

Top scores are possible, bottom scores are certain (and middle scores are not worth mentioning): A pragmatic view of verbal probabilities

Marie Juanchich*

Karl Halvor Teigen[†]

Amélie Gourdon[‡]

Abstract

In most previous studies of verbal probabilities, participants are asked to translate expressions such as *possible* and *not certain* into numeric probability values. This probabilistic translation approach can be contrasted with a novel which-outcome (WO) approach that focuses on the outcomes that people naturally associate with probability terms. The WO approach has revealed that, when given bell-shaped distributions of quantitative outcomes, people tend to associate *certainty* with minimum (unlikely) outcome magnitudes and *possibility* with (unlikely) maximal ones. The purpose of the present paper is to test the factors that foster these effects and the conditions in which they apply. Experiment 1 showed that the association of probability term and outcome was related to the association of scalar modifiers (i.e., it is certain that the battery will last *at least...*, it is possible that the battery will last *up to...*). Further, we tested whether this pattern was dependent on the frequency (e.g., increasing vs. decreasing distribution) or the nature of the outcomes presented (i.e., categorical vs. continuous). Results showed that despite being slightly affected by the shape of the distribution, participants continue to prefer to associate *possible* with maximum outcomes and *certain* with minimum outcomes. The final experiment provided a boundary condition to the effect, showing that it applies to verbal but not numerical probabilities.

Keywords: probability, language, verbal, judgment.

1 Introduction

Traditionally, verbal probabilities have been examined with a focus on their probabilistic meanings (e.g., *a chance* means on average a 30% probability). This venture is labelled the Probabilistic Translation approach. The present paper presents findings suggesting that this approach should be supplemented with a pragmatic approach that focuses on the outcomes people expect to be associated with a particular verbal expression (e.g., *a chance* is typically used to describe a top outcome). We have called this endeavour the Which Outcome (WO) approach.

When do people use verbal probability expressions such as “it is possible” or “it is certain”, and what do they mean? This question has traditionally been approached with a focus on the numeric probabilities these terms are conveying (e.g., Budescu & Wallsten, 1995; Clarke, Ruffin, Hill & Beamen, 1992; Juanchich, Sirota & Butler,

2013). In studies inspired by this probabilistic translation approach participants are typically given a “How Likely” task where they read a verbal probability qualifying a specified outcome and are asked to assess the numeric probability value the speaker has in mind. In other words, the verbal expression is “translated” into a number. Figure 1 provides an example of a How Likely Task. The translation approach has revealed that verbal probabilities are numerically vague, as participants suggest numbers that vary considerably (Budescu & Wallsten, 1995; Karelitz & Budescu, 2004). Despite this fuzziness, most people will agree that *possible* generally corresponds to a higher probability than *unlikely*, and that *likely*, *highly likely*, and *certain* correspond to still higher probabilities. This suggests a rough hierarchy of phrases that enables verbal probabilities to successfully compete with numeric probabilities as input to the decision process (Wallsten, Budescu & Zwick, 1993). In the present paper special attention will be paid to the expressions *possible*, *certain*, and *not certain*. Translation studies agree that *certain* indicates for most people probabilities close to 100%, whereas *not certain* corresponds on average to probabilities in the 40–50% range (Brun & Teigen, 1988; Reyna, 1981). This is also the probability range typically assigned to the term *possible* (Clarke et al., 1992; Theil, 2002).

This research was supported by an internal grant from Kingston Business School to the first author.

Copyright: © 2013. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

*Kingston Business School, Kingston University, Kingston Hill, KT2 7LB, UK. Email: M.Juanchich@kingston.ac.uk.

[†]University of Oslo, Norway.

[‡]University of Birmingham, UK.

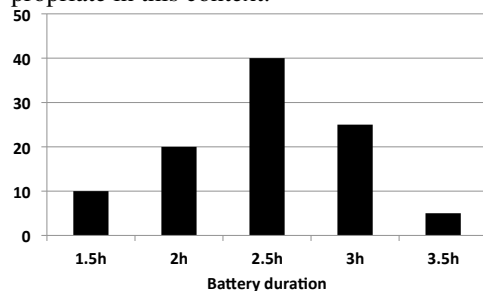
Figure 1: Examples of How Likely and Which Outcome tasks (Teigen, Juanchich & Filkuková, in press).

Example of How Likely Task

“It is possible that the battery will last 2 hours”
 What is the probability that the battery will last 2 hours?
 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Example of Which Outcome Task

A sample of computers of the brand “Comfor” were tested to check how long the batteries last before they need to be recharged. The figure below shows how many batteries lasted how many hours. Please complete the prediction below with the outcome that seems most appropriate in this context.



“It is possible that the battery will last hours”

Verbal phrases have also been studied from a more pragmatic perspective, which has focused on an alternative feature of the meaning of probability terms, namely their directionality (Gourdon & Beck, 2012; Honda & Yamagishi, 2006; Teigen, 1988; Teigen & Brun, 1999; for an attempt to reconcile the two approaches, see Juanchich, Sirota, Karelitz & Villejoubert, 2012). Expressions with a positive directionality (such as *a chance* and *possible*) draw attention toward the occurrence of an outcome and are typically associated with reasons explaining why the outcome should happen (e.g., “It is possible that Blacky will win the race because *he is in excellent shape*”). In contrast, expressions with a negative directionality (such as *not certain* and *unlikely*) focus on the flip side of the coin and are typically associated with reasons supporting the non-occurrence of the target outcome (e.g., “it is not certain that Blacky will win because *he was injured recently*”). The directionality of verbal probabilities allows complementary ways of framing a probabilistic message, which can “leak” information (to use an expression coined by Sher & McKenzie, 2006) about the speaker’s communicative intentions (e.g., encouraging someone to bet on Blacky or not to bet on Blacky). More generally, positive expressions are used whenever a probability is compared to lower chances, for instance, when an exaggerated probability statement has to be corrected upward. By contrast, negative expressions are used when a probability is compared to a higher one,

for instance, when a belief has to be adjusted downwards (Juanchich, Teigen & Villejoubert, 2010).

In the present study, we approach the usage and meaning of probability expressions from yet another pragmatic perspective. Whereas the two previous approaches describe how verbal probabilities qualify an outcome, in this approach we ask *which* outcomes do they qualify. This approach was developed in a context of outcomes that vary in terms of magnitude (Teigen & Filkuková, 2013; Teigen et al., in press; Teigen, Juanchich & Riege, 2013). For example, weather forecasters predict not only the possibility or likelihood of rain but also *how much* it will rain (e.g., “It is possible that it will rain 10mm”). Similarly, doctors do not only predict the chances of recovery of a patient, but a likely recovery time (e.g., “The patient will certainly recover in 10 weeks”). In studies using the *Which Outcome* (WO) approach, participants do not choose a numerical probability associated with the verbal expression, but indicate the *outcome* suggested by the verbal expression in an Outcome Completion Task. For example, in a Battery life vignette, participants received a unimodal, bell-shaped distribution of outcomes showing the duration of a sample of laptop batteries (as depicted in the lower panel of Figure 1), and were asked how long a battery *can* last, or will *possibly* last (“It is possible that a Comfor battery lasts for ... hours”). The WO approach revealed patterns of association between probability terms and outcome magnitude inconsistent with the traditional probabilistic translation approach.

Typically, *certain* was often associated with minimum outcomes, whereas *can*, *possible*, *a chance*, and *not certain* were all used to describe maximal outcomes (Teigen et al., in press). For example, in the battery vignette, where batteries lasted between 1 hour and 3.5 hours with a peak located at 2 hours, most of the participants claimed they were *certain* that the battery would last 1 hour and judged that it was *possible* that the battery would last 3.5 hours. Note that these two values were infrequent (5–10%) and should accordingly have been perceived as quite unlikely.

