Probablement, Wahrscheinlich, Likely? A Cross-Language Study of How People Verbalize Probabilities in Icon Array Visualizations

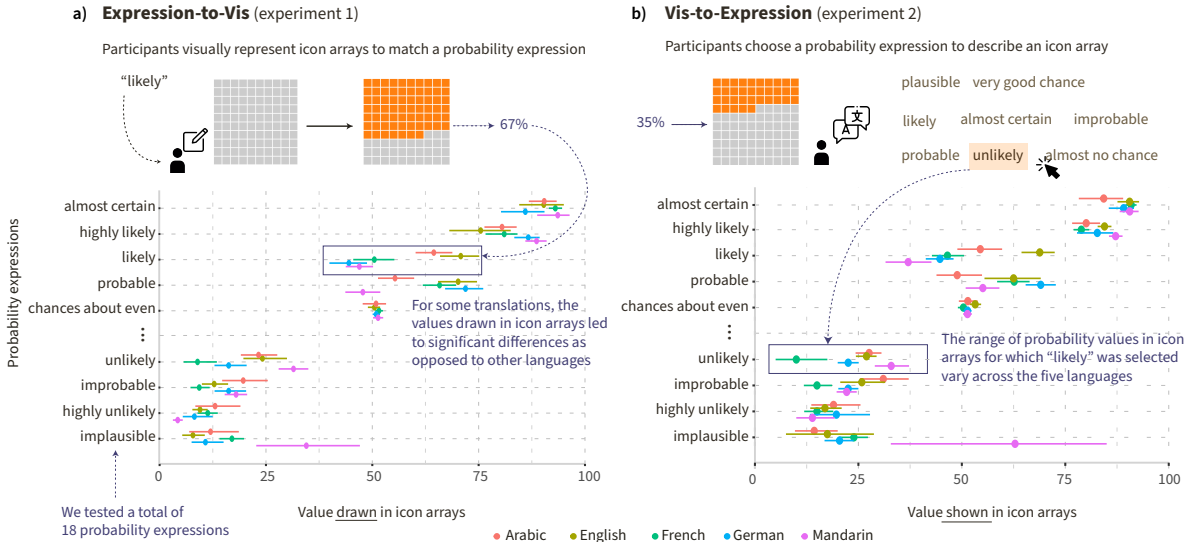Noëlle Rakotondravony, Yiren Ding , and Lane Harrison

Fig. 1: An overview of our cross-language study consisting of two experiments. In (a) **Expression-to-Vis**, we explore how people draw icon-array visualizations when given particular probability expressions, showing 95% CIs mean responses across Arabic, English, French, German, and Mandarin. In (b) **Vis-to-Expression** we investigate how people choose probability expressions to describe icon-array visualizations depicting different values. 95% CIs show the icon-array ranges to which each expression was assigned by participants, across the studied languages.

Abstract— Visualizations today are used across a wide range of languages and cultures. Yet the extent to which language impacts how we reason about data and visualizations remains unclear. In this paper, we explore the intersection of visualization and language through a cross-language study on estimative probability tasks with icon-array visualizations. Across Arabic, English, French, German, and Mandarin, $n = 50$ participants per language both chose probability expressions — *e.g. likely*, *probable* — to describe icon-array visualizations (Vis-to-Expression), and drew icon-array visualizations to match a given expression (Expression-to-Vis). Results suggest that there is no clear one-to-one mapping of probability expressions and associated visual ranges between languages. Several translated expressions fell significantly above or below the range of the corresponding English expressions. Compared to other languages, French and German respondents appear to exhibit high levels of consistency between the visualizations they drew and the words they chose. Participants across languages used similar words when describing scenarios above 80% chance, with more variance in expressions targeting mid-range and lower values. We discuss how these results suggest potential differences in the expressiveness of language as it relates to visualization interpretation and design goals, as well as practical implications for translation efforts and future studies at the intersection of languages, culture, and visualization. Experiment data, source code, and analysis scripts are available at the following repository: `https://osf.io/g5d4r/`.

Index Terms—Visualization, Cross-Language Study, Icon-Arrays

---◆---

1 INTRODUCTION

English remains the dominant language in the study and practice of visualization, but the landscape is changing. Creators from diverse languages and cultures are producing more visualizations, in part due to better access to visualization authoring tools and publishing ecosystems. International newsrooms are producing visualization-laden journalism,

• N. Rakotondravony, Y. Ding, and L. Harrison are with Worcester Polytechnic Institute. E-mail: {ntrakotondravony | yding5 | ltharrison}@wpi.edu .

and citizens on social media share and discuss visualizations in their native languages. Global challenges highlight the importance of better understanding of the interplay of language, culture, and visualization. Climate change, pandemics, and misinformation— all will require a global collective engagement with data to navigate.

Efforts in Human-Computer Interaction show how effects of language and culture might emerge in visualization. Several HCI studies and design guidelines that focus on WEIRD populations (Western, Educated, Industrialized, Rich, and Democratic) have failed to generalize when considering other languages and cultures [42]. Other studies have shown that HCI design processes can successfully integrate language and culture as an influence for interaction mechanisms and interface design [13]. More broadly, some research agendas have included multicultural populations and their needs from the outset [37]. These findings raise questions for the visualization community: to

what extent do studies centering on WEIRD populations generalize to broader global populations?

There are comparatively few studies examining language and culture in visualization. A notable exception is Kim *et al.*'s study of color names across languages [24], which found that some languages have more distinct names within certain color ranges, potentially influencing visualization color palette design. Related to visualization, studies have found that different languages and culture impact the use of color [19], and visual forms of how people represent time [15].

Studies focusing on language and statistics offer a promising means for exploring visualization across languages. In a widely replicated study, Kent showed that intelligence analysts gave different numerical estimations to the probability expressions used in intelligence reports, such as *"likely"* or *"almost certain"* [23] (see Figure 2). Later studies examined probability expressions across languages. Renooij and Witteman elicited numerical values for probability expressions with Dutch speakers [38]. Doupnik *et al.* studied how German and English-speaking accountants interpret verbal probability expressions in International Accounting Standards, finding significant differences depending on language [10]. Recently, visualization researchers Henkin and Turkay extended similar methodologies to study expressions related to correlation estimation, concluding with an explicit call to examine possible effects across languages [21].

In this paper, we explore the intersection of languages, probability expressions, and visualization. Extending the expression-to-probability methodologies of Kent [23], Renooij and Witteman [38], and others, we have participants specify values given an expression by drawing icon-array visualizations. We then invert this procedure by having participants choose expressions for a given icon-array visualization, for a two-part randomized within-subjects study. We collect expressions from prior studies, resolving issues like phrase asymmetry, ending with a list of $n = 18$ base expressions in English. To extend to other languages, specifically French, German, Arabic, and Mandarin, we recruit native speakers in each language for a collaborative translation activity, using inter-coder agreement measures to finalize a set of translations. Using the crowdsourcing platform Prolific, $n = 250$ participants ($n = 50$ native speakers for each language) completed both Vis-to-Expression and Expression-to-Vis sections.

Results suggest that people vary in how they visualize a given probability expression, with differences both within and across languages. Across languages, participants appear to agree more (*i.e.* the response ranges are tighter) when given expressions that indicate higher and lower probability values, such as *very good chance* and *highly unlikely*. Exceptions exist between languages, however, with some expressions producing substantially different value ranges from corresponding expressions in other languages, see Figure 5. People also vary in how they choose expressions when given an icon-array visualization. In Arabic for example, participants chose 15/18 possible expressions when given an icon-array depicting a 40% chance, compared to 7/18 expressions for Mandarin-speaking participants, see Figure 9. Additional analyses between experiments reveal differences in elicitation method, where people across languages tended to draw values for a given expression that were more extreme, while less extreme values were common when expressions were chosen for a given icon-array, see Figure 10.

