuse the message and impede understanding and action by the public [47, p. 69]. Yet, as risk analysts well know, this uncertainty matters. For example, when the Red River was forecasted during the 1997 flood to crest at 49 feet, residents believed their 52-foot dikes would protect them. When the river actually crested over 54 feet, it overtopped dikes and inundated communities in North Dakota and Minnesota leaving them devastated [45]. Residents and local officials seem not to have appreciated that the hydrological prediction had uncertainty, and that this uncertainty was increased by the fact that the event would be a record-breaking event. Forecasters were aware of the uncertainty (although they may have substantially underestimated it), but did not effectively communicate this uncertainty to residents and decision makers. The prediction was incomplete in that it did not highlight various scenarios that might entail higher water levels. No bounding or worst-case predictions were provided [31]. Updates to the prediction were not timely, and the fact that the prediction did not change much may have been interpreted by the public as surety about the forecast. Nevertheless, the subsequent strong public criticism for underestimating the flood seemed to be hard for forecasters to accept because, after all, they had come to within a few feet of predicting this unprecedented event. Political decision makers had expressly requested single forecasts of crest levels [49,47]. The forecasters had believed the public would understand that all such predictions are uncertain.

Of course it is not clear that simply affixing some hedge such as 'about' or 'at least' to the predicted water level could have resulted in better risk communication, or in any way have changed the outcome of the Red River flood. But it would

have allowed for a more easily defendable truth qualification as described by Prince et al. [50]. What is clear from our findings is that hedge-less numerical expressions are very likely to be misinterpreted if the number has any uncertainty, and the direction of the misinterpretation invariably diminishes the uncertainty. The situation is likely worsened when the number is expressed by a perceived authority. If the people who have studied the problem say the answer is 49 feet, consumers of the information usually have scant cause to question the estimate. Shield hedges such as 'we expect that' or 'we estimate that' seem to be automatically discounted by the public because the pronouncements come from authorities, and in some cases the only authority. This is part of what we might call the *authority problem* which is that people tend to overtrust pronouncements delivered by authority figures [58,19,21,10]. It makes little impact how much authorities qualify their statements as opinions, because the whole reason to appeal to authorities is to elicit those opinions which, in the absence of other ideas, have to serve as the only estimates. This makes it all the more important for an honest analyst to convey the uncertainty about a prediction in the expression of the prediction itself.

Given that the public often demands single-number predictions, using hedges may be the only way for an analyst to sneak this uncertainty into the prediction, especially since the significant-digits convention appears to be not communicative and is incomplete in that it cannot express every possible range of uncertainty. Numbers expressed without hedge words are very likely to be commonly misunderstood as being more precise that they actually are. Siegrist et al. [59] call such unqualified expressions "naked numbers" and argue that they can produce erroneous results in cost–benefit analyses, risk assessments and other analyses under uncertainty.

3.4. Roundness of the exemplar number

Round numbers are integers whose numerals end in one or more zeros. Among numbers of the same order of magnitude, having more ending zeros makes the number rounder, so 123 000 is rounder than 123 450. The idea of roundness is often extended to noninteger numbers, so 1.23, which is equal to 1.23000, is rounder than 1.23456. For a given order of magnitude, the log order of the least significant digit directly measures the roundness of the number. For example, the numbers 12 100, 12 342, 12 780, and 12 900 have log significance orders of 2, 0, 1, and 2 respectively, despite all the numbers having the same order of magnitude of 4. For integers, the log significance order counts the non-significant trailing zeros. The regression analyses discussed above reveal that increasing roundness of the exemplar number is associated with larger width of the reported interval, even after controlling for the magnitude of the exemplar number. The regression coefficients for roundness (log significance order) were positive for every approximator studied.

Among English[2] speakers, zero is not the only digit that can be used to make round numbers. It is often presumed that a numeral ending in the digit '5' is a rounder number than one ending in another nonzero digit. Five is, in a sense, halfway to zero. It has been assumed in the linguistics literature (see, for example, Krifka [35]) that approximators such as 'about', 'around', and the null approximator would allow wider ranges of possible values if applied to an exemplar value with a least nonzero digit of '5' rather than another nonzero digit. As an example, one can imagine that there might be a difference between two answers to the question "What time is it?", one answer where the number of minutes ends with 5, such as "It's 12:35" vs. an answer where the number of minutes ends with a nonzero number different from 5, such as "It's 12:37." It has been argued that in the former case, the interval is larger than in the latter case: it could be [12:33, 12:37] for the former case, an interval with the width is four minutes, compared to the interval [12:36 and 30 seconds, 12:37 and 30 seconds] for the latter case which is an interval with the width one minute. Regression analysis can also address the question of whether numerals ending in '5' imply wider intervals than those that end in other nonzero digits.

We tested this question using the data for the symmetric approximators 'about', 'around', 'approximately', and 'roughly'. In anovas that controlled for the effects of magnitude and roundness of the exemplar value, we found that there was a positive and statistically significant effect only for 'around' ($P = 0.024$). For the approximator 'roughly', we observed a significant statistical interaction that precludes a simple interpretation of the effect, and we observed no significant effect for either 'approximately' ($P = 0.097$) or 'about' ($P = 0.36$). However, because the set of these approximators is homoscedastic and the original claim by linguists was about symmetric approximators as a class, it is perhaps reasonable to test for the effect on the pooled data. This test of pooled data revealed a positive and statistically very significant ($P < 0.0002$) effect. If the intervals for the null approximator are also included in the analysis, the effect is still highly significant ($P < 0.0003$). We thus confirm that ending the numeral in the exemplar value with a '5' is interpreted to imply a wider interval of possible values, at least for the symmetric approximators, although the size of the difference is small compared to the effect of either the magnitude or roundness of the exemplar number.