These findings were replicated in several studies, showing that people have strong associations between probability terms and outcome magnitudes that are not consistent with previous studies of probability terms’ numerical meanings. Indeed, participants selected rare outcomes, which had a frequency of 5–10%, to complete statements about possible and certain outcomes, despite the fact that possible and certain are believed to convey probabilities around 50% and 90% respectively. Thus, when the meanings of verbal probability phrases are explored with *which outcome* (in this case: *how much*) questions, the results can differ considerably from those obtained by the conventional *how likely* approach, leading to HL-WO discrepancies.

The present article investigates the conditions in which participants infer the intensity of an outcome from a probability term and aims to gain a better understanding of the dynamics that underpin the Which Outcome (WO) findings. A series of five studies aims to test the occurrence of the WO findings with different scalar modifiers, different outcome distributions (e.g., magnitude distributions of different shapes, categorical distributions) and to new formats (numerical probabilities). These studies contribute to the mapping of factors that are responsible for the WO effects.

1.1 At least or at most quantities?

The discrepancy between the probabilistic translation and Which Outcome approaches may be explained by the way participants construe the different tasks. Teigen et al. (2013) suggested that in the How Likely task (i.e., where participants provide a numerical probability to translate the meaning of a verbal probability) participants adopt a probabilistic reading of the task. In contrast, when they are asked to make a prediction, it appears that participants rely more heavily on pragmatic rules of communication of quantities that depart from a frequentistic uncertainty interpretation. When conveying quantities in natural language, it is common practice to report the minimum quantity expected. For example, a person saying that a crop will give 100 kg of potatoes may in fact make the prediction that the crop will give *at least* a 100 kg of potatoes. Jou, Shanteau and Harris (1996) and Mandel (2001) made the same suggestion that *certain* predictions are interpreted as communicating the minimum outcome to be expected in the context of framing studies (e.g., “If Program C is adopted, 400 people will die” would be understood as “If Program C is adopted, *at least* 400 people will die”). In fact it seems that the association between quantities and scalar modifiers (Nouwen & Geurts, 2007) may be more generalised and complex than expected: *certain* predictions could be associated with “at least” quantities, whereas predictions conveying different degrees of certainty would be associated with other scalar modifiers. So for instance, predictions of *possible* outcomes could be given “at most” readings, as suggested by a recent study by Teigen et al. (in press).

Teigen and Filkuková (2013; study 4) tested this possibility for the modal auxiliaries *will* and *can*, and showed that participants associated *will* (close to certain) with the modifier “at least”, whereas *can* (a modal with similar meaning to possible) was more often combined with the modifier “up to”. A Norwegian corpus search showed the same conclusion: *will* was often associated with the scalar modifier “at least”, whereas *can* was often associated with the modifier “up to”. Experiment 1 was designed to test whether the verbal probability terms *certain*

and *possible* are also linked to such scalar modifiers indicating extremity.

1.2 Extremeness or uniqueness?

With the novel *Which Outcome* approach we found that extreme outcomes were more often selected than could be expected from an analysis based on the frequencies of occurrence. For example participants selected a battery duration of 3.5 hours (frequency of 5%) to be *possible* whereas *possible* is usually considered to convey a 50% probability (Teigen et al., in press). Moreover, maximum extremes were selected more often than minimal ones, especially for outcomes with scalar properties where higher values entail the occurrence of lower ones (e.g., before lasting 3 hours, a battery lasted 1.5 hours). Participants thus preferred outcomes that were extreme, but also unique; which of these two factors drives the preference so far remains unclear.

The preference for extreme outcomes is supported by the cognitive psychology literature showing that extreme probabilities are judged as more informative than moderate ones (Keren & Teigen, 2001), and by the social psychology literature showing that information about extreme cases (e.g., extreme personality traits) has a greater impact than less extreme ones (Fiske, 1980). However, all outcome distributions studied so far (Teigen & Filkuková, 2013; Teigen et al., in press; Teigen et al., 2013) were bell-shaped and approximately symmetrical (see bottom of Figure 1), making extreme outcomes at the same time more infrequent than intermediate outcomes. Hence, we do not know whether people selected maximum outcomes because of their extremeness or their uniqueness—or conversely, whether a higher frequency would make extreme outcomes even more attractive, and perhaps increase the choice of bottom values. Experiments 2 and 3 were designed to replicate and extend the original findings of the WO effect for outcome distributions that are monotonically increasing or decreasing (Experiment 2) or U-shaped with a mode located either to the bottom or to the maximum of the distribution (Experiment 3).

1.3 Ruling out outcome extremeness with categorical outcomes

Which Outcome studies have so far been conducted with outcomes that can be ordered on a magnitude scale from low to high (e.g., short to long battery duration) where it makes sense to ask “how much?”. However, one can also ask participants to choose a *possible* or an *uncertain* outcome among a set of discrete categorical outcomes that differ only in frequency. For instance, if a car is randomly selected from a collection where 5% of the cars are white, 10% red, 20% grey, and 65% black—what is a

possible colour of the car? Which colour is *not certain*? There are in this case no top or bottom outcome values, as the colours can be arranged in any order, so the WO effect might seem to not apply. However, relative probability could be used as an alternative cue, suggesting that the most frequent alternative (black) should be mentioned most frequently as a *possible* colour, whereas the least frequent should be chosen for the *certain* colour. If instead uniqueness plays a role, more participants should select white than red or grey. Similarly, all outcomes in this example are, by definition, *not certain*. But is this primarily a characteristic that people will use to describe the most frequent or the least frequent alternative, or will they choose among those in between? In Experiment 4, the quantitative *how much*-question is replaced with a more general *which outcome*-question to ascertain whether, even under these circumstances, people are prompted to make choices that deviate from those suggested by the quantitative translation approach.

1.4 Is the Which Outcome (WO) effect limited to verbal probabilities?

The WO effect was demonstrated with modals (i.e., *will* and *can*; Teigen & Filkuková, 2013) and with positive and negative probability terms (e.g., *a chance* and *not certain*; Teigen et al., in press; Teigen et al., 2013). In their concluding comments, Teigen et al., (in press) speculated that the effect observed so far with linguistic materials might be extended to uncertainty communication in general. This hypothesis entails that the WO effect could be observed with numerical probabilities as well. In agreement with this possibility, results of Sirota and Juanchich (2012) show that even numerical probabilities can be interpreted as pragmatic devices to soften bad news, an interpretation that was previously believed to be limited to verbal probabilities (Bonneton & Villejoubert, 2006; Juanchich, Sirota & Butler, 2012). Yet other pieces of evidence support that verbal probabilities are more influenced by pragmatic considerations. For example, Windschitl and Wells (1996) posited that verbal probabilities are governed by associative processes (and thus more sensitive to contextual influences), whereas numerical probabilities are more rule based.

In Experiment 4 we asked the WO-questions based on numerical rather than verbal probabilities. Participants were asked which outcome is most appropriate to complete a statement such as “There is around a 10% [50%; 90%] chance that a battery of the brand Comfor will last for hours”. A literal interpretation of a 10% probability would require that the lowest outcome in the distribution is selected (which occurs in exactly 10% of the cases), whereas the WO effect suggests answers in the high end of the distribution, as these are most often cho-

sen in the verbal probability studies as examples of an “unlikely” outcome. As numerical probabilities encourage an analytic, mathematical mode of thinking, rather than a pragmatic approach, this could be a boundary test of the WO effect under circumstances not in its favour.

2 Experiment 1

Teigen and Filkuková (2013) showed that the completion of *can* and *will* sentences (e.g., a battery can last for ... hours) were related to specific scalar modifiers (i.e., *up to* and *at least*). The completion of sentences with a different outcome in the Which Outcome task may be based on a different perception of the questions. For example, people asked to complete the *certain* sentence may think that it is about providing the minimum outcome that could occur, the larger outcome entailing this one. On the other hand, *possible*, as *can*, could be interpreted as a question focusing on the maximal outcome that could occur.