Taken together, the experiments and results reveal substantial differences in the expressiveness of translated language as it relates to how people interpret visualizations. We discuss these findings, and how such differences may impact aspects of the visualization design process, particularly as it relates to communication or visualization translation efforts. Among other findings, we contribute:

- Evidence of no clear mapping between drawn visualizations and probability expressions across languages, suggesting cross-language differences, see Figure 5.
- Results suggesting that different languages can exhibit varying degrees of expressiveness for associated icon-array visualizations, see Figure 9.
- Cross-language experiment materials in 5 languages and datasets reflecting judgments of $n = 250$ participants, including 4,500 *Expression-to-Vis*, and 4,750 *Vis-to-Expression* judgments.
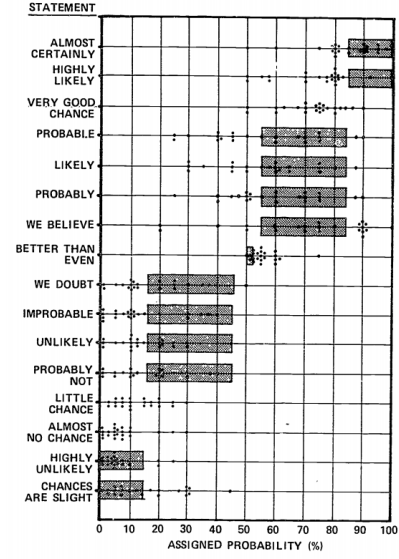

Fig. 2: Results from Kent's survey of 23 intelligence officers. Each dot is a probability assigned to an expression. The shaded areas indicate scale ranges that Kent proposed for the verbal expressions [2]. For comparison, we superimpose the shaded areas on our results, see Figure 5.

2 BACKGROUND

In crafting an experiment targeting estimative probability expressions spanning multiple languages, we draw on methodologies from studies on statistics and language, as well as considerations from cross-cultural studies in the HCI community. For design choices related to the icon-array visualizations, we refer to several visualization studies using icon-arrays in various contexts.

2.1 Probability Expression Interpretation

Numerical formats can facilitate probability comparison [46]. Yet because probability is not understood in the same way by everyone, studies have shown that numbers can provide illusory precision [7]. Other studies have shown that people sometimes prefer to communicate uncertainty using verbal expressions in conversation [48]. These verbal expressions of probability, however (*e.g.* "*highly likely, probable*"), can be interpreted differently. In a survey of 23 intelligence officers, Kent found variation in the numerical values and ranges that participants assigned to probability expressions that were commonly used in intelligence reports [23]. Kent's work is an early study of the interpretation of probability expressions, and highlighted the uneven relationship between the meaning that a communicator intends and the meaning that the audience may perceive (see Figure 2). We aim to see if this effect can be found for visualizations of similar values, and whether different patterns are found in languages beyond English.

Empirical studies on the numerical estimation of probability expressions have used elicitation methods comprising word-to-number translations [7], number-to-word conversion [36], and rank ordering of expressions [30, 38]. In these studies, probability expressions are generally studied by giving people probabilistic outcomes for specific scenarios. Results from these studies have isolated several potential factors that may impact how people understand, assess, and communicate probabilistic data. For example, the combination of verbal and numerical formats like percentage, frequency or numerical range, have been shown to aid peoples' probabilistic reasoning [6,48]. Other studies have explored factors related to culture and language. Doupnik and Richter find that German accountants' interpreted probability expressions in international accounting standards as reflecting significantly lower values than that of their American counterparts [11]. Follow up studies have speculated that this may reflect differences in cultural values where German accountants express more conservatism and stronger risk-avoidance [4, 9].
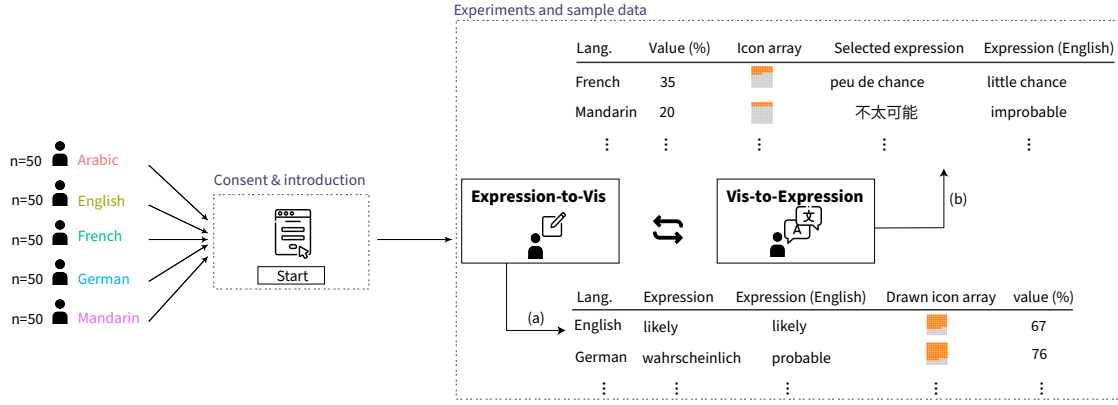
Fig. 3: The methodology that we follow in our studies. After signing the consent form and viewing an introduction, participants start the study with either Expression-to-Vis or Vis-to-Expression. **(a)** and **(b)** point respectively to sample data sets collected from the two experiments.

2.2 Uncertainty Visualization and Icon-Arrays

Visual depictions of probability are widely considered to be effective in communicating uncertainty, aiding audiences of different backgrounds to improve decisions, trust and judgment [34, 41]. Today, uncertainty visualization is widely studied and applied in both scientific domains [5] and in communication with general audiences [22]. One of the most common approaches in uncertainty visualizations implements *frequency framing*, in which the probabilistic information are displayed in frequency or ratio format [40]. In a frequency-based representation, the chance of occurrence of an event is shown as a part-to-whole proportion, considered to align better with how people naturally think of probability [20, 44]. Studies have found that uncertainty visualizations using frequency formats tend to be effective in communicating risks, especially for people with low numeracy [16, 17, 34, 50].

The icon-array is a common visualization type that implements frequency framing. Icon-arrays typically include one shape (or icon) repeated a number of times, with some of the shapes colored or otherwise marked to represent a proportion (*e.g.* 35/100). Several studies have shown that icon-arrays are an effective method for communicating risk, such as simple ratio-based probability values [33]. This part-to-whole representation of proportion reflects the frequency of events and chances, providing a visual affordances for audiences with different statistical and visualization experience to grasp. Studies have identified potential additional benefits of icon arrays including increased accuracy risk estimation tasks [31], reduced denominator neglect [17], and better understanding of medical risk severity [16].

In this study, we identify a direct connection between probability expressions from studies that focus on statistics and language, and icon-array visualizations. Building on icon-array designs used in prior work, we use $10x10$ icon arrays and similar encoding methods, following studies by Kreuzmair *et al.* [26], Bancilhon *et al.* [1], Garcia-Retamero *et al.* [17], and Galesic *et al.* [16].

2.3 Visualizations, Text, and Language

While there is little prior work in visualization across languages and culture, topics of visualizations and text or visualizations and natural language provide perspectives that inform the current work. Verbal, text-based answers visualization tasks are a common methodology, from Cleveland and McGill's graphical perception experiment where people specify a text-based answer [8], to open-ended conversational methodologies such as Peck *et al.*'s study of visualization perceptions in a local farmer's market [35]. Visualization is also considered in context with the text that surrounds it, for example Ottley *et al.* combined text and visualizations for Bayesian reasoning tasks [32]. Kong *et al.* show that the text surrounding a visualization impacts how people engage with the visualization itself [25].

The alignment between the perception of data through visualization and the language used to talk about the data has been sought by researchers who investigate the expressiveness of data visualizations.

In their study of the "Words of Estimative Correlation", Henkin and Turkay analyzed utterances and verbal descriptions from experiment participants to find how people reason and talk about different levels of correlation seen in scatter-plot visualization [21]. Their study highlights variations between how people use correlation terms to describe a visualizations and how they actually choose to visualize the terms or phrases. Drawing on this and other prior work, we focus on language in the sense of probability expressions, translating them across multiple languages, and determining how people associate these expressions with icon-array visualizations (and vice versa).