What is the quantitative implication of this significant effect of five-rounding? Fig. 5 depicts the effect of rounding the exemplar number to end in a '5' on the perceived uncertainty of numerical expressions for all the symmetric approximators. The figure shows simultaneously the effects of magnitude, roundness and five-rounding. The lengths of the vertical line segments represent the possible effect (in terms of log range) of having or not having a '5' as the last significant digit. They are the possible values predicted from the fitted regression model for different nonzero last significant digits. The predictions

[2] People tend to exhibit preferences for the last digit of numbers they report according to their respective language. French and Italian speakers favor 0 and 5 as the last digit, whereas 2, 4, 6 and 8 are more favored by German speakers [9], and this language-specific pattern seems to be irrespective of nationality. English speakers seem to favor 0, but apparently also 8 and 5 de [40].

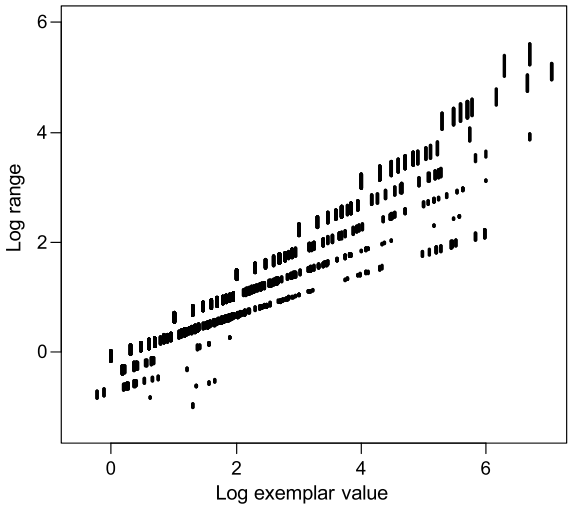

Fig. 5. Breadths (vertical segments) of uncertainties about the log range of intervals implied by exemplar numbers ending in a '5' versus another nonzero digit predicted for all symmetric approximators.

were made at every value of magnitude and roundness in the observed data set. Given a magnitude and roundness for the exemplar number, the upper endpoint of the depicted line segment corresponds to the exemplar number ending in '5', and lower endpoint corresponds to it ending in another nonzero number. The rows apparent in the scatter of the line segments correspond to the various integral log significance orders (roundnesses), with small orders at the bottom of the graph and larger orders toward the top. The statistically significant interaction between the log magnitude and the log significance order of the exemplar value is reflected in the graph. We see the interaction in the fact that the rows corresponding to different log significance orders have noticeably different slopes. The scattergrams for the individual approximators are qualitatively similar to the pattern in Fig. 5.

4. Encoding uncertainty into hedged numerical expressions

Having considered the direct problem of interpreting the quantitative meaning of a hedged numerical expression that has been produced by a native speaker, we now address the reverse problem of formulating a hedged numerical expression to encode uncertainty for communication. Analysts compute uncertainties from their risk analyses which then need to be communicated to decision makers and sometimes the public at large, and it is incumbent on the analyst or the risk communicator to express these uncertainties in ways that can be understood and appreciated.

There are three inputs for this process. The first and main input is the interval of uncertainty to be conveyed, which might be a confidence interval or a dynamic range of some variable or some other window of uncertainty in which a quantity has been isolated. The second input is the dimension of the units. The units themselves may be selectable. For instance, when the analyst is given that the quantity is a time, several possible units are available, including second, minute, hour, day, week, month, year, decade, etc. The unit can be selected to most readily convey the uncertainty. Besides the dimension and the interval range, the only other input is whether the uncertainty should be described symmetrically or asymmetrically and, if the latter, from which direction. With these three inputs, there are constraints on the magnitude of the exemplar number to be used in the constructed numerical expression, although there is some leeway in its particular value.

In practice, perhaps the simplest way to find an appropriate hedged numerical expression for conveying a given interval is to use a simple trial-and-error strategy varying the approximator, the units, the exemplar value, and its roundness (or number of significant digits) to find the combination that minimizes the difference between the log width of the interval computed using the regression analyses and that of the actual interval. Alternatively, this fitting can be constrained to among intervals that encompass the given interval so as to never understate the uncertainty. There may not be a single optimal way to express the uncertainty from a risk analysis. Indeed, the similarities we observed among several of the hedges suggest that, in many cases, multiple approximators may be essentially equivalent for a particular task. The search strategy therefore need only be satisficing and not optimizing.

The latest regression coefficients from analyses conducted as part of the research described in this paper are embedded in software written in R [51] to interpret hedged numerical expressions. The code is available from the authors or directly from https://sites.google.com/site/numericalhedging/hedge-code. In some applications it may be reasonable to undertake a special assessment, perhaps via traditional questionnaires, Amazon Mechanical Turk, or other on-line survey approach [4]. Such an assessment can directly characterize the role played by different individual units that could be used to express results on the quantitative implications of hedging.

5. Conclusions

Different English speakers understand approximators such as ‘about’, ‘nearly’, and ‘no less than’ in different ways, and we find considerable inter-individual variation, but there are consistent patterns in the perceived implications of the various approximators. These patterns are strongly modulated by context such as magnitude of the number, its roundness, and even possibly its units.