2.1 Method

In a pre-test designed for the present studies, two classes of Norwegian undergraduate psychology students ($N = 43$) were asked to estimate the probabilities speakers would have in mind when saying “It is possible¹ [certain²; not certain³] that the laptop battery will last 3 hours” (along with six other verbal expressions). In this context, which comes close to the vignettes used below, *possible* was given a mean probability value of 52.6% ($SD = 17.9$), *certain* received a mean score of 92.1% ($SD = 14.7$), and *not certain* a mean score of 48.1% ($SD = 20.1$).

Participants. A total of 50 participants from Amazon Mechanical Turk completed the web questionnaire (Buhrmester, Kwang & Gosling, 2011; Paolacci, Chandler & Ipeirotis, 2010). The sample included 40% females and had a mean age of 33.18 years (18-65, $SD = 12.80$). Most participants were White Caucasian (88%) and had a higher education (86%).

Materials and Procedure. Participants read vignettes describing the frequency of outcome magnitudes, for example different battery durations (in hours) or different weight loss (in pounds) after following a diet. In each vignette the frequency distribution of the outcome magnitudes was presented as a bell shaped bar chart similar to the one depicted in Figure 1.

¹Norwegian: “mulig”

²Norwegian: “sikkert”

³Norwegian: “ikke sikkert”

Table 1: Choices (percentages) of modifiers to complete statements about three different products based on two different verbal probabilities (It is certain and It is possible), Experiment 1.

| | It is certain | | | | It is possible | | | |
|---------|---------------|-------|--------|---------|----------------|-------|--------|---------|
| | At least | Up to | Around | Exactly | At least | Up to | Around | Exactly |
| Battery | 78.7 | 6.4 | 14.9 | 0.0 | 4.3 | 72.3 | 19.1 | 4.3 |
| Diet | 74.5 | 10.6 | 14.9 | 0.0 | 6.3 | 72.9 | 18.8 | 2.1 |
| Jeans | 76.7 | 7.0 | 16.3 | 0.0 | 9.1 | 79.5 | 11.4 | 0.0 |

We used the vignettes Computer battery, Jeans shrinkage, and Diet developed by Teigen and Filkuková (2013). Only the vignette Diet was adapted by providing a unimodal and bell shaped bar chart instead of a table. Participants read the vignettes presented on different pages and in a single order. The vignettes featured a *certain* and a *possible* sentence to be completed, presented on the same page and in a randomised order.

The completion of the sentence was done by a drag and drop process, from a box containing four modifiers: at least, exactly, about, and up to, and the five numerical outcomes described in the distribution of outcomes. The completion instructions read as follows: *Please, complete each sentence below with the number that seems most natural in this context. You may also add one of the suggested words (at least, around, exactly or up to) if you feel this could improve the sentence.*

For example, in the Battery vignette, participants were asked to complete the following sentence: “it is certain that the battery will last _____ hours”, after reading a distribution of battery durations ranging from 1.5 to 3.5 hours by increments of 0.5 hour. The drag and drop completion task featured the following list organised in a column and always presented in this order: at least, exactly, around, up to, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours. The material is available in Appendix A.

2.2 Results

Outcome value selection. Between 76% and 80% of the participants completed the certain sentence with the minimum outcome value, whereas between 73% and 77% chose a maximum value to describe a possible outcome. As each scenario contained different ranges of numerical outcomes (e.g., in the battery scenario, the range was 1.5-3.5 hours, while in the diet scenario it was 10-18 pounds), we recoded in each case the lowest value as 1, the second lowest as 2, and so on. Therefore, an outcome of 3 represents the middle value, which was also the peak of the distribution and its most likely outcome value, whereas scores of 1 and 5 represent the bottom and top tail outcomes, respectively. Mean scores close to

5 indicate that participants chose outcomes towards the upper end of the distribution, while a mean score closer to 1 indicates choices towards the lower end of the distribution.

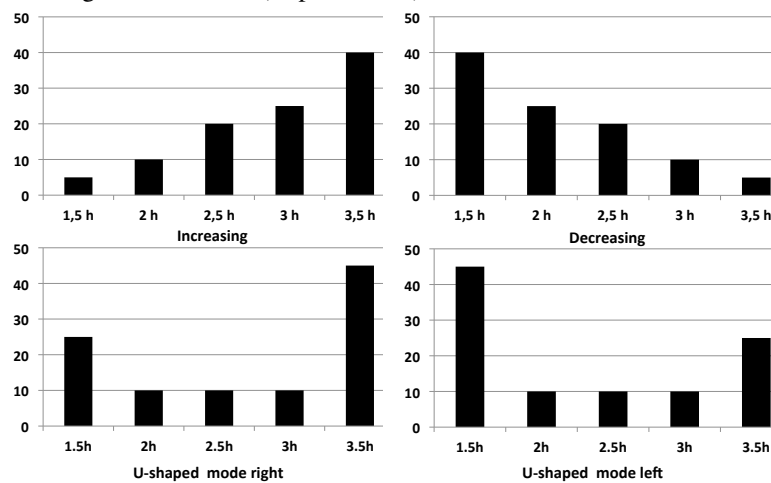
We ran a 3 x 2 within-subjects analysis of variance (ANOVA) on the recoded scores with scenario (battery, diet, and jeans) and probability term (*certain* and *possible*) as within-participants factors. The ANOVA indicated no main effect of the scenario, $F(2,94) = 1.26$, $p = .290$, but a main effect of the probability term, $F(1,47) = 146.85$, $p < .001$, $\eta^2_p = .78$. Participants completed the sentences with *certain* more often with values towards the lower end of the distribution ($M = 1.59$, $SD = 1.06$), and the sentences with *possible* more often with values towards the upper end of the distribution ($M = 4.40$, $SD = 0.97$). Finally, there was no interaction between the scenario and the verbal probability, $F(1,47) < 1$, $p = .586$ (lower bound adjustment).

Modifier selection. Participants provided modifiers for most of the vignettes (in 88% to 98% of the cases). Cases where participants provided a combination of modifiers were excluded (e.g., around up to; at least around up to).

As indicated in Table 1, participants completed the *certain* statement with a modifier stressing the lower bound of an interval (i.e., at least), showing that participants interpret the *certain* sentence completion as indicating the minimum outcome that can occur in the future. In contrast, participants completed *possible* with a majority of modifiers describing the upper bound of an interval (i.e., up to).

We ran analyses of frequencies to investigate the effect of verbal probability and vignette on scalar modifier selection. The analyses showed only an effect of verbal probability. When the verbal probability was certain, the sentence was more often completed with the modifier at least (70% of choices across scenarios), but when it was possible, it was more often completed with up to (69% of choices across scenarios). The effect of the verbal probability was significant in all of the three scenarios (respectively in the Battery, Jeans, and Diet vignettes: $\chi^2(4, N = 100) = 59.63$, $p < .001$, Cramer’s $V = .77$; $\chi^2(3, N = 100)$

Figure 2: Examples of distributions of outcomes that are monotonically increasing and decreasing (Experiment 2) or U-shaped with mode to the right or to the left (Experiment 3).



= 50.09, $p < .001$, Cramer’s $V = .71$; $\chi^2(4, N = 100) = 50.90, p < .001$, Cramer’s $V = .71$). There was no effect of the scenario when the verbal probability was certain ($\chi^2(6, N = 150) = 3.42, p = .755$, Cramer’s $V = .11$) or when it was possible ($\chi^2(8, N = 150) = 6.44, p = .598$, Cramer’s $V = .15$).

These results replicate the effect of verbal probability on outcome selection shown by Teigen et al. (in press), indicating that *certainty* was associated with unlikely small outcomes (frequency of 5%) whereas *possible* was associated with equally unlikely maximum outcomes. The present results further show that the pattern of outcome association is related to the perception of an implicit specific modifier. *Certainty* is associated with lower bound range modifier (at least) whereas *possible* was most often associated with higher bound range modifier (up to).