2.4 Studies Across Languages and Culture in HCI

Studies in human-computer interaction have investigated the impact of languages and culture on interface design and user behavior. In an online experiment, Baughan *et al.* found differences in how Japanese and American participants navigate websites, leading to concrete guidelines in which information might be presented across cultures [3]. Similarly, Evers showed that peoples' understanding of a graphical interface can be influenced by their cultural experience and language, with implications for interface metaphor design and interpretation [13]. Examining the transferability of primarily Western models of design in African contexts, Winschiers and Bidwell conduct information design activities with indigenous populations in South Africa and Namibia. Their findings surface Afro-centric paradigms which can shape interface design, with themes including cultural values such as interconnectedness, spirituality, and language used through oral and performed communication [47]. More closely related to visualization, Gibson *et al.* analyzed the World Color Survey in 110 languages and show that the number of color names are related to how often colors are used within a given culture. They also noted an effect of industrialization, where color becomes an essential part of the identification of objects, impacting how well people identify and name certain colors [19].

While it can be argued that computer interfaces are more widely distributed throughout languages and cultures than data visualizations, visualization appears to be on the rise as well. These findings in human-computer interaction suggest that the visualization community could be doing more to question its assumptions about the universality of approaches and guidelines, particularly as data becomes more global. We aim to take another step towards this goal by designing a study— similar in spirit to the internationalized HCI-focused studies of LabintheWild [37]— to examine probability expressions in relation to icon-array visualizations, across multiple languages.

3 METHODOLOGY

Designing a cross-language visualization study requires addressing several challenges, primarily centered around translation, but also typical concerns such as participant scenarios/prompts, visual encoding design, and interaction. We begin with a baseline methodology, extended from probability expression studies including Kent [23], Renooij and

3

Table 1: List of all expressions used in the studies and their translations. The supplemental materials include this table augmented with results from the experiments, including mean assigned values and bootstrapped 95% CIs for the *Expression-to-Vis* experiment, and the ratio of usage of the expressions for each icon-array for the *Vis-to-Expression* experiment.

	English	French	German	Mandarin	Arabic
1	plausible	plausible	plausibel	貌似可信	معقول
2	almost certain	presque certain	ziemlich sicher	几乎确定	شبه مؤكد
3	highly likely	fort probable	sehr wahrscheinlich	极有可能	من المرجّح جدّا
4	very good chance	de très grandes chances	sehr gute Chance	很有可能	احتمال كبير
5	probable	probable	wahrscheinlich	可能	محتمل
6	likely	possible	möglich	或许	مرجّح
7	probably	probablement	vermutlich	也许	من المحتمل
8	chances better than even	plus d'une chance sur deux	überdurchschnittliche Chancen	超过一半概率	أكثر من متساوي
9	chances about even	chances à peu près égales	ungefähr gleiche Chancen	大约一半	شبه متساوي
10	chances less than even	moins d'une chances sur deux	unterdurchschnittliche Chancen	不到一半	أقل من متساوي
11	probably not	probablement pas	wahrscheinlich nicht	可能不会	من غير المحتمل
12	improbable	improbable	unwahrscheinlich	不太可能	
13	unlikely	invraisemblable		未必	من غير المرجّح
14	little chance	peu de chance	kleine Chance	没什么几率	فرصة ضئيلة
15	almost no chance	presque aucune chance	fast chancenlos	几乎没概率	تقريبا لا توجد فرصة
16	highly unlikely	très peu probable	sehr unwahrscheinlich	极不可能	من المستبعد جدا
17	chances are slight	les chances sont faibles	die Chancen sind gering	机会渺茫	فرص ضعيفة
18	implausible	peu plausible	nicht plausibel	难以置信	غير المعقول

Witteman [38], and Henkin and Turkay's study in the visualization community [21]. These inform two experiments described here that investigate (1) how people visually represent a given expression through icon-arrays, across multiple languages and (2) how people across languages choose expressions to describe a particular icon-array. Figure 3 shows an overview of our experiment methodology.

Click **here** ⑦ for details and instructions.

Your chance of drawing an orange tile and winning a prize can be described as:

"implausible"

Click on the icon array to color the distribution of orange vs gray tiles so that your chance of drawing an orange tile matches the probability expression above:
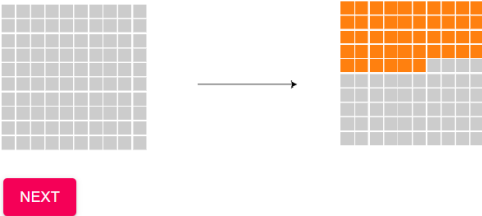

NEXT

Fig. 4: A screenshot of the instructions for experiment 1 *Expression-to-Vis* in English. The orange-colored icon array indicates a sample answer by the participant, which we evaluate numerically as 46%. Participants can access the instructions at anytime during the experiment.

3.1 Participant Prompts/Context

In general, studies targeting probabilistic reasoning and uncertainty give participants a specific context that defines the nature of the task. Such framings are known to impact peoples' behavior [45]. Visualization studies have explored a range of scenarios, from the relatively neutral "when is my bus coming?" [22] to the charged "what is the chance that someone has cancer?" [32]. Because studies have found that language and culture can impact how people perceive risk [10], we adapt a neutral context of a game which we introduce at the beginning of each experiment as follows:

> You are participating in a game which consists of drawing a tile from a set at random. Some of the tiles are orange, and some are gray. The game has two possible outcomes:
> - You draw an orange tile and you win a prize
> - You draw a gray tile and you do not win a prize

3.2 Selecting and Translating Probability Expressions

In studies targeting language and statistics, various ranges of probability expressions have been used. For example, Kent began with five and expanded to 16 [23], while Renooij and Witteman use the seven expressions most suggested by the study participants [38]. Methods for collecting expressions include scanning prior literature, eliciting expressions from study participants [38], and borrowing from specific documents [6, 11]. While studies have examined aspects of estimative probability expressions in other languages, such as Dutch [38] and German [10], we did not identify comprehensive expression lists for all our target languages. We began with Kent's original list of 16 expressions, as they have been adapted across several studies, (*e.g.* [4, 11, 30, 36]). Through pilot studies, we identified two ambiguities that impacted the symmetry of the list. We add "chances less than even" a complement to "chances better than even" and "implausible" to match "plausible". As a result, for this study, we use a list of 18 expressions, and provide options for participants to specify their own if none fit what they would prefer to choose during the *Vis-to-Expression* experiment.

Our goal was to conduct the study in English, French, German, Mandarin, and Arabic. For each language, we recruited three independent translators who were native speakers (US, France, Germany, China, Saudi Arabia, and Tunisia), and also fluent in English. To mitigate the potential for literal (word-by-word) translation, we reminded translators to consider the study scenario during translation, and to provide terms they would use in their native language within the given context.

To measure agreement, we use inter-rater reliability metrics (*e.g.* [29]) and calculate the Fleiss' Kappa κ values [14] for the translations. While $\kappa_{French} = 0.55$, and $\kappa_{German} = 0.585$ are similar, $\kappa_{Mandarin} = 0.187$ and $\kappa_{Arabic} = 0.341$ are lower in agreement. In terms of counts,

the number of expressions for which all 3 translators disagree include $disag_{french} = 3$, $disag_{german} = 1$, $disag_{mandarin} = 8$, and $disag_{arabic} = 4$. The numbers of expressions for which all translators agreed are $ag_{french} = 8$, $ag_{german} = 8$, $ag_{mandarin} = 1$, and $ag_{arabic} = 3$.

Given the high levels of disagreement in Mandarin (*i.e.* only 9/18 expressions had two people agreeing), we engaged native speakers for possible explanations. One potential reason that arose from this discussion is that a group of expressions in the source language (English) can map to a group of expressions in Mandarin in an interchangeable way. For instance, {"probable", "likely"} and {"improbable", "unlikely", "probably not"} were translated to Mandarin as {可能, 很可能, 大概,也许} and {不太可能, 可能不会, 未必}. In such cases, personal preferences might play a role in word/phrase selection.

Next we resolve disagreements in the translation. In cases where two agree, we take the majority as the final expression. When all translators disagree, we use a mediation procedure until agreements are reached about the expressions [29]. However, providing a single translation to each English expression was not always feasible. For example, "unlikely" and "improbable" were both repeatedly translated as "unwahrscheinlich" in German, and "probably not" and "improbable" were both translated as من غير المحتمل in Arabic. In these cases, we reduce the number of expressions in the target language, and mark them accordingly in results. Overall, we expect that some of these differences and similarities in languages will be reflected in the experiment results. Table 1 shows the final list of probability expressions used in this study with their translations in French, German, Arabic, and Mandarin.

