Advances in Survival Analysis and Optimal Scaling Methods

Willems, S.J.W.

Variability in the interpretation of probability phrases used in Dutch news articles - a risk for miscommunication

5

Verbal probability phrases are often used in science communication to express estimated risks in words instead of numbers. In this study we look at how laypeople and statisticians interpret Dutch probability phrases that are regularly used in news articles. We found that there is a large variability in interpretations, even if the phrases are given in a neutral context and even among statisticians. We conclude that experts and the media should be careful in using verbal probability expressions.

5.1 Introduction

Every day people make decisions based on estimated probabilities and risks. These decisions range from choices with little consequences (*"Should I bring my umbrella to avoid getting wet in the rain?"*) to important life decisions (*"Which treatment will most likely cure my cancer without causing too many undesirable side-effects?"* or *"Should I evacuate for the approaching storm?"*). Many of our decisions rely on risks expressed by others (weather forecasters, oncologists, or scientists). Due to this dependence of the decision maker on the information provider, it is important that the message is understood as intended in order to minimize the risk of miscommunication.

Many estimated probabilities are communicated verbally and in that case it is important that the interpretation of the verbal probability phrase is the same for both sender and receiver. For example, an oncologist may predict, based on his or her experience, that there is a 90–95% probability of a cure for a particular patient. The doctor may then use the expression *very likely* to communicate this

This chapter is submitted as Willems, S. J. W., Albers, C. J., and Smeets, I. *Variability in the interpretation of probability phrases used in Dutch news articles - a risk for miscommunication.*

Probability term	Subjective Probability range
Almost certain	99–100%
Extremely likely	95–99%
Very likely	90–95%
Likely	66–90%
About as likely as not	33–66%
Unlikely	10–33%
Very unlikely	5–10%
Extremely unlikely	1–5%
Almost impossible	0–1%

Table 5.1: *Approximate probability scale recommended for harmonized use in the European Food Safety Authority (EFSA) to express uncertainty about questions or quantities of interest (European Food Safety Authority et al., 2019).*

probability. However, if the patient interprets this verbal probability expression as a 75% probability, he or she may unnecessarily have lower expectations.

Some organizations and bureaucrats use probability scales as in Table 5.1 to standardize their risk communication. But how well do these translations match with how people actually interpret these phrases?

Early studies on the interpretation of probability phrases typically used the 'how likely' approach; respondents were asked to give their interpretation of a probability expression as a single value or range on a scale of 0–1 or 0–100%, or were asked to rank them. The phrases were either presented out-of-context or in sentences describing a particular situation. Many of these studies were summarized in the literature reviews by Druzdzel (1989) and Visschers et al. (2009), and the meta-analysis by Theil (2002).

The overall conclusion from these studies was that, although individuals seem to be internally consistent in their ranking of probability phrases (Budescu and Wallsten, 1985) and their perception of them over time (Bryant and Norman, 1980), the interpretation of these phrases varies greatly among individuals. This interpretation variability is especially large for phrases expressing a probability in the range from 20% to 80%. For words that express extreme probabilities, such as *always*, *certain*, *never*, and *impossible*, consensus was highest. This variability of interpretations is represented by the varying widths of the subjective probability ranges in probability scales as in Table 5.1. These wide ranges complicate communication, because it is impossible to express a very specific probability.

Several studies also showed that the numerical interpretations of some prob-

ability phrases overlap or are very similar. For example, Reagan et al. (1989) concluded that *likely* is synonymous with *probable*, and *low chance* with *unlikely* and *improbable*. Synonymous words have overlapping probability ranges which would complicate a probability scale. The codification presented in Table 5.1 seems to avoid this complication by limiting the vocabulary to phrases with nonoverlapping ranges.

Furthermore, translation issues for verbal probability expressions are important for all international organizations that publish their documents in more than one language. For example, a question that may arise within the European Food Safety Authority is whether their probability scale (Table 5.1) translates directly to other European languages, or whether the subjective probability ranges in the second column should be adjusted, and consequently, the expressions in the documentation text.

Most research on the numerical interpretation of probability phrases was conducted in English. There have been some replication studies in other languages, among which the Dutch language. Most of the Dutch studies are over twenty years old. For instance, Eekhof et al. (1992) focused on the interpretation of 30 Dutch phrases. However, all phrases in this study expressed frequencies instead of probabilities. In a later study by Timmermans (1994) some probability phrases were included, usually in combination with an adverb like *quite* or *rather*. Unfortunately, the paper is written in English and does not provide the Dutch expressions used in the study, hence it is unclear exactly which Dutch expressions and adverbs were investigated. In a study by Pander Maat and Klaassen (1996), focus was on the interpretation of uncertainty in information leaflets that come with medicine. Although their main interest was not in the numerical values associated with verbal probability phrases, they did investigate this for three phrases. Renooij and Witteman (1999) did several experiments to develop a probability scale containing both words and numbers. Their focus was on ranking seven probability phrases and developing their corresponding numerical scale. Given that the first study included many phrases but only frequencies, and the other three studies included only a few probability phrases, usually in combination with adverbs, many Dutch probability expressions still needed to be studied.

In addition to replication studies in other languages, several studies have been done to compare the interpretation variability of English probability phrases with the interpretations of their translations to other languages. Three studies, comparing English with French (Davidson and Chrisman, 1994), German (Doupnik and Richter, 2003), and Chinese (Harris et al., 2013), showed that on average the numerical interpretations of the English phrases differ from the interpretation of their counterparts in the three other languages. Additionally,

in French and Chinese, the standard deviations of the numerical values related to the probability phrases were much larger than those of the original English wording.

These results show that the meaning of probability expressions can get lost in translation from one language to another.

5.2 Background

5.2.1 The communication mode preference paradox

Until recently, it was generally believed that information providers, the senders of a message, prefer to express probabilities verbally, namely by using verbal probability expressions as *unlikely*, *usually* and *maybe*, while decision makers favor numeric expressions like percentages. Druzdzel (1989) reasoned that senders prefer verbal expressions because these convey some amount of uncertainty. Including this uncertainty in the expression is favored by senders, because probability estimates are usually based on empirical data and therefore not sufficiently precise to be translated into exact numerical statements. Hence, if a numerical value is given its suggested precision may be misleading. On the other hand, decisions makers prefer this precision of numerical expressions, since numeric values are easier to compare and to draw conclusions from. Erev and Cohen (1990) referred to this difference in preference as the *communication mode preference paradox*. In more recent studies, researchers have challenged this theory, but the results are not conclusive. For example, Juanchich and Sirota (2019) concluded that people favor verbal phrases in general, but in some contexts or for specific purposes numerical expressions are preferred.