3 Experiment 2

When people are asked to select a *certain* outcome from a unimodal, bell-shaped distribution of outcomes (see Figure 1) they often choose the lowest one, or alternatively, the most frequent middle score. As shown in Experiment 1, when they are asked to pick a *possible* outcome they regularly choose the highest value (Teigen et al., in press), despite the fact that outcomes in both tails of the distribution are infrequent and hence quite unlikely. The present experiment was designed to investigate whether this phenomenon could be replicated with monotonic distributions. In monotonically increasing distributions the lowest score is least frequent, whereas the highest score is most frequent. In contrast, in a monotonically decreasing distribution the lowest value is the most frequent, whereas the highest value is least frequent as illustrated by the dis-

tributions in the top panel of Figure 2. If extremeness is a decisive factor, people will continue to describe low values as *certain* and high values as *possible*, regardless of the shape of the distribution. If, on the other hand, outcome frequencies are important then the response pattern might change. So for instance, if the phrase “it is possible” suggests a low probability, participants should be more inclined to select an outcome in the lower than in the upper end of the distribution like the one depicted in the left panel of Figure 2. Conversely, if “certain” is influenced by outcome frequency, they might switch their outcome preferences from low and middle to high values.

3.1 Method

Participants. Students following an introductory class at the University of Tromsø, Norway served as participants; $N = 59$ (89.8% female, 2 did not report gender; median age 20 years). They were randomly assigned to one of two conditions, one with decreasing and the other with increasing distributions.

Design, Materials and Procedure. All participants read four vignettes previously used by Teigen and Filkuková (2013) and Teigen et al. (in press). Two of these high values were associated with positive outcomes (duration of computer batteries and weight reduction with a diet product), and two with negative outcomes (jeans that shrink when washed and letters that are delayed in mail). The Battery, Diet, Jeans, and Mail vignettes were presented in this order, each accompanied with frequency information about the occurrence of different values in a sample of outcomes, either displayed as graphical distributions (as in the top panel of Figure 2) or in a tabular format. All participants completed two statements for

Table 2: What will certainly and what will possibly happen? Choices (percentages) of low, intermediate, and high numbers in statements about four different products based on monotonically increasing and decreasing distributions, Experiment 2.

| Vignettes | Ascending distributions | | | Descending distributions | | |
|----------------|-------------------------|---------|------|--------------------------|---------|------|
| | Low | Interm. | High | Low | Interm. | High |
| It is certain | | | | | | |
| Battery | 86.2 | 6.9 | 6.9 | 96.4 | 3.6 | . |
| Diet | 89.7 | 6.9 | 3.4 | 100 | . | . |
| Jeans | 93.1 | . | 6.9 | 96.4 | . | 3.6 |
| Mail | 48.1 | 7.4 | 44.4 | 67.9 | 7.1 | 25.0 |
| Total | 79.3 | 5.3 | 15.4 | 90.2 | 2.7 | 7.1 |
| It is possible | | | | | | |
| Battery | 10.6 | 3.3 | 86.7 | . | 7.1 | 92.9 |
| Diet | 6.7 | 3.3 | 90.0 | . | 3.4 | 96.6 |
| Jeans | 6.7 | . | 93.3 | . | 3.4 | 96.6 |
| Mail | 46.7 | 13.3 | 40.0 | 20.7 | 10.3 | 69.0 |
| Total | 17.7 | 5.0 | 77.5 | 5.2 | 6.1 | 88.8 |

each vignette, one containing the term *certain* and another with the term *possible*, (e.g., “it is possible [certain] that the battery will last for hours”). Half of the participants received the statements with *possible* before the *certain* statements, and the other half completed the statements in the reverse order.

Half of the participants read distributions of outcomes that were monotonically increasing, with the lowest value as the least frequent one and the highest value as the mode, and the other half read decreasing distributions that dovetailed the increasing version. The vignettes and their associated distributions are presented in Appendix B. This study therefore featured a 4 x 2 x 2 mixed design, with the four vignettes and the two probability terms within-subjects and the shape of the distribution between-subjects.

3.2 Results

Responses from all vignettes were coded in three categories as low (minimum outcome), high (maximum outcome) and intermediate (all scores in between). Results displayed in Table 2 show that most participants preferred to describe a *certain* outcome with the bottom value of the distribution and a *possible* outcome with the maximum value of the distribution, both for ascending and descending distributions. For example, about 90% of the participants judged that it was *certain* that the battery would last 1.5 hours, whereas a similar percentage judged that it

was *possible* that it would last 3.5 hours, given a distribution in which a sample of batteries lasted from 1.5 to 3.5 hours.

A mixed-design variance analysis was conducted, to test the effect of the verbal probability, distribution and vignette on outcome selection. The analysis showed that verbal probability had an effect on outcome selection ($F(1, 52) = 254.44, p < .001, \eta^2_p = .83$). The vignette did not produce a main effect ($F(3, 156) < 1$), nor an interaction with the distribution shape ($F(3, 156) < 1$). Vignette was interacting with verbal probability ($F(3, 156) = 22.90, p < .001, \eta^2_p = .31$). This interaction is due to mail vignette where participants did not select the same outcomes as in the other vignettes. The deviant response pattern of this vignette has been previously observed (Teigen et al., in press) and was presumed to be caused by the fact that the outcome dimension in this vignette is bi-directional. Indeed the time it takes for a letter to go from Norway to the United-States can be construed in terms of quickness or in terms of delay, therefore creating a bimodal preference for either the fastest delivery (i.e., minimum outcome value) or the longest delay (i.e., maximal outcome).

A closer inspection of the results displayed in Table 2 reveals that the association between certain and minimum outcomes and between possible and maximum outcomes are slightly stronger for descending than for ascending distributions, leading to a statistically significant interaction between distribution shape and probability term

($F(1, 52) = 6.10, p = .017, \eta^2_p = .10$). However, the general conclusion to be drawn from these patterns of results is that the tendency to describe low values as *certain* and high values as *possible* is not uniquely dependent upon one particular distribution of outcomes.

These results support the view that preference for maximum outcomes to describe possibility is not determined by their frequency of occurrence. If *possible* was associated with rare outcomes, like in the previous studies, we would have expected a shift of responses toward the low end of the distribution in the ascending series. However, participants continued to select maximal values as the possible ones, even when they were also most frequent. This preference was weaker in the mail context, which has an ambiguous underlying dimension, where both lowest values (i.e., how fast the mail takes) and highest values (i.e., how long it takes) can be perceived as top achievements (for ambiguous dimensions, see Teigen et al., in press, Study 3). The valence of the outcome described did not appear to modify participants' preferences for outcome magnitude.

Participants in the present experiment associated certainty almost exclusively with the lowest value, regardless of its frequency. Even in the ascending distribution the maximum value was rarely selected, although it was the most frequent outcome of the distribution. This pattern is more conspicuous in the present results than for the bell-shaped distributions used previously and contrasts sharply with predictions based on the probabilistic translation approach, where the most probable outcome would also be expected to be the most certain. However, the present results are compatible with a directional, scalar interpretation of numerical magnitudes where low outcomes are entailed by higher ones. For outcomes ordered on such a scale it makes sense to say that a battery of this brand will certainly last for *at least* 1.5 hours (and probably longer) because before lasting 3 hours, a battery has indeed lasted for 1.5 hours.

4 Experiment 3

Experiment 2 demonstrated that participants associate *certain* with the lowest outcome and *possible* with the highest outcome in a monotonically increasing or decreasing distribution, replicating earlier findings based on unimodal symmetrical distributions (Teigen et al., in press). In Experiment 2, extreme outcomes were either the least or the most frequent, thus implying frequency extremeness. The rarity, or conversely, the high frequency of occurrence of the outcome may have contributed to steer participants in selecting these outcomes, as both low and high extreme likelihoods are more informative than middle ones (Keren & Teigen, 2001). Ex-

periment 3 was designed to replicate these findings using bimodal, U-shaped distributions as shown in the bottom panel of Figure 2, where the middle outcomes are the least frequent.