With regression analysis and sample sizes in the thousands obtained through Amazon Mechanical Turk, we are able to construct a *decoding* for each approximator that explains the likely uncertainty that will be understood when it is used as a hedge for a numerical expression. These decodings characterize the quantitative implications of the uncertain expressions, which is important for understanding patient complaints about, for instance, a headache “that’s lasted over 7 days”. The residual variation around each regression prediction can be expressed as a probability box encompassing a distribution of results across respondents and the epistemic uncertainties they reported. These findings can also be applied to *encode* uncertainty for fashioning numerical expressions used in risk communication, so long as the uncertainty is not very wide.

The differences between approximators that we originally expected to see do not seem to be empirically justified by the data obtained from native English speakers. There is far less variation among the hedges than might be imagined, and far less than would be linguistically and practically useful. In fact, the approximators ‘about’, ‘around’, ‘approximately’ and ‘roughly’ are almost identical in distribution. Likewise, the approximators ‘precisely’, ‘exactly’ and the null hedge appear to be very similar, almost to the point of indistinguishability. The asymmetric approximators are similar to one another as well, modulo the direction of asymmetry. Most surprising is our finding that the null hedge entails no uncertainty at all among most English speakers. If this finding persists under different experimental designs and interrogation schemes, it is a fundamentally important lesson for risk communication. It would imply that the significant-digits convention used by many scientists to express numerical results and interpret numerical values reported by others is essentially totally ignored by most people.

6. Discussion

Even though honest accounting of uncertainties can make criticism of risk assessment projections unsportingly easy (e.g., [34]; cf. [2,27]), almost all risk analysts agree it is of fundamental importance. There are a host of special problems associated with risk communication of uncertain values [46,44,23,63,62], and there are good reasons to believe that, as a profession, we are doing it incorrectly, or at least inefficiently.

The ability to quantitatively decode and encode natural linguistic expressions of uncertainty would seem to be a basic facility that must be developed in risk communication. Hedged numerical expressions are extremely common in writing and speech in English. Automated Google numerical searches reveal usage patterns for approximators in the English-language text corpus of the indexable Web (which excludes FTP files, chat, etc.). Searches for all combinations of a list of common units and hedges associated with a number between zero and one million using Google queries like “`about 0.1000000 kilometers`” `yielded` the results depicted in a heatmap in Fig. 6 where darker shades of gray imply higher counts. Base-ten logarithms of the occurrence counts are superimposed on the heatmap. The approximators ‘about’ and ‘over’ were the most common hedges and the results roughly followed the word frequency in the English language. Units of time (especially years) and people were the most common units associated with hedge words.

The experiments described in this paper can be extended and generalized in a variety of ways. For instance, it would be relatively straightforward to consider fractional values, negative values, absolute times such as dates, quantities on special scales such as temperatures, mixed scales such as feet & inches, and perhaps geographical coordinates. It would also be easy to explore possible regional differences in the quantitative implications of approximators. Amazon Mechanical Turk can also be used to extend the experiments to languages other than English. There is considerable similarity among human languages with regard to how approximators are used, although there are some exceptions. For example, Russian uses words such as ‘около’ (‘nearly’), ‘почти’ (‘almost’), ‘примерно’ (‘about’), etc., in ways similar to their English counterparts, but Russian also has a linguistic construction known as approximative inversion in which switching the order of the numeral and the noun creates an approximative reading [68]. The following sentences illustrate this construction:

Иван встретил 40 человек	Иван встретил человек 40
Ivan met 40 people	Ivan met people 40
“Ivan met 40 people”	“Ivan met approximately 40 people”

A number of East Asian languages such as Japanese, Korean and Mandarin use obligate numerical classifier words comparable to measure words even for notions that are considered countable nouns in English. These classifiers also contain meaning. For example, the word ‘*kuai*’ in Manadrin is used for chunky things such as rocks or piece of pork [39]. Apart from these rather minor differences, the systems of approximators used in all natural languages seem to have remarkable congruences, which suggest that they should be amendable to study and cross-linguistic comparison using Amazon Mechanical Turk.