5.2.2 Asymmetry

A complication in the interpretation of probability phrases is asymmetry. For example, based on the discovery of the synonymous pair *low chance* with *unlikely* and *improbable*, Reagan et al. (1989) also expected *high chance* to be synonymous with *likely* and *probable*. However, their data indicated that actually *very likely* and *very probable* are its synonyms. This unbalanced result shows that there is some asymmetry in the interpretation of mirrored probability phrases.

This phenomenon of asymmetry is studied and confirmed by many researchers. In most studies, this imbalance is investigated on a group level by comparing the group means or medians of two complementary phrases. For instance, Lichtenstein and Newman (1967) concluded that the interpretations of *likely* and *unlikely* are asymmetric, since their means sum to (72% + 18% =) 90% and their medians sum to (75% + 16% =) 91% instead of 100%. This asymmetry was

confirmed by both Reagan et al. (1989) (medians sum to 90%) and Stheeman et al. (1993) (medians sum to 80%). Furthermore, Lichtenstein and Newman (1967) focused on the influence of adverbs (such as *very*, *quite* and *fairly*) and found that, for instance, the means of the numeric probabilities given to *quite likely* and *quite unlikely* sum to (79% + 11% =) 90% instead of 100%. Previous studies have also shown that some terms actually are (almost) symmetrical. For example, *very likely* and *very unlikely* (mean interpretations sum to 96%, Lichtenstein and Newman (1967)), and *almost always* and *almost never* (median interpretations sum to 98%, Stheeman et al. (1993)).

Some mirrored terms have a clear linguistic explanation for their asymmetry. For example, Mosteller and Youtz (1990) studied the terms *possible* and *impossible* and found that the interpretation of *impossible* is stable (around 3% for all participants of the study), while *possible* has distinct meanings for different groups of people. Namely, some respondents used the literal interpretation of *possible* and indicated that it could express any percentage between 0% and 100%, and others associated it with rare events that only scarcely occur (as in *barely possible*). Hence, the different interpretations of *possible* causes the strong asymmetry with its mirrored expression *impossible*. The asymmetry in the interpretation of *certain* and *uncertain* can be explained in a similar way.

The asymmetry in the interpretation of verbal probability expressions complicates the development of probability tables. For example, the symmetry of the probability scale in Table 5.1 simplifies the use of the table, but it does not necessarily represent the actual probability ranges of its terms.

All these research results show that the interpretations of verbal probability expressions vary too much to translate them into a (symmetrical) probability scale of which the numerical probability ranges would be supported by everyone. Therefore, many researchers who initially intended to make a translation table, concluded that such a codification is practically impossible (Lichtenstein and Newman (1967); Mosteller and Youtz (1990); Timmermans and Mileman (1993); Weber and Hilton (1990)), or realized that their currently used table was actually not conveying the intended probabilities (Pander Maat and Klaassen, 1996). Yet, still many organizations are using tables like this.

5.2.3 Context dependence

The interpretation of a probability phrase is influenced enormously by its context. For instance, compare your numerical interpretation of the word *likely* in the next two statements:

– *It is likely that it will rain in Manchester, England, next June;*
– *It is likely that it will rain in Barcelona, Spain, next June.*

Probably, your numerical interpretation of *likely* in the first statement is higher

than in the second. Wallsten et al. (1986) used this example and, based on their research, predicted a difference in the numerical interpretation of these statements. Namely, in their study, they showed that an individual's expected base-rate expectation of a context scenario influences this person's interpretation of the probability phrase. In this example the base-rate for the first scenario is higher (in spring rain is more probable in England than in Spain) and this influences the interpretation of the word *likely*.

This hypothesis on the base-rate effect was confirmed by Weber and Hilton (1990), who additionally provided evidence that other variables may be affecting the interpretation as well. According to their findings, the perceived severity or consequentiality of an event and its emotional valence will also influence the judged probability.

Since it was shown that context may influence the interpretation of probability phrases, many researchers decided to investigate them out-of-context. However, it was argued by Druzdzel (1989) that, if no specific context is provided, participants may invent their own context. Due to these self-created contexts, participants' responses will portray the interpretation of the probability phrases in many completely different contexts instead of out-of-context. These different scenarios may cause extra variability in the data which makes it more difficult to draw conclusions from the results.

5.2.4 Differences between sub-populations

In most studies, data on the interpretation of probability phrases was gathered within specific sub-populations. Participants were, for instance, physicians (Bryant and Norman, 1980), science writers (Mosteller and Youtz, 1990), radiologists (Stheeman et al., 1993), biological scientists (MacLeod and Pietravalle, 2017), or patients (Pander Maat and Klaassen, 1996). Although all these studies showed variability in the perception of probability phrases within these sub-populations, one might wonder whether there are any differences between these groups as well. For example, Theil (2002) argued that there may be a difference between professionals, who regularly make and communicate probability estimations, and persons who are inexperienced in this respect. However, his meta-analysis did not provide evidence for this hypothesis.

In studies on the use of jargon, it has been shown that there is a significant difference in the interpretation of medical terms between doctors and patients (Boyle, 1970) and of hydrological vocabulary between experts and laypeople (Venhuizen et al., 2019). Experts may be unaware of this difference (Castro et al., 2007) and, hence, their use of jargon may cause a miscommunication of information.

Given these results on the different interpretations of jargon, there is reason

to believe that there may be differences between the numerical interpretations of probability expressions of experts and laypeople as well, as Theil (2002) suggested. If this hypothesis is correct, experts may be misunderstood if they express probabilities verbally.

5.2.5 Gaps in the literature

Summarizing, we see that despite ongoing interest in and usage of verbal probability expressions, there are large gaps in the literature. Furthermore, in most studies on this topic, the sample sizes were quite small. For instance, the number of participants in the Dutch studies lay between 78 (Timmermans, 1994) and 101 (Eekhof et al., 1992). The English studies have comparable sample sizes, for example, in the nine studies mentioned by Theil (2002) the median number of participants is 52 and the mean is 170.