Moreover, in Experiment 2, statements about what is certain and what is possible were completed by the same participants. This might have contributed to polarizing participants' assessment of these two terms, in line with findings suggesting that people tend to exaggerate the difference between concepts they are asked about in a joint presentation format, compared with their assessment of the same concepts presented on separate occasions (Schwarz, 1999). To control for this effect, in the present experiment assessments of certain and possible outcomes were performed separately in a between-subjects design. Finally, to avoid a potential influence of outcome valence, one presumably neutral "nonsense" or "utility free" vignette was included in the set, replacing the ambiguous Mail vignette.

4.1 Method

Participants. Psychology students from the University of Birmingham (UK) took part in this experiment in exchange for course credits, $N = 102$ (81.1% females, median age: 19 years). Participants were randomly allocated to four different conditions.

Design, Materials and Procedure. In a 2 (Probability term) \times 2 (Mode location) \times 4 (Vignette) mixed design, participants read the Battery, Jeans, and Diet vignettes previously used (see Vignettes 1-3 in Appendix B) and the new Shmulp (utility free) vignette (described as Vignette 5 in Appendix B). The utility free vignette was composed with fantasy outcomes in a non-existing context. This vignette aimed to test whether the reference for minimum or maximal outcomes was derived from information about the utility of the outcomes. If this was the case, the preference for minimum or maximal outcome should not be observed in this vignette. The probability term and the mode location of the distribution were manipulated in between-subjects.

In each vignette the distribution of outcome values was presented by means of a bar chart featuring a U-shape with two peaks, the highest being located either to the left or to the right (as in the bottom panel of Figure 2). In all vignettes participants were asked to complete one statement describing either a "possible" or "certain" value (e.g., "it is possible that the battery will last ... hours").

4.2 Results

For each vignette, answers more than 3 standard deviations from the mean were discarded. This left 95 partici-

Table 3: What will certainly and what will possibly happen? Choices (percentages) of low, intermediate, and high numbers in statements about four different products, for U distributions with the mode located either on the left or on the right (N=95), Experiment 3.

| Vignettes | Mode on the left | | | Mode on the right | | |
|----------------|------------------|---------|------|-------------------|---------|------|
| | Low | Interm. | High | Low | Interm. | High |
| It is certain | | | | | | |
| Battery | 90.5 | 9.5 | -- | 69.6 | 4.3 | 26.1 |
| Diet | 61.9 | 14.3 | 23.8 | 78.3 | 8.7 | 13.0 |
| Jeans | 81.0 | 19.0 | -- | 56.5 | 17.4 | 26.1 |
| Shmulps | 76.2 | 14.3 | 9.5 | 60.9 | 8.7 | 30.4 |
| Total | 77.4 | 14.3 | 8.3 | 66.3 | 9.8 | 23.9 |
| It is possible | | | | | | |
| Battery | 40.0 | 4.0 | 56.0 | 11.5 | -- | 88.5 |
| Diet | 44.0 | 4.0 | 52.0 | 7.7 | -- | 92.3 |
| Jeans | 32.0 | 16.0 | 52.0 | 3.8 | 3.8 | 92.3 |
| Shmulps | 40.0 | 4.0 | 56.0 | 7.7 | 3.8 | 88.5 |
| Total | 39.0 | 7.0 | 54.0 | 7.7 | 1.9 | 90.4 |

pants with answers in each scenario, to be taken into account in the analysis. Remaining values were recoded as low, intermediate, and high, as in previous experiments. Answers outside the distribution range, if not outliers, were recoded as either low or high values (3 in the battery vignette, 4 in the diet vignette, 3 in the jeans vignette, and 11 in the Shmulp vignette).

Results displayed in Table 3 show that most of the participants preferred describing *certain* outcomes with the bottom values of the distributions. However this trend was more pronounced when the mode of the distribution was low (on the left) than when it was high (on the right). For example, 92% of the participants judged that it was certain that the battery would last 1.5 hours, given a distribution in which the mode was 1.5 hours, whereas 67% judged that it was certain that it would last 1.5 hour, given a distribution in which the mode was 3.5 hours.

In contrast, most of the participants described *possible* outcomes with the maximum value of the distribution. Again, this trend was more pronounced when the mode of the distribution was high (on the right) than when it was low (on the left). For example, 89% of participants judged that it was possible that the battery would last 3.5 hours, given a distribution in which the mode was 3.5 hours, whereas 54% judged that it was possible that it would last 3.5 hours, given a distribution in which the mode was 1.5 hours.

For the purpose of the analysis of variance, low answers were given a value of 1, intermediate answers received a value of 2 and high answers were given a value

of 3. The main and combined effects of the probability terms (i.e., possible vs. certain) and of the mode location (i.e., left vs. right) on outcome preferences were tested by a 4 (vignette) x 2 (verbal probability) x 2 (distribution) mixed design variance analysis.

Overall, the vignettes had no effect on the magnitude of the answer, ($F(3, 273) < 1, p = .358$; lower bound adjustment). The analysis revealed a main effect of the location of the modal value on the outcome preference ($F(1, 91) = 12.86, p = .001, \eta^2_p = .12$) and a main effect of the probability term on the outcome magnitude ($F(1, 91) = 63.20, p < .001, \eta^2_p = .41$). The analysis did not reveal an interaction between the probability term and the location of the modal value or between the probability term and the vignette (respectively, $F(1, 91) = 2.43, p = .122$ and $F(3, 273) < 1, p = .564$). There was no significant interaction between the location of the modal value and the vignette ($F(3, 273) = 2.32, p = .076, \eta^2_p = .03$). There was an interaction between the probability term, the location of the modal value and the vignette ($F(3, 273) = 4.36, p = .005, \eta^2_p = .05$). This interaction may illustrate that in the *certain* Diet vignette the preference for the minimal outcome was reinforced when the mode was on the right whereas in all the other vignettes this preference was reinforced when the mode was on the left.

These results replicate the WO effect, showing that participants associate certainty with the minimum outcome and possibility with the maximum one. It is worth noting that this finding holds even when *possible* was not

contrasted with *certain* in a within-subject design such as in Study 2. Moreover, the effect seems not to be utility dependent as it also was found in the Shmulp vignette, which describes a non-existent entity with nonsense outcome values (Shmulps can have different numbers of Glomps). In the Shmulp context it is hence difficult to claim that extreme outcomes are preferred because they are considered more (or less) useful than middle ones. Finally, results based on the U-shaped distributions suggest that the preference for extreme outcome values is not caused by the shape of the distribution, although participants preferred the extremes that were also most likely. This applies both to *certain* and *possible*. Bottom outcomes are more certain when they are frequent, and similarly, maximum outcomes are more possible when they are also the most frequent ones.

5 Experiment 4

Experiments 1, 2 and 3 confirmed previous findings (Teigen et al., in press) that people typically complete *possible* statements with extreme outcome values, and disregard outcomes in the middle range. Moreover, results indicate that extreme outcomes were selected to qualify *possible* or *certain* irrespective of whether those outcomes were the least, the most or only moderately likely. Building on these findings, we can assume that the extremeness of the outcomes in itself may be partly responsible for the preference for maximal and minimal outcome in the outcome completion task. To test the role of outcome quantitative extremeness the next experiment features outcomes that cannot be ordered on a quantitative dimension.

In Experiment 4 participants were presented with frequency distributions of multiple categorical outcomes (e.g., the distribution of red, black, white, and grey cars), which cannot be ordered according to magnitude. For example, a black car cannot be ranked as surpassing cars of a different colour, and does not imply the previous occurrence of a red one in the same way that a battery lasting 3.5 hours implied that the battery has also lasted for 1.5 hour. In categorical outcome distributions, it makes less sense to speak of maximum or minimum scores, and the concept of outcome extremeness is not applicable, unless redefined. This could be obtained by ordering categories from least to most frequent, or the other way around, from common to rare.