3.3 Visual Encodings for the Icon-Arrays

Visualization studies have suggested that the type and arrangement of an icon-array can impact reader perception and engagement with the underlying data [40, 50]. To align with the scenario described in our study, we use a 10 x 10 grid of square icons, a typical ratio in the literature for problems with a population of 100 items (used in *e.g.* [32] and [49]) . Color is used to denote icons representing different outcomes in the event of interest: *Drawing an orange square and winning a prize*, and icons were arranged consecutively.

3.4 Participants and Procedure

Study participants were recruited using the online platform Prolific. Participants must have the target language of the experiment defined as their native language in their Prolific profile, and could only take the experiment once. In each study, Vis-to-Expression and Expression-to-Vis were assigned in random order, where half of the participants see Experiment 1 first while the other half see Experiment 2 first. 50 participants were recruited for each version of the study, making a total of 250 participants for all languages. Participants were paid $1.90 based on pilot studies estimating a 10-15 minute completion time, and following Prolific platform guidelines. Pay was fixed, *i.e.* it was not impacted by participant choices or performance in the experiment.

4 EXPERIMENT 1: EXPRESSION-TO-VIS, FROM PROBABILITY EXPRESSIONS TO ICON-ARRAYS

In this experiment, we aim to understand how people visualize probability expressions through icon arrays, and how that varies across English, French, German, Mandarin, and Arabic. Unlike existing studies about numerical estimation of probability expressions, we ask participants to represent their estimations graphically.

4.1 Procedure

Participants see an initial icon array with only gray icons, along with a probability expression describing their chance of *picking an orange tile and winning a prize*. They are asked to click or click-and-drag to draw the proportion of orange icons that matches the given probability expression. Figure 4 shows an example question in Experiment 1. The Arabic, and German versions have 17 questions, whereas the English, French, and Mandarin version have 18 questions. For each question, we collect the data format highlighted in the sample data in Figure 3 (a).
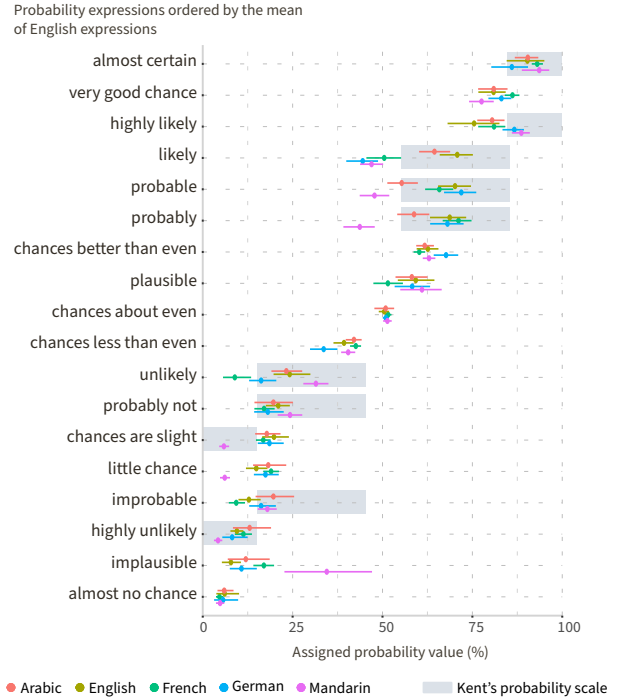

Fig. 5: Bootstrapped 95% confidence intervals of mean responses in *Expression-to-Vis*. The range of responses across the five languages are tighter for expressions indicating high and low values. Shaded areas indicate the scale range of probabilities proposed by Kent [23].

4.2 Results

In total, we collected 4,400 answers from participants across English, French, German, Mandarin, and Arabic. Because there were two instances where translators agreed about a 2-to-1 mapping from an English expression to the target languages, we duplicate these confidence intervals and perform statistical comparisons separately for each. These include the entries for unwahrscheinlich for the English expressions "unlikely" and "improbable", and من غير المحتمل for the English expressions "probably not" and "improbable".

Figure 5 shows bootstrapped 95% confidence intervals of the means for all expressions across the five languages. There appears to be general alignment with Kent's suggested ranges, though some such as highly likely and improbable deviate somewhat. This may be due to sample differences, *i.e.* Kent studied 23 intelligence analysts in the 1960s. More generally, we notice several differences for a given expression across languages. For example, visualizations drawn for expressions aligning with "likely", "probable" and "probably" in French, German, Mandarin and Arabic deviate lower than English, in some cases below the 50% mark.

While we generally align our analyses with recommendations in the VIS and HCI communities to move beyond dichotomous statistics [12], we provide statistical comparisons here to go along with analysis shown in Figure 6. Our aim is to identify expressions that are substantially above or below the associated English translations. While comparisons between other languages are possible, we focus on English since the expressions were originally translated from English.

Analyzing the between-language variance of participant-drawn visualizations with a one-way ANOVA, we find that only five expressions do not show at least one significant difference across the five languages. These stable expressions include "plausible" (and its translations: "plausible", "plausibel", "معقول", "貌似可信"), "almost certain" (presque certain, ziemlich sicher, شبه مؤكد, 几乎确定), "chances about even" (chances à peu près égales, ungefähr gleiche Chancen, شبه متساوي,

Expression-to-vis
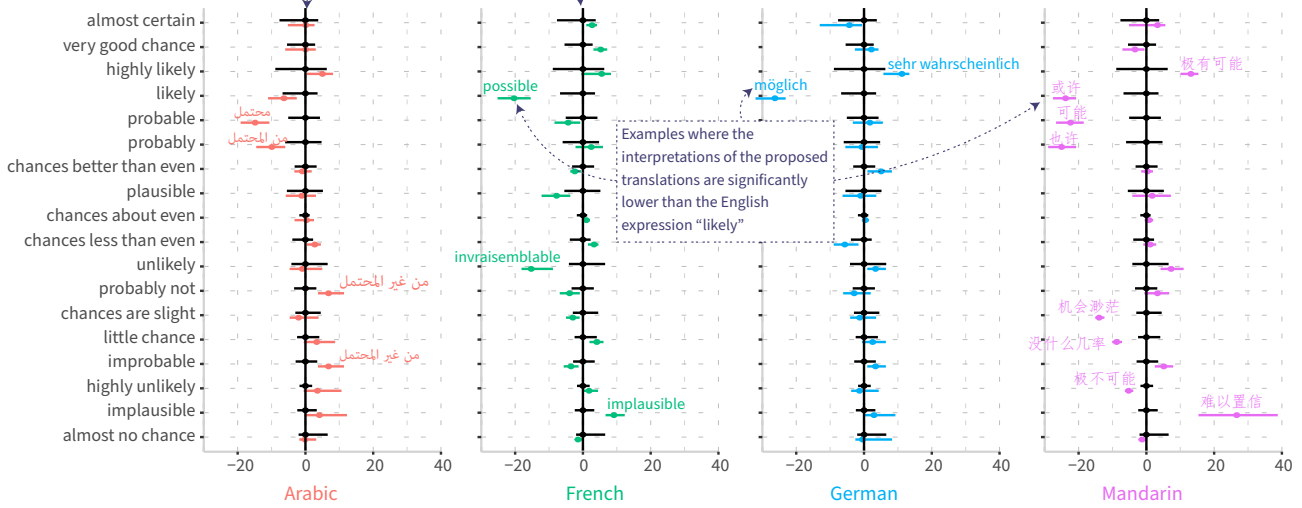
Comparison centered around the mean value for English

Fig. 6: 95% confidence intervals for differences in mean of English expressions and their translations. Values are centered on the mean value in English. Intervals on the left indicate that the mean value for an expression is lower than its associate in English. We label translations that are significantly misaligned with the English expressions. In German, entries for "unlikely" and "improbable" are duplicated (they repeatedly translated as the same German unwahrscheinlich); similarly in Arabic , entries for "improbable" and "probably not" are duplicated (they are translated in the same Arabic expression من غير المحتمل).

大约一半), "probably not" (probablement pas, wahrscheinlich nicht, من غير المحتمل, 可能不会), and "almost no chance" (presque aucune chance, fast Chancenlos, تقريبا لا توجد فرصة, 几乎没概率).