Therefore, we set up a large-scale study for the interpretation of Dutch verbal probability expressions, presented in a neutral context which are based on ordinary events. In this way, we investigate whether the variability in interpretation is also high when the context is barely susceptible to prior beliefs.

Additionally, we check for synonymous phrases and asymmetry since these two characteristics are well studied in English but have not yet been analyzed in Dutch studies. Furthermore, we compare the results of statisticians with those of laypeople to check whether experts use different interpretations.

5.3 Methods

We used a survey design were probability phrases were presented in a neutral sentence to participants, and they could give their interpretation as a point estimate on a 0–100% scale.

5.3.1 Choice of phrases

There are many Dutch probability and frequency phrases that can be studied. To make a selection for our study, we first listed the phrases used in the English studies and translated them to Dutch. For translation Google Translate (Google, 2018) and the leading Dutch dictionary Van Dale (Van Dale Uitgevers, 2018) were used. If more than one translation was appropriate, both were added to the list. Then we added the expressions from previous Dutch studies (Eekhof et al., 1992; Pander Maat and Klaassen, 1996; Renooij and Witteman, 1999). This resulted in a list of 131 phrases.

This list was too long to use in one survey, so a selection had to be made. Since the most frequently used phrases are also the most relevant, we selected

the verbal probability expressions that were used at least 100 times in all online available articles of the popular Dutch news website *nu.nl*. To prevent too much overlap with the research by Eekhof et al. (1992), only the ten most commonly used frequency phrases were selected. Furthermore, some combinations of adverbs with a probability phrase were removed from the list to prevent too much overlap with the study by Timmermans (1994), and to prevent repetitions of very similar phrases. Additionally, the word *undecided* was removed, since it was mostly used in sport results where it has a different meaning.

This method of phrase selection resulted in a list of 29 frequency and probability expressions. These phrases, and their English translations, are given in Table 5.2 of the supplementary material (subsection 5.7.1). In the text of this paper, we will use the English translations. Please keep in mind that all given numerical interpretations for these phrases are actually for their Dutch counterparts.

5.3.2 Context

As described before, the interpretation of a probability expression may be influenced by a person's prior expectations of the phrase's context. To avoid these base-rate effects, our aim was to formulate sentences that are neutral in the sense that everyone can imagine the situation but has little prior expectations about it. Some examples of the statements, formulated with the probability phrase *likely*, are

– *It is **likely** that this plan succeeds.*
– *It is **likely** that this hotel is fully booked.*
– *It is **likely** that the team wins a match.*

We tried to minimize the base-rate effect by not specifying a specific plan, hotel, or team. We developed twelve sentences like these. The complete list of these contexts is given in Table 5.2 of the supplementary material (subsection 5.7.2). In each sentence the verbal probability expression was printed in bold to direct more attention to it.

5.3.3 Numeric interpretations

For each probability expressions in the survey, participants gave the point estimate of their numerical interpretation in percentages (0–100%) by using a slider. After the statement, each survey item was formulated as a question. For example, the questions related to the three statements above were formulated as:

– *What is the probability (expressed in percentages) that this plan succeeds?*
– *What is the probability (expressed in percentages) that this hotel is fully booked?*
– *What is the probability (expressed in percentages) that the team wins a match?*

All probability phrases were presented individually and in a random order, and participants were required to answer each question before continuing to the next. In this way, missing data was prevented.

5.3.4 Randomization

To prevent a systematic influence of the context on the interpretation of the probability phrase, 12 different versions of the survey were created. In every version, the probability phrase was formulated in a different context and contexts were repeated two or three times in each survey version (since 29 is not divisible by 12). All survey versions were evenly and randomly distributed among the participants by the survey software Qualtrics (2005-2018).

5.3.5 Personal characteristics

After giving their interpretation of the 29 phrases, participants were asked for some personal information. This included whether they are a statistician, their highest completed education level, their age, and their gender. Statisticians were self-reported and this was questioned as *"Are you a statistician or do you perform statistical analyses on a weekly or monthly basis?"*. Education was categorized in six common categories of degrees in the Netherlands. Age was categorized in intervals of 20 years. These wide intervals were chosen to protect the anonymity of the participants and because the exact ages were not of particular interest for this research. However, age was included to check whether both young and older people participated. As with age, gender is not of particular interest for this study, but it was included to check whether participants are almost equally distributed among the genders.

All personal characteristics were asked as multiple-choice questions and participants could select one of the given categories. Participants were allowed to refrain from providing their age and gender.

5.3.6 Pilot

A pilot study showed that the length of the survey was reasonable (approximately ten minutes) and that the explanation was clear. We noticed that some participants had the tendency to base their interpretation of a phrase on their interpretations of previous phrases. This confirms that randomization of the phrases is necessary. Additionally, it supported our decision to present one phrase at the time and to not allow participants to change their answers to previous questions. If we had permitted this, participants may have ranked their answers instead of giving the interpretations individually, which may have influenced the results. Based on the pilot study, we decided to make the original question *"Are*

you a statistician or do you perform statistical analyses on a regular basis?" more specific by changing *"on a regular basis"* into *"on a weekly or monthly basis?"*.

5.3.7 Survey distribution

We obtained permission to distribute this survey from the ethical committee of the Faculty of Behavioural and Social Sciences of the University of Groningen in the Netherlands (17451-O). Since we wanted to compare the interpretations of Dutch-speaking statisticians with those of non-statisticians, the survey was distributed among both groups. Statisticians were invited to participate via the mailing list of the Netherlands Society for Statistics and Operations Research (VVSOR) and the Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS). To reach non-statisticians, the survey invitation was distributed via the personal Twitter (Twitter Inc., 2018) accounts of the three authors (one of the authors is a public figure and has over 60,000 followers, many of which are not in the academic community). Their followers were asked to participate and to share the survey in their network.

5.4 Results

5.4.1 Participants' characteristics

The survey was open for participation for almost four months, namely between July 18th, 2018 and November 8th, 2018. During this time, 1004 persons started the survey, of which 115 did not finish it. These incomplete observations were removed from the data. Another 8 participants were excluded from the analysis, because their native language was not Dutch. As a result, the data contains the responses of 881 participants.