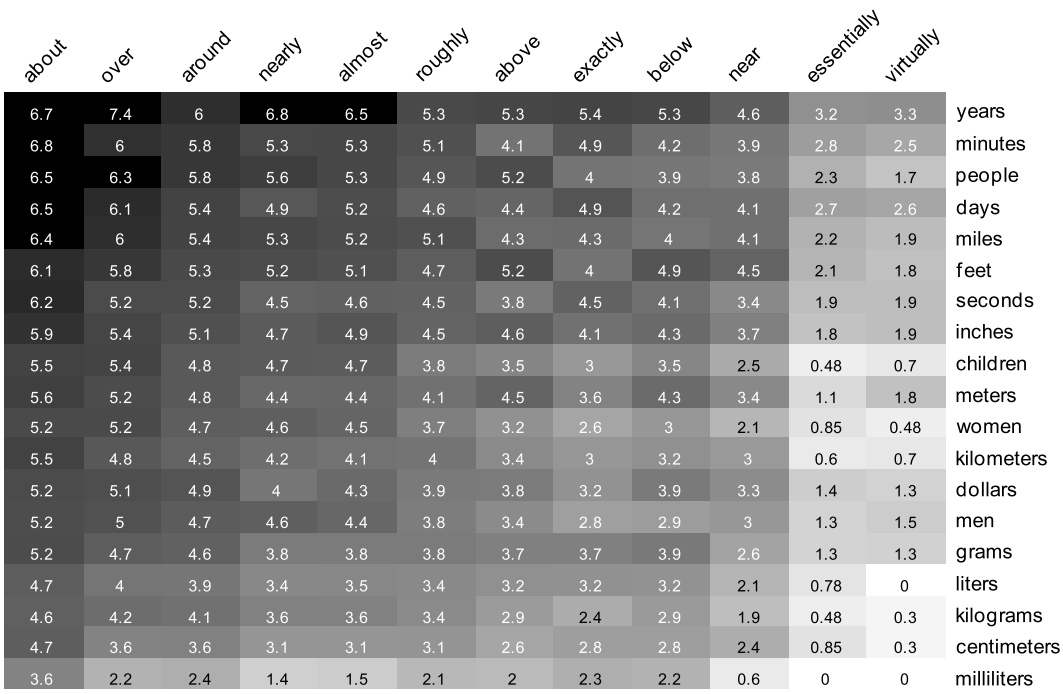

Fig. 6. Grayscale heatmap of the base-ten logs of the number of occurrences for various approximators (columns) and units (rows) in numerical expressions found by Google.

6.1. Other language systems conveying uncertainty

Channell [14] described a variety of ways English speakers communicate their uncertainties. Besides the unit–number–approximator paradigm, English has at least two other systems to convey imprecision, or equivalently, to qualify the precision that an expression about a quantity should be understood to have. Both systems use a unit–quantifier paradigm that folds imprecision and magnitude together. In the first system the quantifiers are pseudonumbers such as 'dozens of', 'millions of', 'bajillion', 'eleventy', 'umpteen', and placeholders borrowed from mathematics such as 'X', 'Y', 'Z', 'n-tuple', and 'a number of' that denote unspecified numbers often to be determined later. Pseudonumbers already combine imprecision with magnitude, but they too can be hedged as in the phrase 'about a bajillion dollars', although it is unclear whether the approximator makes a difference. The second system uses an elaborate array of generally non-numerical words such as 'much', 'many', 'little', 'few', 'some', 'both', 'a couple of', 'several', 'all', 'any', 'each', 'every', 'most', 'not any', 'not all', 'a handful of', 'a lot of', 'lots of', 'plenty of', 'innumerable', etc., which grammatically are determiners that express a relative or indefinite indication of quantity. Although these constructs are called quantifiers in linguistics, they convey only qualitative information about a magnitude or quantity.

The study of quantifiers dates to the invention of logic itself by Aristotle [67]. Modern first-order logic recognizes only the universal quantifier (∀, 'for all') and the existential quantifier (∃, 'there exists'), perhaps corresponding to 'every' and 'some' respectively. But logicians and linguists today recognize a substantially generalized notion of quantifiers [3,67], which includes approximators in a special category. Among these quantifiers are primitive and derived forms (e.g., 'not any'). Unlike approximators which are continuously generated as language evolves, the list of primitive quantifiers is fixed and consistent across all human languages, although it is not exhaustive logically (e.g., no natural languages have 'not all' expressed in a single word). Nearly two hundred years ago the philosopher De Morgan attempted to develop a calculus for interpreting and making quantitative inferences from semi-quantitative expressions using linguistic quantifiers [53]. It would certainly be helpful to bring this work into flower by establishing rules to decode the quantitative implications of generalized quantifiers, and thus allowing automated quantitative interpretation of natural-language utterances that involve words like 'some', 'several', 'most' and 'not all'. It may be possible that a direct approach using MTurk such as we have used to study the quantitative implications of approximators in numerical expressions may also be useful in studying the quantitative implications of linguistic quantifiers.

6.2. Fuzzy sets or other structures

It is possible that some or most people have more complex conceptions of hedged numerical values than simple ranges. For instance, it might reasonable to assert that the numeral 7 with some particular approximator corresponds to a trapezoidal fuzzy number [32]. In particular, one person might interpret it to be $[6, 6.5, 7.5, 8]$ (which has a flat top over the

range [6.5, 7.5] representing entirely possible values, and a base spanning the range [6, 8] as the conceivable values). Other people might contrarily assert for the same hedge that the flat top and range are considerably wider.

Channell [14] claimed that individuals interpret numbers with approximators as implying only interval ranges, rather than some more complex characterization. In our own preliminary work, we found subjects did not provide any more detailed information than intervals, even when asked to. Because intervals are fundamental structures for both imprecise probability and fuzzy-set approaches, the results of our study can be applied in both. For these reasons, we only collected intervals in the MTurk experiments.

Fig. 3 depicted the collections of intervals reported by a population of workers as probability boxes [20], but these collections could also be summarized by fuzzy numbers [32]. Many ways to construct fuzzy membership functions from empirical data have been described (e.g., [24,65,60,64,6,42,43]). In principle, the MTurk approach could be extended to elicit fuzzy numbers or probability boxes or other structures directly from individuals. These experiments could be designed to detect whether any individuals conceptualize hedged numerical quantities with more information content than simple intervals.