To further analyze differences between languages, we use Tukey posthoc tests to identify pairs where the expressions significantly differ from English. All reported confidence intervals are bootstrapped 95% CIs. In French, we find two deviating expressions: possible (mean: -20.32, 95% CI: [-29.23, -11.40], p.adj = 1.69E-8, English: likely), and invraisemblable (mean: -15.26, 95% CI: [-23.80, -6.75], p.adj = 1.65E-5, English: unlikely).

Two expressions in Arabic also deviate from the English counterpart: محتمل (mean: 14.84, 95% CI: [6.24, 23.44], p.adj = 3.52E-5, English: probable), and من المحتمل (mean: 9.92, 95% CI: [0.83, 19], p.adj = 0.0246, English: probably).

For German, we find three deviating expressions: sehr wahrscheinlich (mean: 11.2, 95% CI: [2.2, 20.19], p.adj = 0.0064, English: highly likely), möglich (mean: -26.3, 95% CI: [-35.22,-17.38], p.adj = 3.03E-13, English: likely), and unterdurchschnittliche Chancen (mean: -5.66, 95% CI: [-10.88, -0.44], p.adj = 0.026, English: chances less than even)

Mandarin has the highest number (seven) expressions that differ from English: 极有可能 (mean: 13.14, 95% CI: [4.15, 22.13], p.adj = 0.00075, English: highly likely), 可能 (mean: -22.4, 95% CI: [-31, -13.8], p.adj = 9.5E-11, English: probable), 或许 (mean: -23.84, 95% CI: [-32.76, -15], p.adj = 3.02E-11, English: likely), 也许 (mean: -25.02, 95% CI: [-34.10, -15.93], p.adj = 7.76E-12, English: probably), 没什么几率 (mean: -8.78, 95% CI: [-15.13, -2.42], p.adj = 0.0017, English: little chance), 机会渺茫 (mean: -13.94, 95% CI: [-19.9, -7.98], p.adj = 6.44E-9, English: chances are slight), and 难以置信 (mean: 26.66, 95% CI: [13.9, 39.42], p.adj = 2.76E-7, English: implausible).

These results show multiple instances where participants in a particular language consistently draw icon-arrays that align with different probability ranges than the associated English expression, both above and below. We will discuss possible reasons behind these differences, including implications for visualization, in section 7. Interestingly, translations for the duplicated entries both in German and Arabic did not significantly differ from the original English expression. This sug-

gests that the suggested translation in German and Arabic do align with both expressions in English.

Other patterns in the analysis reflect possible drawing affordances in the icon-array that persist across languages. Most responses end in 0 or 5, similar to rounding behavior in graphical perception studies [43]. These include 51% of total answers for English, 54% for French, 52% for German, 45% for Arabic, 48% for Mandarin. This pattern also aligns with previous results where people tend provide numerical estimations that are multiples of 10 for verbal probabilities [27, 39] (see Figure 7)

We can also view each base expression from the perspective of its range across languages, see Figure 7. For example, although there is no particular pattern across the five languages for the expressions, the Mandarin translation of "implausible" shows the largest IQR with 91%, as well as a bimodal distribution of responses. Expressions with narrow ranges suggest that, across associated expressions in other languages, participants will draw similar ranges in icon-arrays. Another pattern is that expressions near extreme low/high and center values appear to have smaller interquartile ranges. Expressions at the extremes are consistently evaluated, while mid-range expressions (but not central) may convey less precise estimates of probability.

5 EXPERIMENT 2: VIS-TO-EXPRESSION, FROM ICON-ARRAYS TO PROBABILITY EXPRESSIONS

This experiment is essentially an inversion of Experiment 1. Here, we aim to understand how people choose probability expressions to describe a given icon-array visualization.

5.1 Procedure

Using the same neutral scenario, participants are given an icon-array of a specific value, along with a list of probability expressions (see Figure 8. Participants are asked to select the expression that they believe best describes the icon-array shown. Expression lists consist of 18 expressions in English, French, and Mandarin, 17 expressions in German, and Arabic. Participants are also encourage to provide their own answer, if desired. To cover the probability space, we encode 19 values between 5% to 95% with a step of 5%. For each trial, we collect the pair {icon-array value, selected probability expression}.
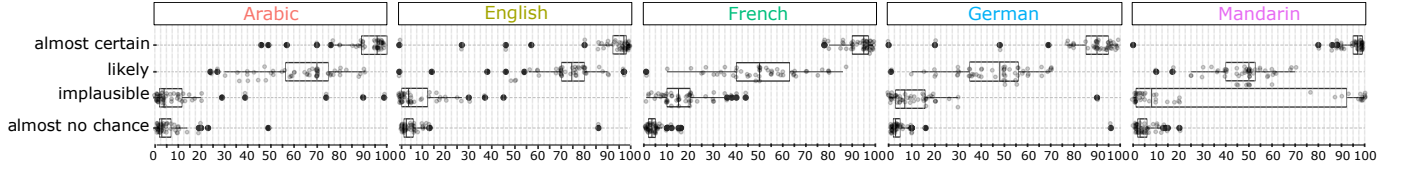
Fig. 7: Range plots of example expressions and values drawn on the icon-arrays (Expression-to-Vis). There seems to be rounding behavior around multiples of 0 and 5. Value ranges for some expressions also vary across translations (*e.g.* "likely" differs between English and other languages).

Fig. 8: A screenshot of the instructions for experiment 2 *Vis-to-Expression* in German. Participants were also given the option to input their own expression to describe the icon-array.

5.2 Results

In total, we collect 4,750 icon-array to expression pairs from participants across the five languages (50 participants per language). Icon arrays across steps of 5% are described fifty times by participants, either by selecting an expression from the suggested list, or by typing their answers. Similar to Expression-to-Vis, we duplicate entries for expressions that map to two expressions in English. These include من غير المحتمل in Arabic and unwahrscheinlich in German. Results for Vis-to-Expression are shown in Figure 9.

The upper bar graph in Figure 9 shows the count of unique expressions for each icon-array value. Across languages, there appears to be consistency in that higher values (90%, and 95%) are described using fewer unique expressions. This may reflect a more consistent expressiveness of languages of probability expressions in higher ranges. However, given the average counts across all value possibilities, it is clear that few, if any perfect matches of probability expressions and visualization exists.

Arabic has the highest average number of unique expressions. On average, 10.89 expressions (stdv = 2) were used to describe each probability value. For example, for icon-arrays of 40%, participants in Arabic chose 15 out of 17 possible expressions. 10 of these expressions were selected at least twice. Looking back at the Expression-to-Vis results in Figure 5, there appears to be a gap above 25% and below 45%, although أقل من متساوي or "chances less than even" did end up being the most selected expression for this value (19 times).

In contrast, Mandarin has the lowest average count of unique expressions per icon array 8.10 (stdv = 1.73) than the other versions of the experiment. Implausible is a notable exception, which was rarely used and had a wide range associated with it. To a lesser degree, German, French, and English show the central values around 50% with a low number of unique expressions (below their averages), while mid-range values above and below 50% show more variance.

Figure 9 also shows the intersection of icon-array values and expression counts. The size of the circles indicate the ratio at which an expressions has been used to describe a given value.

Looking vertically, there appears to be variance in how icon-arrays depicting particular values are described, with some expressions preferred over other. For example, values around 30-40% and 60-80%

occasionally show multiple possible preferred expressions. Across all values and expressions, larger circles appear to exist for central and high probability values, reflecting some of the consistency seen in the first experiment. People across languages seem to have clear preferences for translations of "chances about even" for an icon-array at 50%. Other patterns show variance and disagreement. The translations for "probable", "likely" appear to be used to specify a large range of probability values in a high range. Notably, however, similar patterns do not appear to exist across languages for wide ranges of low probability values. For instance, while "implausible" translations potentially fits this low-and-wide range role is German, French, and Arabic, it is scarcely chosen at all in English and Mandarin.

During the experiment, participants had the option to suggest additional probability expressions whenever needed. Participants provided their own expressions 55 times for Arabic, 99 for English, 145 for French, 65 for German, and 14 for Mandarin. Example trends in these include people using the listed expressions in a full sentences (*e.g.* Il est *improbable* de gagner ("it is improbable to win")), or with varying qualifiers (*e.g.* فرصة عالية، فرصة عالية جدا ("high chance, very high chance")). Among the new expressions and phrases that were suggested, we notice some referring directly to the proportion shown in the icon array arrangement (*e.g.* Etwa jeder Vierte gewinnt for a 25% icon array). We provide these data in the supplement, as an extended analysis of written answers by participants may give an opportunity to explore additional language or cultural factors that people refer to when making judgments about icon-array visualizations.