The participants are evenly distributed among the genders (430 male vs. 440 female). There were many more non-statisticians than (self-reported) statisticians (655 vs. 226). Their distribution among the age groups and education levels is displayed in Figure 5.1. The first bar plot indicates that most participants were equally distributed among the two middle age groups (20–39 years and 40–60 years). The second bar plot shows that many of the participants were highly educated. This is partially explained by the fact that most statisticians have an academic education (94%), but even among the non-statisticians, the proportion of academically educated persons is large (58%). Furthermore, there are more males than females among the statisticians (59% male) and more females among the non-statisticians (55% female).

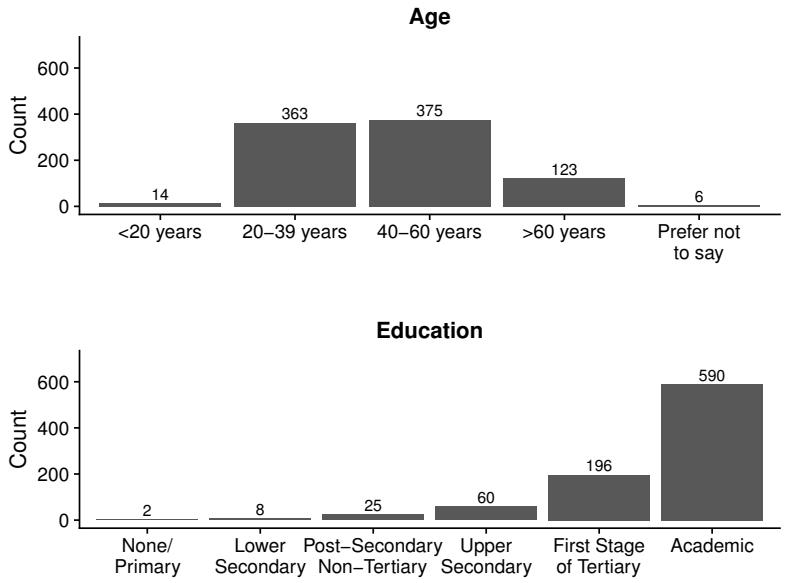

Figure 5.1: *Bar graphs of the number of participants in each of the category levels of variables* Age *and* Education.

5.4.2 Interpretation of probability phrases

The distributions of the interpreted percentages of each probability phrase are displayed by the density plots in Figure 5.2 and the mean values and 5% and 95% percentiles are listed on the right side of the plots. The 5% and 95% percentiles indicate the range of interpretations of 90% of the participants.

There seems to be some consensus about the interpretation of extreme words like *always*, *certain*, and *impossible*. Namely, the intervals between their 5% to 95% percentiles have a width of about 20 percentage points. Surprisingly, the 95% percentile of the extreme phrase *never* is at 32%, which seems high for this expression.

There is even less consensus for phrases that do not represent an extreme probability. Namely, their numerical interpretations have percentile ranges with widths up to 50 percentage points. For example, 90% of the respondents interpreted the verbal probability expressions *sometimes*, *probable*, and *almost always* between, respectively, 11–55%, 41–86%, and 70–96%.

Other things to notice are the small peaks in the density plots which indicate that participants often express probabilities as multiples of ten. Also, there was no phrase in our survey that represents 50%. The candidates *liable to happen*, *chance*, *uncertain*, *maybe*, and *possible*, for which 50% is the most frequently

chosen interpretation, all have a large tail to the left and percentile ranges of 42–50 percentage points.

5.4.3 Asymmetry

For the usability of verbal probability expressions, (a)symmetry in the interpretation of mirrored verbal probability expressions are of interest. The imbalance in their interpretation is often investigated by reviewing whether the group means or group medians of the interpretations of two complementary words sum to 100%. The groups means from our data are listed in Figure 5.2, and show that, as in English, asymmetry is present for the Dutch translations of *likely* and *unlikely*. Namely, the mean interpretation of *likely* in our data is 75% and the mean for *unlikely* is 16%, and hence these sum to 91%. Symmetry is found for phrases as *very likely* and *very unlikely* (sum to 95%), *almost always* and *almost never* (sum to 100%), and *often* and *not often* (sum to 97%).

The results from previous studies and those listed above are based on the results on a group level (group means). We also looked at the results on an individual level by plotting the density of the sums of complementary phrases, see Figure 5.3. These plots show that there are some mirrored pairs which interpretation sums up to about 100% for most participants, for example *(almost) always* and *(almost) never*, and *very likely* and *very unlikely*. Other complementary phrases were interpreted asymmetrically by many participants and usually sum up to slightly less than 75% to 100%, for example *likely* and *unlikely*, and *often* and *not often*.

As explained in the introduction, in some cases asymmetry has a linguistic cause. Our results on the interpretation of possible and impossible confirm the findings of Mosteller and Youtz (1990). Namely, Figure 5.2 shows that *impossible* has a stable interpretation that is close to 0%, while *possible* has a broad interpretation from 20% to 70% which peaks around 50%. The asymmetry is also confirmed by the distribution of their sums in Figure 5.3.

A similar pattern is found for *certain* and *uncertain*; there is a consensus on the interpretation of *certain* (around 100%) while the perception of *uncertain* varies a lot and is comparable to *maybe*'s interpretation, namely some value between 20% to 50% (see Figure 5.2). As a result, the percentages of *certain* and *uncertain* always sum to more than 100% and together peak at 150% (see Figure 5.3).