There is both a clear connection as well as a distinction between this work in interval-valued approximators and the linguistic hedges associated with Zadeh [71], fuzzy logic, and the computing-with-words trend that has unfolded over the last four decades. In his seminal work on the topic, Zadeh [70] defined a linguistic hedge as "an operator which acts on the fuzzy set representing the meaning of the operand". We can think of the approximator as special type of linguistic hedge that acts a *constructor* operator to convert a precise numerical value to an imprecise quantity (interval, fuzzy set or probability box). In reaction to Zadeh's work, Lakoff [37] affirmed that a better understanding of modifiers is important in natural language semantics. In questioning some of the shortcomings of specific modifier valuations, which include quantifications of hedges, he posited a perceptual model for interpretating modified values that depend on contextual factors. The MTurk crowdsourcing experiment is an attempt to elicit these valuations and to find commonalities in perceptual models across informants. Such a research program would expect to find semantic invariants that could be mapped as uniformities across situations [54]. An attempt to characterize such a mapping with approximators is reflected in the diversity in context data used in the MTurk experiment.

Fuzzy methods offer an alternative and more comprehensive strategy for quantifying the implications of hedged expressions. In particular, they promise to characterize both shields and approximators. Indeed, it is common for natural language statements to include a shield hedge in combination with an approximator. For instance, "it's most likely around 5 days old" may translate into an imprecise probability distribution centered around 5 days.

6.3. The special case of probability

Expression of probabilistic risk is a fundamental special case. Humans and maybe primates generally seem to have an innate probability sense [25], although it may fail to engage for ill-formatted sensory data [15].

A wide variety of verbal, graphical and other techniques have been suggested for conveying a probability of a well defined event. One fundamental difficulty in this most basic risk communication task may be that most strategies presume the probability is precisely characterized as a real number [62]. In fact, probabilities are usually estimated from data limited in abundance and precision. Likewise, risk analyses often yield imprecisely specified probabilities because of measurement error, small sample sizes, model uncertainty, and demographic uncertainty. In contrast, human cognition seems to account for the effect of sample size automatically so that '10 out of 100' is perceived to be more precise than '1 out of 10'. Gigerenzer and his colleagues [22,23,36] argue that "natural frequency" in expressions like 'k out of n' is an effective tool for conveying an event probability, including the reliability of the estimate embodied in the n-value.

This facility for natural frequencies can be exploited for communicating calculated risks. Under the theory of confidence structures [1], the probability of an event estimated from binary data with k successes out of n trials is associated with a structure that has the form of a probability box. When n is large, this structure approximates the beta distribution obtained by Bayesians under a binomial sampling model and Jeffreys prior, and asymptotically it approximates the frequentist scalar estimate k/n. But when n is small, it is imprecise and cannot be approximated by any single distribution because of demographic uncertainty that arises from estimating continuous variables from discrete data. When a risk analysis yields a result in the form of a precise distribution or imprecise probability box for an event's probability, we can approximate the result with a binomial probability estimated for some values of k and n. Thus we can characterize the event probability from the risk analysis with a terse, natural-language expression of the form 'k out of n', where k and n are nonnegative integers and $0 \leq k \leq n$. Note this natural frequency is one of several numerical hedges using paired numbers. It is comparable to, and composable with, 'x or y', and 'x to y'.

6.4. Future work: ludic elicitation

There are in principle several ways to approach understanding how uncertainties from numerical summaries expressed in English phrases should be decoded quantitatively to inform a risk analysis. Questionnaires asking readers the meanings of hedges were found to be cumbersome and even dizzying to the informants. Such questionnaires could not conveniently include potentially important background information that would give context to each use of a hedged numerical expression. However, using Amazon Mechanical Turk, the same kinds of questions can be contextualized and presented as separable

individual tasks posed to human workers. We found this to be a convenient approach that produces good sample sizes, but it may yield biased results if the human workers are not sufficiently motivated to produce good, thoughtful answers.

In principle, the quality of a worker's answers can be checked and rejected if they are incomplete or poor, in which case the worker is not paid for the effort and is stigmatized within the accounting system of Amazon Mechanical Turk. In practice, however, it would be quite difficult to check the responses, especially because we are interested in the variation in responses to questions that have no clear answers. Although the hourly rates and the payments they receive are pitifully small, there may be a tendency among workers to work fast rather than carefully, in order to maximize their incomes. How can we be sure that they are not providing slapdash responses that do not reflect what they actually believe about the hedges under study? How can we prevent workers from gaming our system and not giving us their considered responses? Interestingly, the best strategy may be to turn our system into a game.

An alternative strategy to collect more realistic information more easily uses *ludic elicitation* based on Luis von Ahn's idea of internet games to harvest human intelligence [38] in which, as a part of on-line game play, humans make decisions that reveal interpretations and preferences by their decisions. Artfully designed games can elicit information as a side-effect of play so that the elicitation process is enjoyable to participants who will thus play longer and share more of their intelligence. This approach parasitizes human play, but it has relatively few moral or ethical problems since information is unlike resources that can only be shared in a zero-sum way. Sharing information does not diminish one's own supply of information. Although this process is a bit different from the normal approach of directly asking questions of human informants, scientific expertise about the design of questionnaires and experiments in general should nevertheless help us to avoid some pitfalls, misfires, and ambiguities in the results.

Game play is typically composed of three broad phases: a pre-ludic learning phase in which the participant is discovering the rules of play and coming to understand strategies to achieve high scores, a ludic phase in which the player actively plays the game because of its intrinsic interest and challenge, and a post-ludic phase in which a player either no longer plays the game at all or plays in a teasing or meta-play way that flaunts or ignores the game's prescribed rules and goals. Clearly, only the middle ludic phase is generally useful for information collection since players' responses are trustworthy reflections of their beliefs only in this phase. This means that attention must be devoted to discerning which phase each player is in.