6 COMPARING EXPRESSION USAGE ACROSS EXPERIMENTS

Given the within-subjects design of both experiments, it is possible to make comparisons across them. In experiment 1 *Expression-to-Vis*, people were given each expression and drew a specific icon-array design (stored as a percentage value). Similarly, In experiment 2 *Vis-to-Expression*, icon-arrays of particular values were given and participants chose expressions. Shown in Figure 10, one observation between the two experiments is that people tend to draw extreme values for given expressions, but when people are given icon-arrays depicting similar values, they choose different expressions. A concrete example is that when someone receives an expression *"almost no chance"*, they typically draw an icon array with very few colored squares. However, when showed with an icon-array with more colored squares than are typically drawn for this expression, people still describe this icon-array as "almost no chance".

With these values for each participant, we can also explore how consistent people are in assigning expressions to visualizations versus visualizations to expressions. While several distance metrics are possible, in an exploratory analysis we define a straightforward distance measure for each probability expression as *Distance D = exp2vis_val - mean(vis2exp_val)*. Where *exp2vis_val* is the value that a participant drew for an expression, and *mean(vis2exp_val)* is the average of all values for which participants assigned this expression. A large distance indicates that a participant is less consistent their mapping of language to visualization, while a low distance suggests more consistency. We find several instances where participants exhibit consistency across the two experiments, and other instances where participants differ between themselves widely. Example visualizations and data for these are included in the supplement. These also motivate potential individual-level modeling efforts for future work.
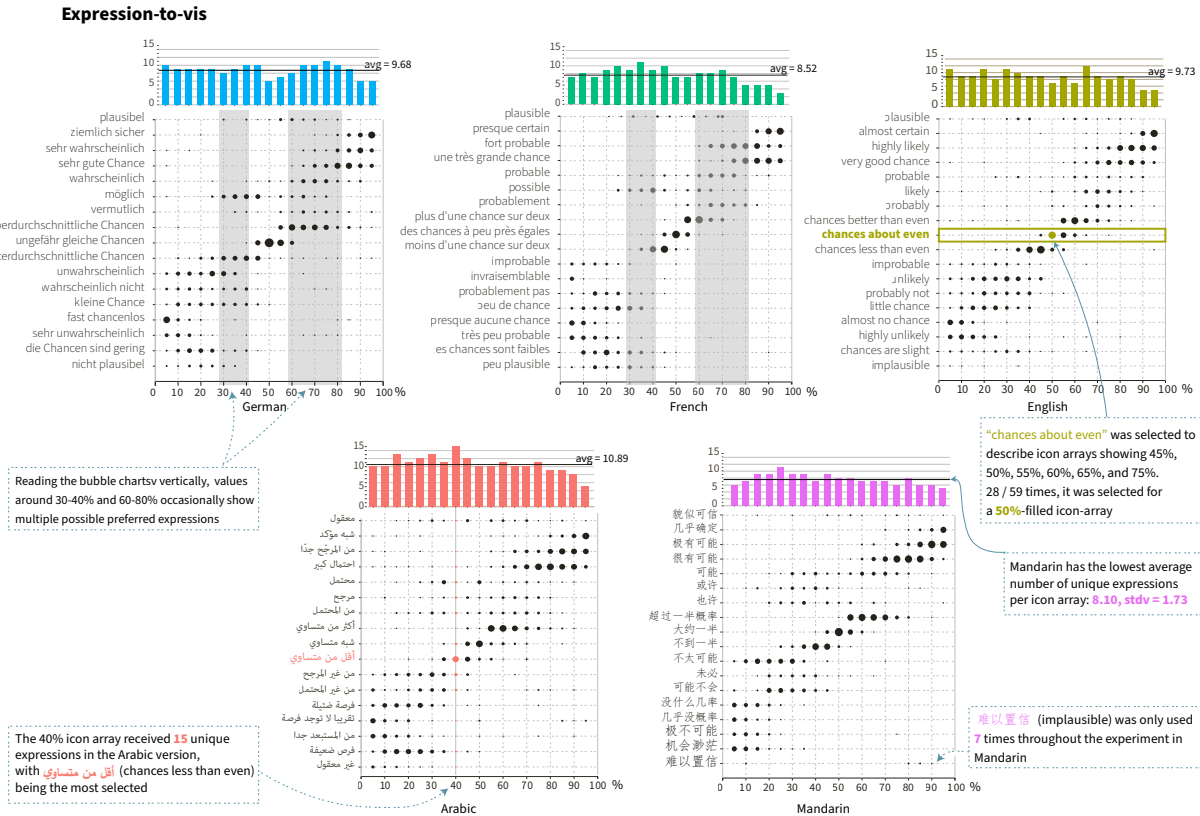
7

Fig. 9: Results from experiment 2 *Expression-to-Vis*. Barcharts represent the number of unique expressions participants selected for a given icon-array. Bubbles indicate count of each probability expression and value pair. Results show similarities and differences across languages.

7 DISCUSSION

The results of the *Expression-to-Vis* and *Vis-To-Expression* experiments suggest that language plays an important role in the specification and interpretation of icon-array visualizations. Differences are found within languages. For example, in Mandarin, there was no overlap in Kent's suggested scale range of 55%-85% for the expressions *likely, probable*, and *probably* and the visualizations participants drew when given translations of these expressions.

There are other differences across languages. Using English as a source translation, we identified instances in every tested language that deviated significantly above or below the ranges for the associated English expression (Figure 5). These and other reported findings raise questions about the interplay between language expressiveness, visualization, and translation efforts.

7.1 Implications for Visualization Translation

The observed differences across languages may hold implications for visualization translation efforts. Consider a case where a climate change visualization originally crafted in English, but is intended to be translated into other languages for broader distribution. Translation may use a similar visualization technique, but requires substantial changes to the accompanying textual elements to facilitate target audiences' understanding of the communicated data in the specific context of the visualization. Results here suggest that estimative probability expressions, which may be used both in visualizations themselves and in texts referencing the visualization, should be carefully translated so as to preserve the intended numerical range. Difficulties in translation may persist across both translators and automatic translation tools, such as Google Translate. Future studies might build on these results and others examining the impact of text on visualization interpretation, *e.g.* [25], to investigate the extent to which alignments and misalignments between expressions and visualizations impact peoples' judgments.

The translation process itself is another factor that may play a role

in the observed differences. Although multiple native speakers were consulted and mediation procedures were used to reach agreement, it is possible that other speakers would have generated somewhat different expressions. Even with other translators, however, there is no guarantee that the chosen expressions would have matched the ranges participants chose for the source expression. A different resulting expression list would also not necessarily cover equal spans of the probability value ranges in the target language.

These translation challenges raise one possible application of the methodology and the results here. It may be possible to *align* expressions based on the resulting participant-driven probability ranges. For example, while the expression plausible in French differs significantly from the English plausible, another expression in French, plus d'une chance sur deux (originally translating *chances better than even*), does align better with this range in the observed data. Computational methods could be designed to construct these translations for the tested languages and others. As an initial exploration, we iteratively paired up expressions across two languages until we found pairs that do not significantly differ following a Tukey test comparison. The results of this approach showing multiple possible "aligned" translations are available in supplemental material.

Computational approaches to language-expression alignment could address challenges in cross-language statistical reporting scenarios. The IPCC (Intergovernmental Panel on Climate Change) report, for example, specifies guidelines for its writers to use certain probability expressions for certain ranges [6]. As medical tests and associated symptom displays also rely on an intersection of icon-arrays and language (*e.g.* [17, 18, 32]), these efforts may also aid medical or pandemic risk communication. Our results provide a possible path towards further refining these standards, helping ensure the intended meaning of statistics and charts is communicated faithfully across languages.