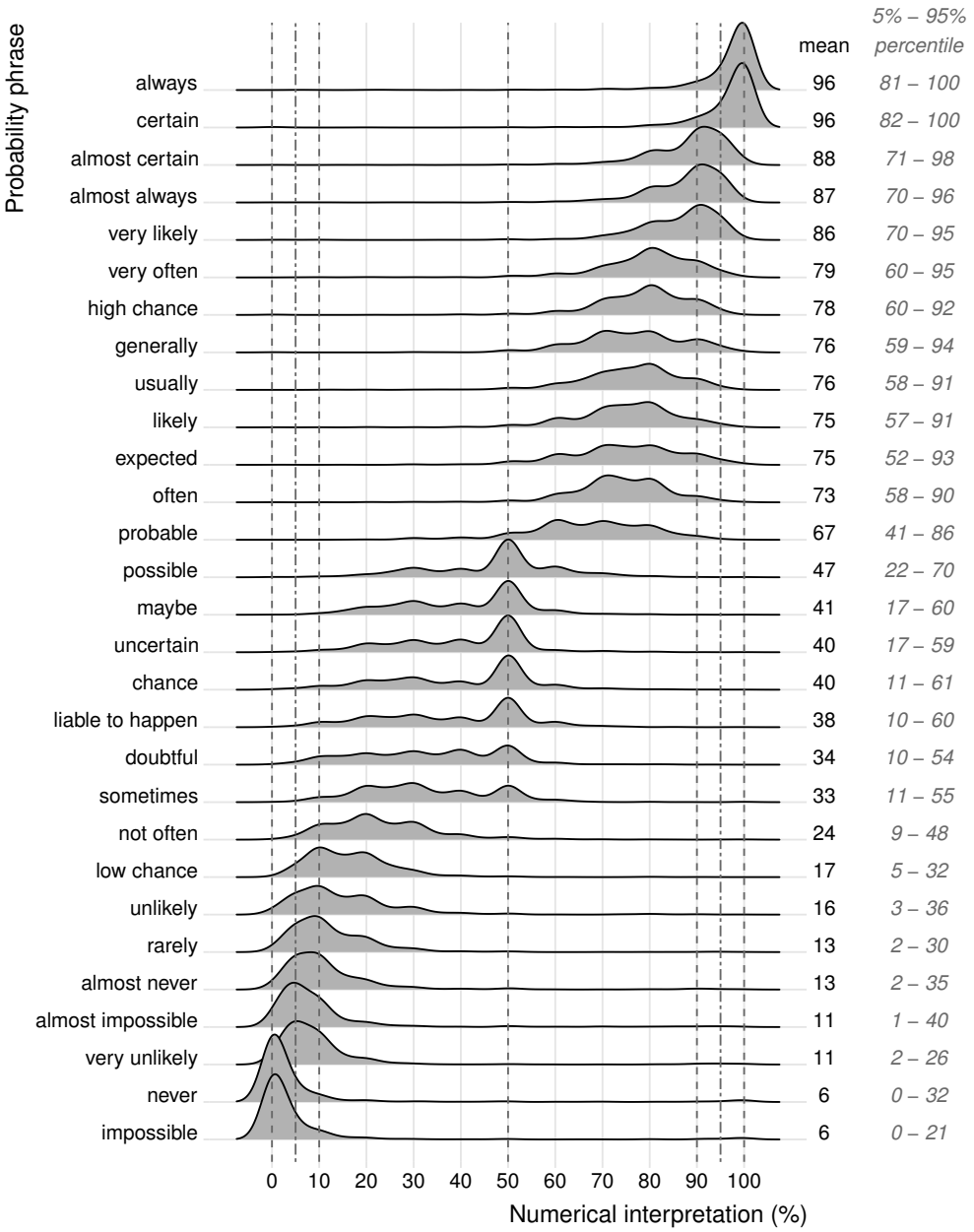

Figure 5.2: *Density plots and mean values of the numerical interpretations (in percentages) given by all participants for each phrase in the survey. Note that density plots are a smooth variant of histograms and may therefore be positive outside the data range of 0–100%.*

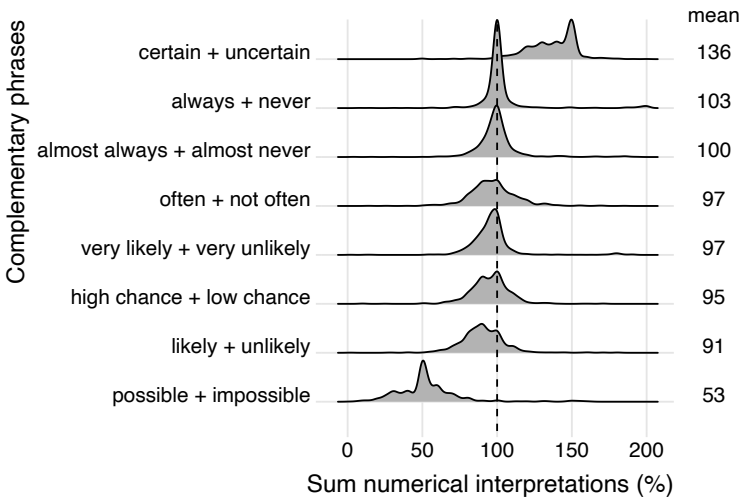

Figure 5.3: *Density plots and mean values of the sums of the numerical interpretations (in percentages) given by all participants for the complementary phrase pairs in the survey.*

5.4.4 Context

One of our concerns was that the context of the sentences influences the perception of the probability phrases. To avoid the base-rate effect, we tried to formulate the context sentences as neutral as possible.

To check whether we succeeded in our intention, we investigated the variability of the interpretation of phrases among different contexts. Figure 5.4 shows the mean percentages given by the participants to each probability phrase, grouped by context. This plot shows that, in general, the means of a phrases are very similar for each context, with a maximum of 20 percentage points difference between contexts. Most of this variability appears for words that represent 30% to 80%.

Most importantly, although the plots show some influence of context on the interpretations, they do not suggest that any of the sentences is systematically interpreted differently (higher/lower or more/less extreme) from the others.

5.4.5 Differences between sub-populations

One of the aims of this research was to make a comparison of the interpretation of probability phrases of different sub-populations, namely to compare interpretations of experts (statisticians) with those of laypeople. Figure 5.5 shows the density plots of the statisticians and non-statistician for a selection of five probability phrases. These expressions were selected from different ranges of