Three verbal probabilities will be examined here: It is *possible*, it is *entirely possible* and it is *not certain*. If *possible* is still associated with “maximum scores”, it would now be expected to be used to characterize either the most frequent or the least frequent category, as opposed to the intermediate ones. The emphatic expression

entirely might be expected to magnify the tendency to associate *possible* with a highly probable outcome, but from a conversational perspective, *entirely* could rather call attention to the occurrence of a relatively improbable outcome. Finally, the phrase *not certain* is, by definition, applicable to all outcomes that have a probability of occurrence below 1.0. However, for magnitude distributions, Teigen et al. (in press) found that *not certain* statements were primarily associated with maximum or middle outcomes, rather than low ones. In other words, *not certain* differs from *possible* by being applicable to intermediate values, but was similar in the sense that high extremes were preferred to minimum scores. For categorical outcomes, as in the present study, it remains an open question whether *not certain* will be associated with the most frequent category, the least frequent category (which both can be regarded as “top” scores, albeit on different dimensions) or categories in the intermediate range.

The verbal probability *certain* was not used here as the *at least* interpretation does not apply to categorical variables, and none of the outcome categories were comprehensive enough to be regarded as *certain*. Participants would accordingly feel that the task of picking a *certain* outcome to be a meaningless one. In contrast, all categories with a non-zero probability of occurrence would in principle be *possible* as well as *not certain*.

5.1 Method

Participants. Students from the University of Birmingham (UK) took part in the experiment on the internet in exchange for course credits, $N = 98$ (89.8% females, median age: 19 years). Participants were randomly allocated to three conditions with 32-34 participants in each.

Design, Materials and Procedure. The participants read four vignettes presented in randomized order, and for each vignette they completed a statement starting with either “it is possible that ...”, “it is entirely possible that ...” or “it is not completely certain that ...”. Participants completed the statement with the categorical outcome they judged most appropriate.

Each vignette described 3 to 4 categorical outcomes with occurrence frequencies in percentages. For example in the car vignette participants read that Jason has won a car and that he is wondering about the car colour. Jason is then informed that among the cars to be won 5% are white, 10 % red, 20% grey, and 65% black. The order of the categorical outcomes was counterbalanced so half of the participants read the frequencies presented in an increasing order (e.g., 5% to 65%) and the other half in a decreasing order (e.g., 65% to 5%). The order of presentation of the outcomes was based on their frequency, either decreasing or increasing to prevent participants to

Table 4: Percentages of participants choosing the least frequent, moderately frequent or most frequent category to describe a possible, an entirely possible or a not certain outcome, Experiment 4.

| Vignettes | Possible ($n = 32$) | | | Entirely possible ($n = 34$) | | | Not certain ($n = 32$) | | |
|-------------|-----------------------|------|------|--------------------------------|------|------|--------------------------|------|------|
| | Least | Mod. | Most | Least | Mod. | Most | Least | Mod. | Most |
| Car | 12.5 | 3.1 | 84.4 | 32.4 | 2.9 | 64.7 | 40.6 | 9.4 | 50.0 |
| Supervision | 3.1 | 3.1 | 93.8 | 20.6 | 0.0 | 79.4 | 34.4 | 18.8 | 46.9 |
| Transport | 9.4 | 12.5 | 78.1 | 14.7 | 2.9 | 82.4 | 34.4 | 0.0 | 65.6 |
| Shmulp | 15.6 | 3.1 | 81.3 | 32.4 | 5.9 | 61.8 | 37.5 | 6.3 | 56.3 |
| Total | 10.2 | 5.5 | 84.4 | 25.0 | 2.9 | 72.1 | 36.7 | 8.6 | 54.7 |

infer an outcome preference based on their order of presentation. The order of presentation did not affect the participants' preferences and will not be discussed further.

In addition to the car vignette described above, two vignettes illustrated real life situations (Supervision and Transport) and the final vignette used a nonsense context (Shmulp). In the Supervision vignette a teacher is wondering whether she would tutor a student attending program A (10%), B (20%), or C (70%). In the Transport vignette a man is guessing whether his wife would go to work the next day by car, bus, or bike. Finally, in the Shmulp vignette, participants read that shmulp are of three kinds: gclering (10%), laurding (20%) or glimpsing (70%). The vignette was introduced as follows: *Please read the following vignette; don't worry if you do not understand the meaning of some words. Try to complete the statement with the expression that sounds most appropriate given the context.* The complete set of vignettes is presented in Appendix C.

5.2 Results and Discussion

Choice distributions are presented in Table 4. The car vignette (Car) involved a choice among four options. For this vignette, the selections of the two "moderately frequent" categories were pooled together. The three other vignettes involved a choice between three categories (e.g., students from programme A, B, or C in the Supervision vignette). In the table, the categories are ordered according to size from least to most frequent.

As depicted in Table 4, most participants completed the statements with the most frequent outcome across all the conditions and vignettes (from 46.9% to 93.8%). Nevertheless, some participants chose the least frequent outcome to describe a *possible* outcome (10.15%). This preference was magnified by the more colloquial *entirely possible* statement, for which one participant in four chose the least frequent outcome. It was speculated that this more emphatic expression would fit a conversational

context where it is important to announce that this outcome, despite its rareness, should not be overlooked. Finally, more than one in three participants chose the least likely outcome to describe a *not certain* outcome.

Overall, only around 5% of the participants chose the intermediate outcome, even if the frequency of this outcome (20-30%) should be within the range of probabilities corresponding to *not certain* and *possible*, according to probabilistic translation studies (Brun & Teigen, 1988).

The main and interaction effects of the probability terms (i.e., *possible*, *entirely possible*, *not certain*) and of the vignettes on outcome choices were tested with a mixed-design variance analysis. The order of presentation of the categorical outcome (i.e., increasing vs. decreasing order) was also included as an independent variable. The tests of between-subjects effects revealed a main effect of the probability term on the outcome preference, $F(2, 92) = 6.07, p = .003, \eta^2_p = .12$. The order of presentation of the categorical outcome did not affect participants' preference, ($F(1, 92) < 1$), nor did it interact with the probability term ($F(2, 92) < 1$). The probability term and the vignette did not interact significantly ($F(6, 276) = 1.88, p = .085, \eta^2_p = .04$). There was also a main effect of the vignette on the outcome choice, $F(3, 276) = 17.49, p < .001, \eta^2_p = .16$. This was related to a stronger preference for the most frequent outcome in the supervision and the transport vignettes.

Participants appear to have avoided moderately frequent outcomes and preferred the least and most frequent ones. Most of the participants chose to describe *possible*, *entirely possible* and *not certain* statements with the most likely categorical outcome. In the present vignettes, these outcomes were assigned frequencies of 60-70%, which is somewhat higher than the probabilities (around 40-50%) suggested to correspond to *possible* by participants in translation studies, but not incompatible with these values. However, intermediate outcomes were almost never mentioned as possible, despite being assigned frequencies of 20-30%, which are about equally close to typical

Table 5: Choices (percentages) of low, intermediate and high outcome values for predictions featuring a 10%, 50% or 90% probability, Experiment 5.

| Vignettes | Around 10% ($n = 30$) | | | Around 50% ($n = 32$) | | | Around 90% ($n = 31$) | | |
|-----------|-------------------------|--------|------|-------------------------|--------|------|-------------------------|--------|------|
| | Low | Inter. | High | Low | Inter. | High | Low | Inter. | High |
| Battery | 66.7 | 20.0 | 13.3 | – | 96.9 | 3.1 | 9.7 | 87.1 | 3.2 |
| Diet | 21.4 | 32.1 | 46.4 | – | 93.3 | 6.7 | 35.5 | 54.8 | 9.7 |
| Jeans | 53.3 | 13.3 | 33.3 | 6.3 | 87.5 | 6.3 | 13.3 | 86.7 | – |
| Mail | 13.3 | 6.7 | 80.0 | 9.4 | 90.6 | – | 9.7 | 63.4 | 26.9 |
| Total | 38.7 | 18.0 | 43.2 | 3.9 | 92.1 | 4.0 | 17.1 | 73.0 | 9.9 |

numeric translations of possible. In contrast, a larger proportion of individuals felt drawn toward the least likely outcome, which had an occurrence frequency of only 5–10%. For *not certain* the least likely outcome was chosen even more frequently. This could be explained by different membership functions for negative than for positive probability expressions (Budescu, Karelitz & Wallsten, 2003), but even the membership approach would find it difficult to explain the absence of intermediate outcomes, as membership curves are consistently drawn as single-peaked (cf. Dhimi & Wallsten, 2005, Figure 3).