7.2 Towards Better Elicitation for Cross-Language Visualization

Methodology was a key challenge in the study. While we aimed to carefully adapt and extend prior studies to begin exploring the intersection of visualization and languages, possibilities emerged through the design and resulting analysis. In the *Vis-to-Expression* experiment, participants offered 86 (out of 4750) additional expressions. While some of these overlap with existing expressions in the study, it is likely that there are other expressions or phrases that the studied languages and associated cultures use in talking about probabilistic events. Finding ways to elicit these could be a challenging but rewarding effort for visualization. For instance, a participant in the Mandarin version commented:

> "貌似可信"在汉语中不是一个好的表达，我作为母语者都不能完全理解你们用这个词想表达什么" (It seems "plausible" is not a good expression in Chinese. As a native speaker, I can't fully understand what you are trying to express with this word)

In English, "plausible" is common, but might there be other translations or similar expressions in other languages that fill a similar role?

One possibility is to move beyond English as a source, and instead develop in-language elicitation methodologies. These might be graphically based, using interaction and visualization with input capability to allow participants to specify ranges and expressions. Alternately, they may be large crowdsourced studies, following similar scenario and trial-based methods. In either case, the goal would be to elicit a wider range of expressions and visualization descriptions from participants. Such efforts could reveal additional ranges and expressive capabilities within languages, beyond those studied here.

Beyond the results presented here, there are other useful starting points for exploring possibilities in cross-language elicitation for visualization. One source would be to consider the history of large-scale color elicitation studies such as the World Color Survey [28]. Other language and statistics studies such as Renooij and Witteman [38], and the NLP-driven analysis of Henkin and Turkay [21] might inform approaches to scale.

7.3 Limitations and Future Work

One limitation of the current work is the restriction to five languages. While these were chosen as an initial step in the space, and intended to cover several major language families, there are thousands of languages in the world and various language groups that could be explored. In the context of a global pandemic, for example, it is important to support effective data-focused communication as broadly as possible. Another limitation is the use of English as a source language. While the intended focus of this study is to highlight difficulties surrounding translation, the use of English still somewhat centers findings on WEIRD populations. Other possible study designs are within reach, however. Future work might design expression elicitation methods directly for native speakers, or using both visualization and numerical probabilities, which might reveal additional expressive features between languages. Elicitation methods might also examine the language people use when describing visualizations of probability distributions, instead of individual probabilities. The neutral scenario / context given to participants and sole use of icon-array visualizations are other practical limitations to explore in future work. Cultural differences such as risk avoidance (*e.g.* [4, 10]) may be less pronounced in neutral scenarios, but become more pronounced with carefully designed contexts. We might refer to studies targeting medical reasoning (*e.g.* Ottley *et al.* [32]) or natural disaster risk (*e.g.* Padilla *et al.* [34]) for promising scenarios and visualizations to explore in future work. Finally, to more directly target implications for visual representations, future studies might explore the extent to which *different* visual representations elicit responses from native speakers, which might uncover effects of language and culture between visualizations.
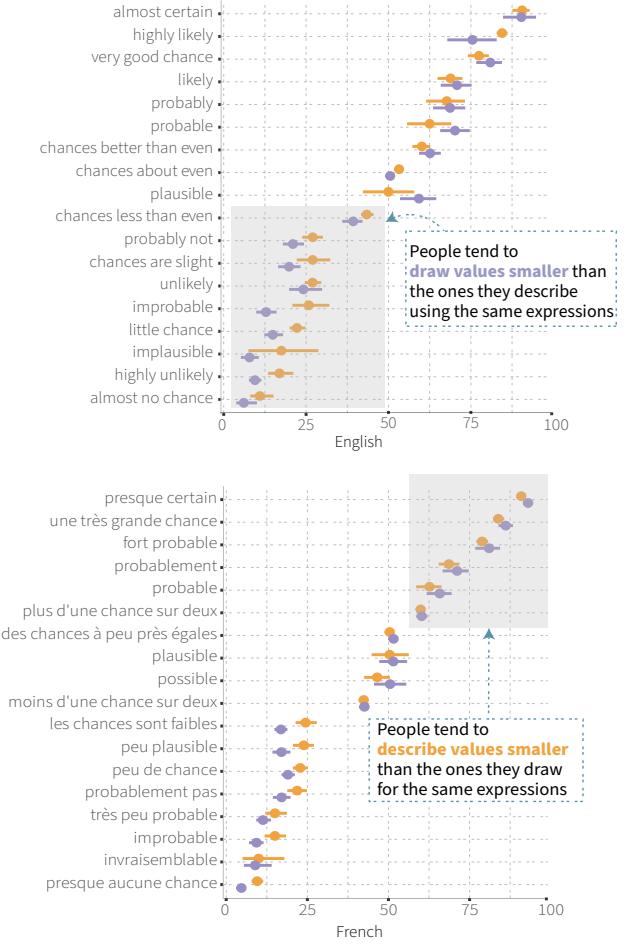

Fig. 10: Comparing the two experiments, results show differences in elicitation methodology that hold across languages. People appear to draw visualizations with more extreme values when given an expression. But when given a visualization of similar value, people tend to choose different expressions. Charts for all five languages and for individual participants are available in supplementary material.

8 CONCLUSION

With the changing global landscape of data and visualization practice, it is important that the research community explores the intersection of language and visualization. We present two experiments with the goal of understanding how people across five languages draw icon-array visualizations given probability expressions and assign probability expressions given icon-array visualizations. Results of these experiments show several differences both across and within languages, with no clear mapping across languages, and several instances of possible "gaps" between expressions in a given language. We discuss implications of these results for ongoing efforts such as data and visualization translation, targeting areas such as climate change and pandemic communication. Taken together, these studies and results are intended to offer a limited yet useful step in broadening the focus of the visualization community beyond traditional WEIRD populations.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Bancilhon, Z. Liu, and A. Ottley. Let's Gamble: Uncovering the Impact of Visualization on Risk Perception and Decision-Making. *Computing Research Repository (CoRR)*, abs/1910.09725, 2019.

[2] S. Barclay et al. Handbook for decisions analysis. 1977.

[3] A. Baughan, N. Oliveira, T. August, N. Yamashita, and K. Reinecke. Do Cross-Cultural Differences in Visual Attention Patterns Affect Search Efficiency on Websites? In *CHI '21: CHI Conference on Human Factors in Computing Systems, Virtual Event / Yokohama, Japan, May 8-13, 2021*, pp. 362:1–362:12. ACM, 2021. doi: 10.1145/3411764.3445519

[4] F. Bocklisch, A. Georg, S. Bocklisch, and J. Krems. Do You Mean What You Say? The Effect of Uncertainty Avoidance on the Interpretation of Probability Expressions-A Comparative Study between Spanish and German. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 35. cognitivesciencesociety.org, 2013.

[5] P. G. Brust-Renck, C. E. Royer, and V. F. Reyna. Communicating Numerical Risk: Human Factors That Aid Understanding in Health Care. *Reviews of Human Factors and Ergonomics*, 8(1):235–276, 2013. doi: 10.1177/1557234X13492980

[6] D. V. Budescu, H. H. Por, S. B. Broomell, and M. Smithson. The Interpretation of IPCC Probabilistic Statements around the World. *Nature Climate Change*, 4(6):508–512, 2014. doi: 10.1038/nclimate2194

[7] D. V. Budescu and T. S. Wallsten. Consistency in Interpretation of Probabilistic Phrases. *Organizational behavior and human decision processes*, 36(3):391–405, 1985.

[8] W. S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[9] T. S. Doupnik and E. L. Riccio. The Influence of Conservatism and Secrecy on the Interpretation of Verbal Probability Expressions in the Anglo and Latin Cultural Areas. *The International Journal of Accounting*, 41(3):237–261, 2006.

[10] T. S. Doupnik and M. Richter. Interpretation of Uncertainty Expressions: a Cross-national Study. *Accounting, Organizations and Society*, 28(1):15–35, 2003. doi: 10.1016/S0361-3682(02)00010-7

[11] T. S. Doupnik and M. Richter. The Impact of Culture on the Interpretation of "In Context" Verbal Probability Expressions. *Journal of International Accounting Research*, 3(1):1–20, 2004. doi: 10.2308/jiar.2004.3.1.1

[12] P. Dragicevic. Fair Statistical Communication in HCI. In *Modern statistical methods for HCI*, pp. 291–330. Springer, 2016.

[13] V. Evers. Cross-Cultural Understanding of Metaphors in Interface Design. *Proceedings CATAC'98, Cultural Attitudes towards Technology and Communication*, pp. 1–11, 1998.