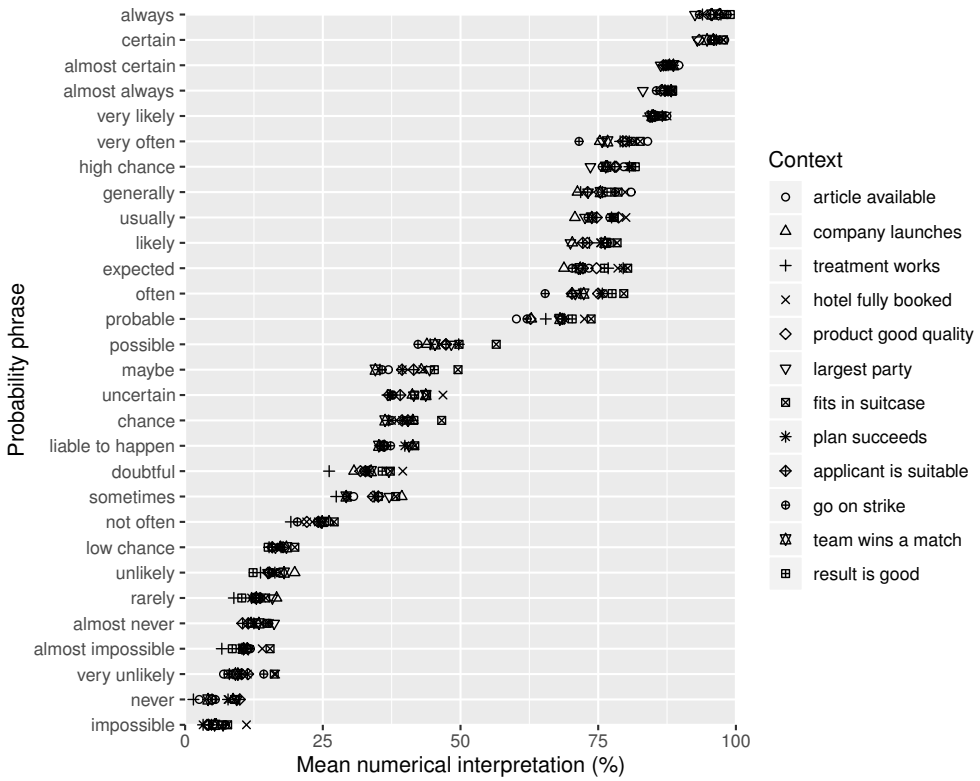

Figure 5.4: *Means of the numerical interpretations (in percentages) given by all participants for each phrase in the surveys, grouped by the context of the sentences. Listed contexts in the legend are abbreviations of the originals (see Table 5.3 of the supplementary material (subsection 5.7.2)).*

numerical interpretations. Results for all phrases are shown in Figure 5.6 of the supplementary material (subsection 5.7.3).

These density plots show that the interpretations of the probability phrases are very similar for both statisticians and non-statisticians. This similarity is represented by the overlapping regions of the plots. The nonoverlapping regions are relatively small, which suggests that there are no big differences between the groups. This is supported by the group means, since the maximum difference between statisticians and non-statisticians is four percentage points.

Although the differences are small, the density plots of *very likely* and *almost never* in Figure 5.5 may suggest that statisticians agree more on the interpretation of verbal probability expressions expressing an extreme probability. This phenomenon is also seen for other extreme phrases (see Figure 5.6 of the supplementary material, subsection 5.7.3), but not for phrases expressing

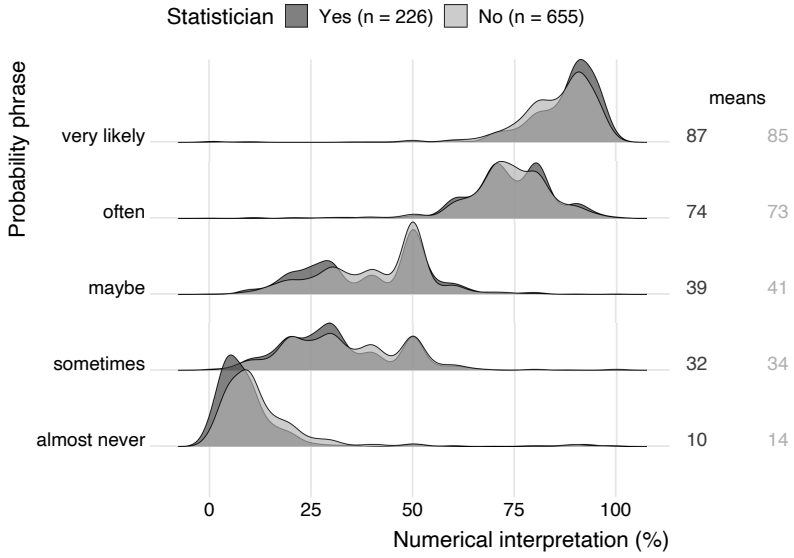

Figure 5.5: *Density plots and mean values of the numerical interpretations (in percentages) given by statisticians and non-statisticians for a selection of five phrases from the survey. Note that density plots are a smooth variant of histograms and may therefore be positive outside the data range of 0–100%.*

percentages closer to 50%. However, the difference between the group means is small for these phrases, so the group effect (if present) is weak.

We also investigated whether there were differences in responses by men and women but found no notable differences (see Figure 5.7 of the supplementary material, subsection 5.7.3).

Note that we do not statistically test for group differences because the interpretations have very irregular distributions which makes statistical testing complicated.

5.5 Discussion

In this study we have investigated the variability of the interpretation of Dutch probability and frequency phrases. The set-up of our survey was comparable to previous surveys on the interpretation of English phrases, but it filled some gaps in the research on Dutch probability phrases. For example, we included many Dutch expressions that were not studied before and represented them in a neutral context. Furthermore, we verified asymmetries in the interpretation of mirrored phrases, and checked for differences in interpretation between statisticians and non-statisticians.

Our results showed that, as in English, there is a large variability in the interpretation of Dutch probability and frequency phrases. Although there is some agreement about extreme words as *always*, *certain*, *never*, and *impossible*, there is no consensus about words that describe a less extreme probability.

As mentioned before, Eekhof et al. (1992) already studied many Dutch frequency expression. Ten of those were also included in our study so we could compare the results. For nine of these phrases, the mean interpretations differed a maximum of three percentage points. Only the interpretation of *sometimes* differed more, namely a difference of 8 percentage points (mean of 33% in our study vs. 25% in their study).

Besides comparing our results to those of previous studies in Dutch, we can verify our results with studies that included the English counterparts of the phrases in our survey. Theil (2002) listed the mean interpretations for ten probability phrases found in ten studies, seven of which overlapped with our list of verbal probability expressions. The mean interpretations that we measured for these phrases are all between the lower and upper bounds of the means measured in these ten studies. However, the ranges of those means were quite wide for some expressions. Due to this large amount of variability in the English results, it is not possible to conclude from this that there are no differences between the interpretations of Dutch phrases and their English translations.

Additionally, our data confirms the previous results on asymmetry in the interpretation of verbal probability expressions, also on an individual level. For example, usually an individual's numerical interpretations of *likely* and *unlikely* do not sum to 100%.