We have implemented multiple games for Facebook to improve our assessments of the quantitative meanings of English-language approximators, and also to assess whether the linguistic encodings are effective tools for risk communication. The point is to confirm that the approximator interpretations produce inputs consistent with what was intended by users, and also that uncertainty projection routines produce computed answers that are justifiable with those inputs, or, if they are not, to correct the scheme and routines to improve the consistency.

Acknowledgements

We thank William McGill of Penn State, Lev Ginzburg and Dan Rozell of Stony Brook University, Mark Burgman and Louisa Flander of University of Melbourne, Nick Friedenberg of Applied Biomathematics, and Robert O'Connor of the National Science Foundation. This work was supported by the National Library of Medicine, a component of the National Institutes of Health (NIH), through a Small Business Innovation Research grant (award number RC3LM010794) to Applied Biomathematics funded under the American Recovery and Reinvestment Act, and by the National Science Foundation, through a grant (award number 0756539) to Adam Finkel at The University of Pennsylvania, and a subcontract with Applied Biomathematics. The views expressed should not be considered those of the National Library of Medicine, the National Institutes of Health, or the National Science Foundation.

Appendix A

Below is the Python code used to generate reasonable alternative numbers for constructing experimental sentences. The function generateNumbers takes as its first argument a numeric value and, as its second argument, a string code indicating the type of the value (integer, decimal).

```
import random
def generateNumbers( num, mode ):
     if mode == 'int':
          numLength = len(str(num))
          numList = []
          numList.append(str(num))
          order = 0
          digits = num
          for i in range(numLength):
                lastDigit = digits%10
                order = lastDigit* 10**i + order
                digits//= 10
                if order != 0 and order != num:
```

```
                numList.append(str(num - order))
            a1 =  num - order + random.choice ([1,2,3,4,6,7,8,9])*10**i
            if a1 > 0:
                numList.append(str(a1))
            b1 =  num - order  + 5*10**i
            if b1 > 0:
                numList.append(str(b1))
    return sorted(list(set(numList)), key = float)
elif mode == 'dec':
    numLength = len(str(num)) - 1
    factor  = 1
    for i in range(numLength):
            num*= 10
            factor*=10
    numList=[]
    numList.append(str(num/factor))
    order = 0
    digits = num
    for i in range(numLength):
            lastDigit = digits%10
            order = lastDigit*10**i+order
            digits //= 10
            if order! = 0 and order != num:
                numList.append(str((num - order)/factor))
                #numList.append(str(num - order + 10**(i + 1)))
            a1 = num - order + random.choice([1,2,3,4,6,7,8,9])*10**i
            a1/=  factor
            if a1 > 0:
                numList.append(str(a1))
            b1 = num - order + 5*10**i
            b1/= factor
            if b1 > 0:
                numList.append(str(b1))
 return sorted(list(set(numList)),key = float)
```

References

[1] M.S. Balch, Mathematical foundations for a theory of confidence structures, Int. J. Approx. Reason. 53 (2012) 1003–1019.
[2] T. Ball, Wrong prediction, wrong science; unless it's government climate science. Watts Up With That? [website], edited by A. Watts, http://wattsupwiththat.com/2013/01/08/wrong-prediction-wrong-science-unless-its-government-climate-science/.
[3] J. Barwise, R. Cooper, Generalized quantifiers and natural language, Linguist. Philos. 4 (1981) 159–219.
[4] J. Bethlehem, S. Biffignandi, Handbook of Web Surveys, Wiley Handbooks in Survey Methodology, vol. 567, John Wiley & Sons, 2012.
[5] P. Bevington, D.K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill Science, 2002.
[6] T. Bilgiç, I.B. Türkşen, Measurement of membership functions: theoretical and empirical work, in: D. Dubois, et al. (Eds.), Fundamentals of Fuzzy Sets, Kluwer Academic Publishers, 2000, pp. 195–227.
[7] R.P. Binzel, A near-Earth object hazard index, Ann. N.Y. Acad. Sci. 822 (1997) 545–551.
[8] P.P. Bonissone, K.S. Decker, Selecting uncertainty calculi and granularity: an experiment in trading-off precision and complexity, in: L.H. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence, North-Holland, Amsterdam, 1986, pp. 217–247.
[9] M. Bopp, D. Faeh, End-digits preference for self-reported height depends on language, BMC Public Health 8 (2008) 342, http://www.biomedcentral.com/1471-2458/8/342.
[10] M.A. Burgman, M. McBride, R. Ashton, A. Speirs-Bridge, L. Flander, B. Wintle, F. Fidler, L. Rumpff, C. Twardy, Expert status and performance, PLoS ONE 6 (7) (2011) e22998, http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0022998.
[11] D. Burr, J. Ross, A visual sense of number, Curr. Biol. 18 (2008) 425–428, http://www.pisavisionlab.org/teaching/burr/CB_number.pdf.
[12] J. Channell, More on approximations, J. Pragmat. 4 (1980) 461–476.
[13] J. Channell, Vague language use: some vague expressions, Doctoral dissertation, University of York, England, 1983 (in English).
[14] J. Channell, Vague Language, Oxford University Press, Oxford, England, 1994.
[15] L. Cosmides, J. Tooby, Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty, Cognition 58 (1996) 1–73.
[16] T. Dantzig, Number: The Language of Science, MacMillan Company, New York, 1930.
[17] M. Delgado, F. Herrera, E. Herrera-Viedma, L. Martínez, Combining numerical and linguistic information in group decision making, Inf. Sci. 107 (1998) 177–194.
[18] J. Denker, Measurement and uncertainties. Physics documents, http://www.av8n.com/physics/uncertainty.htm, 2011. See especially http://www.av8n.com/physics/uncertainty.htm#sec-execsum-sigfig, 2011 and http://www.av8n.com/physics/uncertainty.htm#sec-more-about-sigfig, 2011.
[19] E. Ert, I. Erev, The rejection of attractive gambles, loss aversion, and the lemon avoidance heuristic, J. Econ. Psychol. 29 (2008) 715–723.
[20] S. Ferson, RAMAS Risk Calc 4. 0 Software: Risk Assessment with Uncertain Numbers, Lewis Publishers, Boca Raton, FL, 2002.
[21] D.H. Freedman, Wrong: Why Experts* Keep Failing Us—And How to Know When Not to Trust Them: *Scientists, Finance Wizards, Doctors, Relationship Gurus, Celebrity CEOs, High-Powered Consultants, Health Officials and More, Little, Brown and Company, New York, 2010.