6 Experiment 5

This study aimed to test whether the results of the new approach also applied to numerical probabilities or whether numerical probabilities were associated with outcomes according to strict quantitative rules. We studied which of the outcome values taken from a bell-shaped distribution participants associated with three numerical probability magnitudes: around 10%, around 50%, and around 90%. The term “around” was used to relax impossible probabilistic constraints (e.g., associating a 50% probability to a 40% probable outcome).

Based on the new approach findings, we expect low and moderate probabilities (10% and 50%) to be associated with maximum outcome values (in agreement with low and moderate verbal phrases, such as *unlikely*, *a chance*, and *possible*), whereas 90% could be expected to suggest minimum values (like *certain*). On the other hand, the probabilistic translation approach predicts that low probabilities will be associated with extreme outcome values (e.g., low or high), whereas 50% and 90% probabilities will be associated with intermediate values that are also the most frequent.

6.1 Method

Participants. Altogether 94 participants were recruited through Amazon Mechanical Turk (53.6% female; median age: 30 years, age range: 18–63 years). Most of the participants were in work (69.2%) and had higher education experience (83.6%). Three participants did not report their socio-demographic characteristics and two participants reported an impossible age (2 and 4 years old), which was considered a typo (or a joke). Participants were randomly assigned to three different numerical probability conditions (around 10%, around 50%, and around 90%). Participants with extreme scores (more than 3 standard deviations away from the mean) were excluded, resulting in $n = 32$, $n = 30$, and $n = 31$ in Conditions 10%, 50%, and 90%, respectively.

Questionnaires. All participants completed a brief web-questionnaire composed of the four vignettes (Battery, Diet, Jeans, and Mail) used in Experiment 2, but with the original bell shaped outcome distributions shown in Figure 1. In each vignette participants read and completed a numerical uncertainty statement (e.g., “there is around a 10% chance that a battery of the brand Comfort will last ... hours”). The vignettes were presented in a randomized order on separate web pages. The uncertain outcome was described with a probability of “around 10%”, “around 50%” and “around 90%” in three different versions of the questionnaire.

6.2 Results

Outcome values chosen by participants were recoded as low, intermediate and high values as in Experiments 1 and 2. The distributions of choices for each probability condition are shown in Table 5.

The outcome values associated with a 10% chance were mostly taken from the low and high tails of the distribution. For example, the 10% chance of battery duration was most often associated with the minimum battery

duration (67%), in agreement with the distribution percentages (in this vignette, the minimum score was obtained by 10%, whereas the maximum score was only achieved by 5% of the batteries). For the other vignettes, both maximum scores and bottom scores were obtained by 1/10 of the samples. In the Jeans vignette participants preferred the lowest value, whereas they associated a 10% chance with the highest values to describe weight loss following a diet, or the time a letter from Norway takes to arrive in the US. A large majority of the participants judged that a 50% chance and a 90% chance best characterised an intermediate value (92.9% and 80.0% respectively). For example a computer battery had a 90% chance to last 2.5 hours given that such batteries were previously found to last between 1.5 and 3.5 hours. Yet it is interesting to note that altogether, 27% of participants chose one of the (rare) extreme outcomes to describe a 90% chance. Of these, 63.3% chose the minimum outcome, in line with the How Much effect observed for *certain* in the previous studies. Numerical probability magnitude had an effect on the outcome chosen in all the four vignettes (respectively in the Battery, Diet, Jeans, and Mail vignettes: $\chi^2(4, N = 93) = 50.89, p < .001, \text{Cramer's } V = .52$; $\chi^2(4, N = 89) = 33.31, p < .001, \text{Cramer's } V = .43$; $\chi^2(4, N = 92) = 48.40, p < .001, \text{Cramer's } V = .51$; $\chi^2(4, N = 93) = 69.80, p < .001, \text{Cramer's } V = .61$).

Overall, participants' choices of outcome values were quite consistent with a literal interpretation of numerical probabilities and appeared not to be affected by the WO effect. These results suggest that numeric and verbal probabilities are pragmatically different in the sense that verbal expressions corresponding to a moderate degree of certainty (*possible, not certain*) are typically associated with the maximum extreme values, whereas a corresponding numerical probability (50%) is associated with outcomes in the middle range. A high verbal probability (*certain*) can be given an *at least* interpretation, and hence be associated with low outcome values. The present results suggest that this makes less sense for a high probability expressed in numbers (90%).

7 General discussion

In studies conducted within the Which Outcome (WO) approach, participants are typically provided with a frequency distribution of outcomes (e.g., computer batteries which last between 1.5 and 3.5 hours) and asked to complete a verbal statement with the outcome value that sounds most appropriate ("It is possible that a computer battery will last ___ hours"). This question format revealed several WO-effects featuring (1) a preference for selecting outcomes in the tails of an outcome distribution, (2) a preference for low end values to describe *cer-*

tain outcomes and outcomes that *will* happen, and (3) a preference of high end values for *unlikely* and *uncertain* outcomes, *possible* outcomes, outcomes that have a *chance* of happening, or outcomes that *can* occur (Teigen & Filkuková, 2013; Teigen et al, in press; Teigen et al., 2013). The objective of the present paper was to study the robustness of these effects and to test their limits with various distribution shapes, outcome types, and with numeric rather than verbal expressions of probability, thereby improving our understanding of the pragmatics of communication about uncertain events.

Results of five studies informed and specified the WO approach on four points.

First, the preference for a specific outcome goes together with the preference for a scalar modifier. *Possible* predictions are likely to be associated with the modifier "up to" and with a maximal outcome, whereas *certain* predictions are most often associated with the modifier "at least" and with a minimum outcome (Experiment 1). Secondly, a preference for extreme outcomes was not limited to symmetrical distributions. This finding indicates that extremes were not preferred because of their uniqueness; on the contrary, results from Experiment 2 and 3 show that extremes that were frequent were chosen more often than less frequent extremes. Thirdly, disregard for intermediate values could also be observed in a set of categorical outcomes (Experiment 4), where the most and the least frequent category were chosen more often than categories in the middle. Finally, the WO effects characteristic of verbal probabilities cannot be extended to uncertainty communication in general as a study conducted with numerical probabilities (Experiment 5) found that they are given literal (mathematical) rather than pragmatic interpretations.

7.1 Predictions and modifiers

Findings of Experiment 1 provided evidence that the term *possible* was often associated with predictions of maximum outcomes and was therefore often associated with a scalar modifier indicating an upper interval bound (i.e., *up to*). *Certain* was often associated with a minimum outcome that had a low frequency of occurrence and was associated with a scalar modifier indicating a lower interval bound (i.e., *at least*). Interestingly, very few participants chose to use the term *exactly* (0-6%), although they were provided the frequentistic information that is deemed to be sufficient to make accurate predictions. These results mirror well the findings of Teigen and Filkuková (2013) with the terms *can* and *will*, as findings with *possible* and *certain* were very similar to findings with *can* and *will*, respectively. Further research should use the Which Outcome method to test the pattern of association of modifiers and outcomes with other verbal prob-

abilities (e.g., a chance, it is uncertain) and attempt to draw some regularities in people's preferences. For example, further research could test whether the pattern of association between verbal probabilities and modifier depends on the average degree of certainty believed to be conveyed by the probability term: low or medium probabilities (e.g., can, possible) could be more often associated with maximum outcome modifiers (i.e., up to) and very high degrees of certainty (e.g., will, certain) are associated with minimum outcome modifiers (i.e., at least). These findings could have strong implications for prediction studies where a medium probability of an outcome occurrence is expected to convey the same information than the medium probability of the outcome non-occurrence. For example, in framing studies, the prediction "If Program C is adopted, 400 people will die" (Tversky & Kahneman, 1981) was originally assumed to be logically equivalent to "If Program A is adopted, 200 people will be saved" given a set of 600 people. Yet, participants may instead consider that saving *at least* 200 people is not the same as having *at least* 400 people dying. The non-equivalence between the different readings of the predictions was hypothesised by Mandel (2001), who showed that a complete description of the different outcomes (e.g., 200 are saved and 400 die) cancels the framing effect. Our present findings suggest that the *at least* interpretation would also hold for descriptions involving certainty, for instance "If Program A is adopted, 200 people will be saved for sure".