[14] J. L. Fleiss. Measuring Nominal Scale Agreement among Many Raters. *Psychological bulletin*, 76(5):378, 1971.

[15] O. Fuhrman and L. Boroditsky. Cross-Cultural Differences in Mental Representations of Time: Evidence from an Implicit Nonlinguistic Task. *Cognitive science*, 34(8):1430–1451, 2010.

[16] M. Galesic, R. Garcia-Retamero, and G. Gigerenzer. Using Icon Arrays to Communicate Medical Risks: Overcoming Low Numeracy. *Health Psychology*, 28(2):210–216, 2009. doi: 10.1037/a0014474

[17] R. Garcia-Retamero and M. Galesic. Communicating Treatment Risk Reduction to People with Low Numeracy Skills: A Cross-Cultural Comparison. *American Journal of Public Health*, 99(12):2196–2202, 2009. doi: 10.2105/AJPH.2009.160234

[18] R. Garcia-Retamero and M. Galesic. Who Profits from Visual Aids: Overcoming Challenges in People's Understanding of Risks. *Social Science and Medicine*, 70(7):1019–1025, 2010. doi: 10.1016/j.socscimed.2009.11.031

[19] E. Gibson, R. Futrell, J. Jara-Ettinger, K. Mahowald, L. Bergen, S. Ratnasingam, M. Gibson, S. T. Piantadosi, and B. R. Conway. Color Naming Across Languages Reflects Color Use. *Proceedings of the National Academy of Sciences*, 114(40):10785–10790, 2017.

[20] G. Gigerenzer and U. Hoffrage. How to Improve Bayesian Reasoning without Instruction: Frequency Formats. *Psychological Review*, 102(4):684–704, 1995. doi: 10.1037/0033-295X.102.4.684

[21] R. Henkin and C. Turkay. Words of Estimative Correlation: Studying Verbalizations of Scatterplots. *CoRR*, abs/1911.12793, 2019.

[22] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is my Bus? User-Centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. *Conference on Human Factors in Computing Systems - Proceedings*, pp. 5092–5103, 2016. doi: 10.1145/2858036.2858558

[23] S. Kent. Words of Estimative Probability. *Journal of the American Intelligence Professional*, 8(4):49–65, 1964.

[24] Y. Kim, K. Thayer, G. S. Gorsky, and J. Heer. Color Names Across Languages: Salient Colors and Term Translation in Multilingual Color Naming Models. In *EuroVis (Short Papers)*, pp. 31–35, 2019.

[25] H.-K. Kong, Z. Liu, and K. Karahalios. Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018.

[26] C. Kreuzmair, M. Siegrist, and C. Keller. High Numerates Count Icons and Low Numerates Process Large Areas in Pictographs: Results of an Eye-Tracking Study. *Risk Analysis*, 36(8):1599–1614, 2016. doi: 10.1111/risa.12531

[27] D. Leclercq. J'en Suis Aussi Sûr que Vous, Mais pas avec le Même Pourcentage de Chances, que ce Soit Hors Contexte ou En Contexte. *Journal international de Recherche en Education et Formation Evaluer. Journal international de Recherche en Education et Formation*, 2(21):89–125, 2016.

[28] D. T. Lindsey and A. M. Brown. World Color Survey Color Naming Reveals Universal Motifs and their Within-Language Diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(47):19785–19790, 2009. doi: 10.1073/pnas.0910981106

[29] N. McDonald, S. Schoenebeck, and A. Forte. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019. doi: 10.1145/3359174

[30] F. Mosteller and C. Youtz. Quantifying Probabilistic Expressions. *Statistical Science*, 5(1):2–12, 1990. doi: 10.1214/ss/1177012242

[31] Y. Okan, R. Garcia-Retamero, E. T. Cokely, and A. Maldonado. Improving Risk Understanding across Ability Levels: Encouraging Active Processing with Dynamic Icon Arrays. *Journal of Experimental Psychology: Applied*, 21(2):178–194, 2015. doi: 10.1037/xap0000045

[32] A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. Han, and R. Chang. Improving Bayesian Reasoning: The Effects of Phrasing, Visualization, and Spatial Ability. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):529–538, 2016. doi: 10.1109/TVCG.2015.2467758

[33] L. Padilla, S. C. Castrob, and H. Hosseinpoura. A Review of Uncertainty Visualization Errors: Working Memory as an Explanatory Theory. *The Psychology of Learning and Motivation*, p. 275, 2021.

[34] L. Padilla, M. Kay, and J. Hullman. Uncertainty Visualization. *Handbook of Computational Statistics and Data Science*, 2020.

[35] E. M. Peck, S. E. Ayuso, and O. El-Etr. Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. *arXiv*, pp. 1–12, 2019.

[36] R. T. Reagan, F. Mosteller, and C. Youtz. Quantitative Meanings of Verbal Probability Expressions. *Journal of Applied Psychology*, 74(3):433–442, 1989. doi: 10.1037/0021-9010.74.3.433

[37] K. Reinecke and K. Z. Gajos. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW 2015, Vancouver, BC, Canada, March 14 - 18, 2015*, pp. 1364–1378. ACM, 2015. doi: 10.1145/2675133.2675246

[38] S. Renooij and C. Witteman. Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, 22(3):169–194, 1999. doi: 10.1016/S0888-613X(99)00027-4

[39] C. A. Sanne Willems and I. S. Verbal. Variability in the Interpretation of Probability Phrases Used in Dutch News Articles — a Risk for Miscommunication. *Orphanet Journal of Rare Diseases*, 21(1):1–9, 2020.

[40] M. M. Schapira, A. B. Nattinger, and C. A. McHorney. Frequency or Probability? A Qualitative Study of Risk Communication Formats Used in Health Care. *Medical Decision Making*, 21(6):459–467, 2001.

[41] D. Spiegelhalter, M. Pearson, and I. Short. Visualizing Uncertainty about the Future. *Science*, 333(6048):1393–1400, 2011. doi: 10.1126/science.1191181

[42] C. Sturm, A. Oh, S. Linxen, J. Abdelnour Nocera, S. Dray, and K. Reinecke. How WEIRD is HCI? Extending HCI Principles to Other Countries and Cultures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, vol. 18, pp. 2425–2428, 2015. doi: 10.1145/2702613.2702656

[43] J. Talbot, V. Setlur, and A. Anand. Four Experiments on the Perception of Bar Charts. *IEEE transactions on visualization and computer graphics*, 20(12):2152–2160, 2014.

[44] C. Till. Fostering Risk Literacy in Elementary School. *International Electronic Journal of Mathematics Education*, 9(1-2):85–98, 2014.

[45] V. H. Visschers, R. M. Meertens, W. W. Passchier, and N. N. De Vries. Probability Information in Risk Communication: A Review of the Research Literature. *Risk Analysis*, 29(2):267–287, 2009. doi: 10.1111/j. 1539-6924.2008.01137.x

[46] T. S. Wallsten, D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth. Measuring the Vague Meanings of Probability Terms. *Journal of Experimental Psychology: General*, 115(4):348–365, 1986. doi: 10.1037/0096-3445. 115.4.348

[47] H. Winschiers-Theophilus and N. J. Bidwell. Toward an Afro-Centric Indigenous HCI Paradigm. *International Journal of Human-Computer Interaction*, 29(4):243–255, 2013.

[48] B. C. Wintle, H. Fraser, B. C. Wills, A. E. Nicholson, and F. Fidler. Verbal Probabilities: Very Likely to be Somewhat More Confusing than Numbers. *PLoS ONE*, 14(4):1–18, 2019. doi: 10.1371/journal.pone.0213522

[49] C. Xiong, A. Sarvghad, D. G. Goldstein, J. M. Hofman, and Ç. Demiralp. Investigating perceptual biases in icon arrays. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2022.

[50] B. J. Zikmund-Fisher, H. O. Witteman, M. Dickson, A. Fuhrel-Forbis, V. C. Kahn, N. L. Exe, M. Valerio, L. G. Holtzman, L. D. Scherer, and A. Fagerlin. Blocks, Ovals, or People? Icon Type affects Risk perceptions and Recall of Pictographs. *Medical Decision Making*, 34(4):443–453, 2014. doi: 10.1177/0272989X13511706