[22] G. Gigerenzer, The art of risk communication: what are natural frequencies? Br. Med. J. 343 (2011) d6386, http://dx.doi.org/10.1136/bmj.d6386.
[23] G. Gigerenzer, Calculated Risks: How to Know When Numbers Deceive You, Simon and Schuster, New York, 2003.
[24] R. Giles, Foundations for a theory of possibility, in: M.M. Gupta, E. Sanchez (Eds.), Fuzzy Information and Decision Processes, North–Holland, Amsterdam, 1982, pp. 183–195.
[25] P.W. Glimcher, D.L. Sparks, Representation of averaging saccades in the superior colliculus of the monkey, Exp. Brain Res. 95 (1993) 429–435.
[26] P. Gordon, Numerical cognition without words: evidence from Amazonia, Science 306 (2004) 496–499, http://www.sciencemag.org/content/suppl/2004/10/15/1094492.DC1/Gordon.SOM.pdf.
[27] T. Harris, Canadian government joins Alberta premier in climate change/pipeline fantasy lobbying, New.ideas@Frontier [blog], hosted by Frontier Centre for Public Policy, http://www.fcpp.org/blog/canadian-government-joins-alberta-premier-in-climate-changepipeline-fantasy-lobbying/, 2013.
[28] HHS [United States Department of Health and Human Services], Federal policy for the protection of human subjects, ('Common Rule') [website], http://www.hhs.gov/ohrp/humansubjects/commonrule/, 2013.
[29] G. Ifrah, From One to Zero: A Universal History of Numbers (transl. from the French by L. Bair), Viking Penguin, New York, 1985.
[30] G. Ifrah, The Universal History of Numbers: From Prehistory to the Invention of the Computer (transl. from the French by D. Bellow, E.F. Harding, S. Wood and I. Monk), John Wiley & Sons, New York, 2000.
[31] L.D. James, S.F. Korom, Lessons from Grand Forks: planning structural flood control measures, Natural Hazards Review 2 (2001) 22–32.
[32] A. Kaufmann, M.M. Gupta, Introduction to Fuzzy Arithmetic: Theory and Applications, Van Nostrand Reinhold Company, New York, 1985.
[33] S. Kent, Words of estimative probability, Stud. Intell. 8 (4) (1964), available at https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html.
[34] V.V. Kharin, Q. Teng, F.W. Zwiers, G.J. Boer, J. Derome, J.S. Fontecilla, Skill assessment of seasonal hindcasts from the Canadian Historical Forecast Project, Atmos.-Ocean 47 (2009) 204–223, see also http://weather.gc.ca/saisons/info_pc_e.html.
[35] M. Krifka, Approximate interpretation of number words: a case for strategic communication, in: G. Bouma, I. Krämer, J. Zwarts (Eds.), Cognitive Foundations of Interpretation, KNAW Publications, Amsterdam, 2007, pp. 111–126.
[36] E. Kurz-Milcke, G. Gigerenzer, L. Martignon, Transparency in risk communication: graphical and analog tools, in: W.T. Tucker, S. Ferson, A.M. Finkel, D. Slavin (Eds.), Strategies for Risk Communication: Evolution, Evidence, Experience, in: Annals of the New York Academy of Sciences, vol. 1128, 2008, pp. 18–28.
[37] G. Lakoff, Hedges: a study in meaning criteria and the logic of fuzzy concepts, J. Philos. Log. 2 (1973) 458–508.
[38] E. Law, L. von Ahn, Input-agreement: a new mechanism for collecting data using human computation games, in: ACM Conference on Human Factors in Computing Systems, CHI 2009, 2009, pp. 1197–1206, see also http://www.youtube.com/watch?v=tx082gDwGcM.
[39] F.-H. Liu, The count-mass distinction of abstract nouns in Mandarin Chinese, in: Theories of Everything, UCLA Work. Pap. Linguist. 17 (2012) 215–221.
[40] S. de Lusignan, J. Belsey, N. Hague, B. Dzregah, End-digit preference in blood pressure recordings of patients with ischaemic heart disease in primary care, J. Hum. Hypertens. 18 (2004) 261–265.
[41] G.M. Mace, R. Lande, Assessing extinction threats: toward a reevaluation of IUCN threatened species categories, Conserv. Biol. 5 (1991) 148–157, http://onlinelibrary.wiley.com/doi/10.1111/j.1523-1739.1991.tb00119.x/pdf.
[42] T. Marchant, The measurement of membership by comparisons, Fuzzy Sets Syst. 148 (2004) 157–177.
[43] T. Marchant, The measurement of membership by subjective ratio estimation, Fuzzy Sets Syst. 148 (2004) 179–199.
[44] M.G. Morgan, B. Fischhoff, A. Bostrom, C.J. Atman, Risk Communication: A Mental Models Approach, Cambridge University Press, 2002.
[45] R.E. Morss, E. Wahl, An ethical analysis of hydrometeorological prediction and decision making: the case of the 1997 Red River flood, Environ. Hazards 7 (2007) 342–352.