Previous studies in English showed that the asymmetry in the interpretation of some mirrored pairs (as *possible* and *impossible*, and *certain* and *uncertain*) has a linguistic cause. Our study confirms that similar asymmetries are found for the Dutch translations of these phrases.

Another phenomenon that has previously been shown to have an influence on the interpretations of verbal probability expressions is context. Therefore, we tried to present the expressions in neutral contexts. Although the mean interpretations varied among the contexts (see Figure 5.4), our results did not show a structural difference between contexts. Hence, there were probably no strong base-rate effects, indicating that our chosen contexts were neutral enough.

Only after analyzing our data we realized that of the twelve contexts that we used, ten presented a positive outcome (for example *this treatment will work* and *this plan succeeds*), but two presented a negative outcome (namely *they will go on strike* and *this hotel is fully booked*). Phrasing a risk positively or negatively may also influence interpretation. However, we found no structural differences between the mean interpretations of these positive and negative contexts.

To test for differences in interpretations between experts and laypeople this survey was distributed among both statisticians and non-statisticians. Our data showed large variability within each group and no structural differences between them. Hence, it seems that regularly making and communicating probability estimations does not increase agreement about the interpretations of probability expressions. This justifies our analyses on the complete sample.

The size of the complete sample (881 participants) is one of the strengths of this study. In most studies sample sizes were quite small; the mean number of participants in the previous Dutch studies listed in this paper was 93 and, for example, in the English studies mentioned by Theil (2002) it was 170.

Participants were invited via Twitter, which is a convenient way to reach a lot of people. Although this resulted in our large sample size, the convenience sample includes a disproportionate distribution of education levels and, hence, our results on the non-statisticians may not generalize to the population. However, the results from this study are still valuable, since they showed that, even within this homogeneous sample, interpretations of probability expressions differed enormously. This indicates that interpretations are dissimilar even among more like-minded persons. If the sample had been more heterogeneous, the interpretations may have varied even more.

5.6 Conclusion

From our results we conclude that the interpretations of Dutch verbal probability expressions are comparable to those of their English translations. Therefore, all challenges regarding communicating with English verbal probability expressions also apply in Dutch. For example, making a translation table from Dutch verbal probability phrases to numeric values is infeasible.

Although it was generally believed that information providers prefer to express probabilities verbally, a solution might be to convey estimated risks using either numerical values instead of verbal expressions, or both. This may prevent the intended probability from getting lost in its translation from one language to another.

Recently, studies have been done on this topic. For example, Jenkins et al. (2018) studied whether the order (either verbal - numeric or numeric - verbal) may influence the interpretation of a verbal probability expression, and Wintle et al. (2019) studied four different methods of presenting numeric probabilities along with a verbal probability expression. Since research on this topic is not finished, we would advise to keep an eye open for them. Hopefully an optimal mode of presentation of estimated probabilities will be found that minimizes the risks of miscommunication.

5.7 Supplementary material

5.7.1 Translations of probability phrases

Dutch phrase	English translation
onmogelijk	impossible
nooit	never
zeer onwaarschijnlijk	very unlikely, very improbable
bijna onmogelijk	almost impossible
bijna nooit	almost never
zelden	rarely, seldom
onwaarschijnlijk	unlikely, improbable
kleine kans	low chance
niet vaak	not often
soms	sometimes, once in a while
twijfelachtig	doubtful
kan gebeuren	liable to happen
kans	chance
onzeker	uncertain
misschien	maybe
mogelijk	possible
vermoedelijk	probable
vaak	often
te verwachten	expected
waarschijnlijk	likely, probably
meestal	usually
doorgaans	generally, usually
grote kans	high chance
heel vaak	very often
zeer waarschijnlijk	very likely, very probable
bijna altijd	almost always
bijna zeker	almost certain
zeker	certain
altijd	always

Table 5.2: *The 29 Dutch frequency and probability phrases used in the survey, with their English translations used in this paper. The phrases are presented in the same order as in Figure 5.2 in this paper.*

5.7.2 Context sentences

Dutch context sentences and their English translations (italic)
Het is waarschijnlijk dat alles in de koffer past.
It is likely that everything fits in the suitcase.
Het is waarschijnlijk dat het team een wedstrijd wint.
It is likely that the team wins a match.
Het is waarschijnlijk dat deze behandeling aanslaat.
It is likely that this treatment will work.
Het is waarschijnlijk dat een sollicitant geschikt is voor de baan.
It is likely that this applicant is suitable for the job.
Het is waarschijnlijk dat bedrijf A het product eerder lanceert dan bedrijf B.
It is likely that company A launches the product before company B does.
Het is waarschijnlijk dat zij gaan staken.
It is likely that they will go on strike.
Het is waarschijnlijk dat de uitslag goed is.
It is likely that the result is good.
Het is waarschijnlijk dat deze partij de grootste wordt bij de verkiezingen.
It is likely that this party will be the largest in the elections.
Het is waarschijnlijk dat dit plan slaagt.
It is likely that this plan succeeds.
Het is waarschijnlijk dat deze producten van goede kwaliteit zijn.
It is likely that these products are of good quality.
Het is waarschijnlijk dat dit hotel is volgeboekt.
It is likely that this hotel is fully booked.

Table 5.3: *The 12 Dutch context sentences used in the survey, with their English translations used in this paper.*