[46] NRC [National Research Council Committee on Risk Perception Communication], Improving Risk Communication, National Academy Press, Washington, DC, 1989.
[47] NRC [National Research Council Committee on Estimating and Communicating Uncertainty in Weather and Climate Forecasts], Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts, The National Academies Press, Washington, DC, 2006, http://www.nap.edu/openbook.php?record_id=11699.
[48] NSF [United States National Science Foundation], Human subjects [website], http://www.nsf.gov/bfa/dias/policy/human.jsp, 2013; see also http://www.nsf.gov/bfa/dias/policy/docs/45cfr690.pdf.
[49] NWS [National Weather Service], Service assessment and hydraulic analysis: Red River of the North 1997 floods, http://www.nws.noaa.gov/oh/Dis_Svy/RedR_Apr97/TOC.htm, 1998.
[50] E.F. Prince, J. Frader, C. Bock, On hedging in physician—physician discourse, in: R.J. di Pietro (Ed.), Linguistics and the Professions, Ablex Publishing, Norwood, NJ, 1982.
[51] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013, http://www.R-project.org/.
[52] S.G. Rabinovich, Measurement Errors and Uncertainties: Theory and Practice, Springer-Verlag, 2000.
[53] A. Rice, 'Everybody makes errors': the intersection of De Morgan's logic and probability, 1837–1847, Hist. Philos. Logic 24 (2003) 289–305.
[54] B. Rieger, Fuzzy computational semantics, in: Joint Japanese–European Symposium on Fuzzy Systems, Ser. 3, vol. 8, 1994, pp. 197–217.
[55] J.M. Sadock, Truth and approximation, in: Berkeley Linguistics Society Papers, vol. 3, 1977, pp. 430–439.
[56] P. Schulte, H. Schlager, H. Ziereis, U. Schumann, S.L. Braughcum, F. Deidewig, NO_x emission indices of subsonic long-range jet aircraft at cruise altitude: in situ measurements and predictions, J. Geophys. Res. 102 (D17) (1997) 21431–21442.
[57] R.A. Serway, Physics for Scientists and Engineers: With Modern Physics, Saunders Golden Sunburst Series, Harcourt School Publishers, 1990.
[58] S. Shackley, B. Wynne, Representing uncertainty in global climate change science and policy: boundary-ordering devices and authority, Sci. Technol. Human Values 21 (1996) 275–302.
[59] J. Siegrist, S. Ferson, A. Finkel, Advanced bias correction: factoring out bias and overconfidence, manuscript, 2014.
[60] M. Smithson, M. Smithson, Fuzzy Set Analysis for Behavioral and Social Sciences, Springer-Verlag, New York, 1987.
[61] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, H.L. Miller (Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2007, see especially "Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties', http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-uncertaintyguidancenote.pdf.
[62] D. Spiegelhalter, M. Pearson, I. Short, Visualizing uncertainty about the future, Science 333 (2011) 1393–1400.
[63] W.T. Tucker, S. Ferson, A. Finkel, D. Slavin (Eds.), Strategies for Risk Communication: Evolution, Evidence, Experience, Annals of the New York Academy of Sciences, vol. 1128, Blackwell Publishing, Boston, 2008.
[64] I.B. Türkşen, Measurement of membership functions and their acquisition, Fuzzy Sets Syst. 40 (1) (1991) 5–38.
[65] T.S. Wallsten, D.V. Budescu, A. Rapoport, R. Zwick, B. Forsyth, Measuring the vague meanings of probability terms, J. Exp. Psychol. Gen. 115 (1986) 348–365, see also http://www.dtic.mil/dtic/tr/fulltext/u2/a196944.pdf.
[66] C. Weiss, Expressing scientific uncertainty, Law, Probability & Risk 2 (2003) 25–46, http://lpr.oxfordjournals.org/content/2/1/25.full.pdf.
[67] D. Westerståhl, Generalized quantifiers, in: E.N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, 2011, http://plato.stanford.edu/archives/sum2011/entries/generalized-quantifiers/.

[68] M. Yadroff, L. Billings, The syntax of approximative inversion in Russian (and the general architecture of nominal expressions), in: Ž. Bošković, S. Franks, W. Snyder (Eds.), Formal Approaches to Slavic Linguistics: The Connecticut Meeting, Michigan Slavic Publications, Ann Arbor, MI, 1998, pp. 319–338.
[69] L. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[70] L.A. Zadeh, A fuzzy set-theoretic interpretation of linguistic hedges, J. Cybern. 2 (1972) 4–34.
[71] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, Comput. Math. Appl. 9 (1983) 149–184.
[72] S.S. Zumdal, D.J. DeCoste, Basic Chemistry, Brooks/Cole, Cengage Learning, Belmont, CA, 2011.