5.7.3 Differences between sub-populations

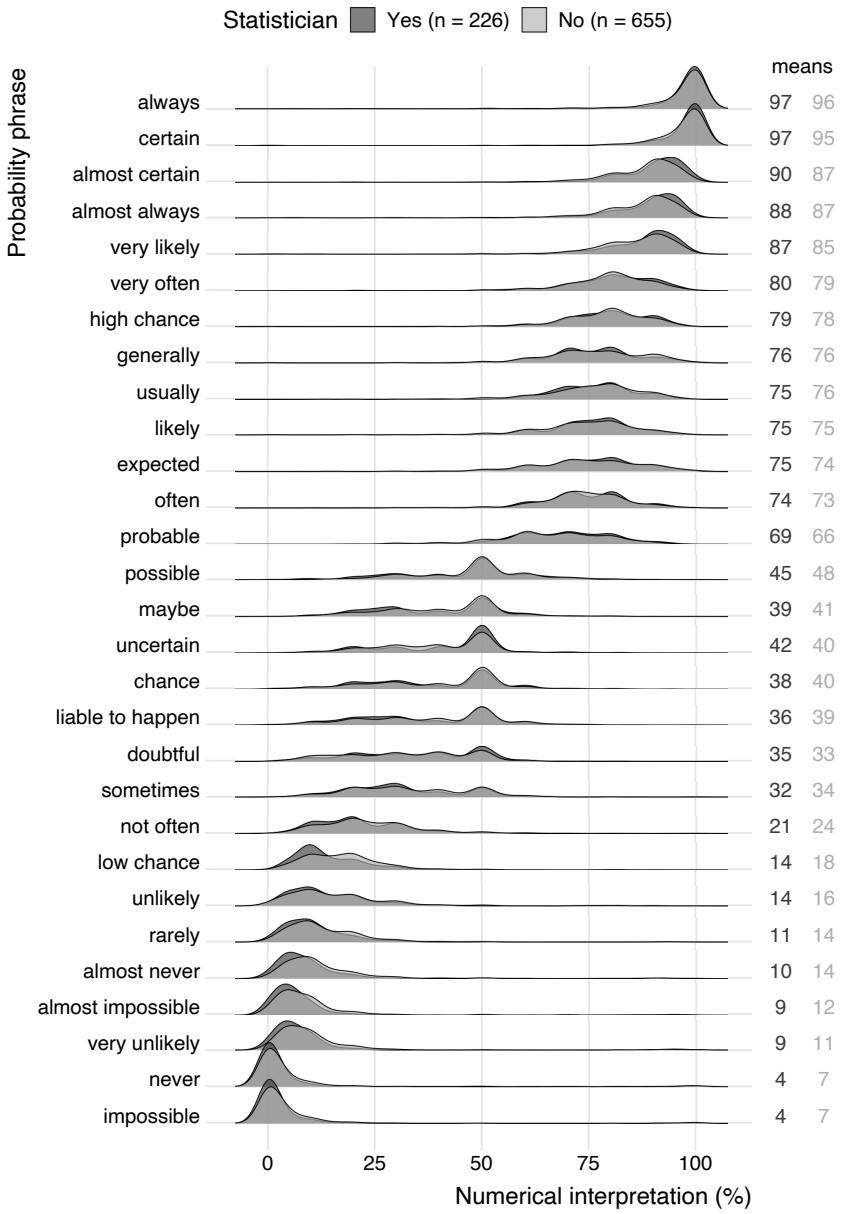

Figure 5.6: *Density plots and mean values of the numerical interpretations (in percentages) given by statisticians and non-statisticians for each phrase in the survey. Note that density plots are a smooth variant of histograms and may therefore be positive outside the data range of 0–100%.*

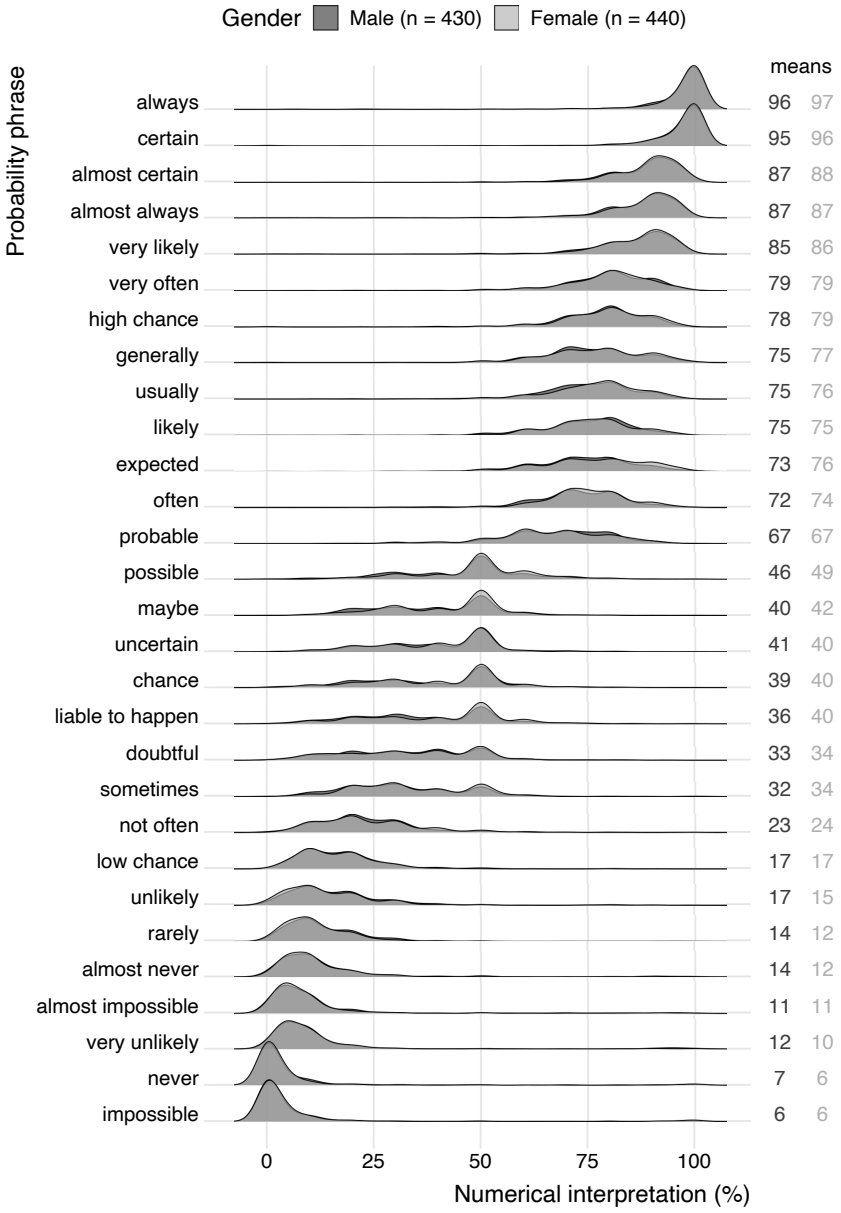

Figure 5.7: *Density plots and mean values of the numerical interpretations (in percentages) given by males and females for each phrase in the survey. Category level* Other/Prefer not to say *was omitted, because there were only 11 observations in this group. Note that density plots are a smooth variant of histograms and may therefore be positive outside the data range of 0–100%.*