7.2 Preference for extreme outcomes

When describing an outcome magnitude, participants chose maximum values to describe *possible* outcomes and minimum values for *certain* outcomes (Teigen et al., in press). These results hold even when the maximum value was the most likely and the bottom value the least likely. In Experiment 2, almost 90% of the participants chose to associate *certain* with the least likely outcome in apparent disregard of the quantitative meaning of what it takes to be certain. For example, in the computer battery vignette, participants read that the battery could last from 1.5 to 3.5 hours, and that only 10 batteries out of the 100 tested lasted 1.5 hours. The WO effects also held in a between-subject design, where participants completed *possible* or *certain* statements with outcome values from skewed U-shaped distributions (Experiment 3). Also in this case, participants chose to complete *certain* statements with low scores and *possible* statements with high scores, particularly when these scores belonged to the most frequent group of outcomes.

7.3 Preferences for large and small categories

When outcomes belonged to qualitatively different categories (e.g., car colours), magnitude extremeness could not be used as a cue for choosing which outcome was *not certain* or *possible*. In this case, most participants selected categories most closely matching the probabilistic meaning of the probability term, choosing for example the most frequent outcome as the one believed to be (*entirely*) *possible* and *not certain*. Yet, the least frequent category was also selected by a substantial minority, who apparently felt this category more appealing than the intermediate categories (which were almost never selected). This result suggests that some participants selected an outcome only based on its rarity. Rare events have in many contexts a special appeal, as testified by the old adage "praeclara sunt rara" (the extraordinary is rare), and by modern studies of the *scarcity principle* in consumer psychology (Ditto & Jemmott, 1989; Lynn, 1991).

7.4 The new approach does not apply to numerical probabilities

No WO effect was found with numerical probabilities, which appeared to be given a frequentist interpretation; participants chose the outcome whose occurrence matched most closely the probability communicated. Participants did not consider that a 90% probability should be associated with the minimum outcome, or that a 10% probability should be reserved for maximum values. Instead, 10% was used by a majority to describe the low tail of the distribution in the Battery and Jeans vignettes, which had exactly a 10% frequency of occurrence, whereas values from the high tail of the distribution were preferred in the Diet and Mail vignettes. Interestingly, these vignettes gave the outcome distributions in tabular form as whole numbers (1 of 10) rather than percentages (10%), which could have made the probability matching strategy less salient.

Numerical probabilities appear here to have the advantage of not being subject to contextual (pragmatic) considerations that affect the interpretation of verbal probabilities. This result is in line with the suggestions and findings of Windschitl and Wells (1996) who proposed that numerical probabilities elicit rule based strategies whereas verbal probabilities rely more on associative processes.

This does not exclude however, that under some circumstances, verbal probabilities could be interpreted more in line with quantitative expectations. Possibly, presenting distributions of probabilities instead of frequencies could elicit a use of verbal probabilities consistent with their probabilistic meaning. For example, when

knowing that a computer has a 10% chance to last 1.5 hours, participants could be less likely to say that this performance is *certain* than when based on frequency distributions.

7.5 Two interpretations of verbal probabilities

When people are asked to translate verbal expressions into numerical probabilities, in accordance with the “how likely” approach, *possible* and *not certain* outcomes are expected to be placed close to the midpoint of the probability scale. In the pre-test reported in the introduction, only one participant suggested that *possible* could mean a 10% probability, and only three participants translated *not certain* with 10% or less. However, in a context of “which outcome” questions, we have seen that outcomes with a 10% probability are frequently selected as representative of possible as well as not certain events.

These results demonstrate that standard numerical “translations” (obtained by the How Likely approach) of verbal probabilities are not applicable to all contexts. One possible source of the discrepancy between the results of the How Likely and a Which Outcome task could be that participants given the “how likely” question have a dichotomous outcome space in mind consisting only of the focal outcome and its complement (a battery lasting or not lasting in 3 hours), whereas those who are given the “which outcome” or “how much” questions are considering multiple outcomes.

One may suggest that the WO findings are caused by a specific design of the tasks. For example, participants were not given the opportunity to choose a 100% certain outcome, as the most frequent outcome had only around a 50% chance of occurring. Participants, finding no alternative that by itself could be regarded as a certain outcome, may have chosen to redefine the task as a question about minimum outcomes. According to this interpretation the WO effect would merely be the result of compliance with the response format, which forced participants to voice certainty in a situation where none of the outcomes is actually certain to occur.

However, several arguments can be marshalled against this view. First, if participants were reluctant to breach the probabilistic interpretation of verbal probabilities, they could choose one of the most likely outcomes (closer to 100%) instead of an unlikely bottom value. Note that participants associated the most likely outcome to a 90% numerical probability, suggesting that they were able to choose the outcome that best fitted a likelihood value. Second, the consistent preference for low values rather than high ones suggests substantial agreement about which values can be considered *certain*, as opposed to those that are more appropriately labelled *possible*. Fi-

nally, WO effects are also demonstrated for *possible* and in previous studies even for phrases like *there is a chance* or *it is unlikely* (Teigen et al., in press). These findings indicate that WO effects also occur where several outcome frequencies match the numerical meanings of the expressions and where people have the opportunity to select outcomes more consistent with the HL approach. The scope and robustness of the WO phenomena indicate that they cannot be dismissed as a purely methodological artefact. Instead, the WO phenomenon appears to rely on strong conversational habits.

7.6 Applied implications

Findings of the new approach highlight that risk and uncertainty communication is a process that is even more complex than expected.

Investigations applying the traditional probabilistic translation approach to verbal probabilities using “How Likely” methods have unveiled many potential risk communication problems. The recurrent between-subjects variability of the probability associated to different terms (e.g., Budescu & Wallsten, 1995) and their contextual dependence (e.g., Harris & Corner, 2011; Juanchich et al., 2012) represents two important and still unresolved challenges. On top of this complexity, we now know that a forecaster may say that “it is *certain* that it will rain 2mm tomorrow”, despite knowing that this exact amount of precipitation is in fact unlikely. It is well known in linguistics that numerals are not always given an “exactly” interpretation, but that “at least” and “at most” readings are also possible (Levinson, 2000; Musolino, 2004). It is for instance acceptable to say: “John has three children, perhaps four”, indicating that the first mentioned number can be considered true even if he has four children. The present research adds to this analysis by showing that *certain* is particularly suited to suggest an “at least” interpretation of scalar quantities. Compare the predictions A and B.

(A) Tomorrow we will have 2mm of rain.

(B) Tomorrow we will *certainly* have 2mm of rain.

Both forecasts A and B are ambiguous in the sense that they can mean that 2mm of rain is the *most likely* amount of precipitation, or that the expected amount is *at least* 2mm of precipitation (and most likely more). Our results suggest that the “at least” interpretation is more compatible with statement (B) than with statement (A).

Yet, misunderstandings will arise if the recipient of statement (B) expects that for certain it will rain just a little, instead of inferring that this will be the minimum amount. The gap between the pragmatic (e.g., *certain* associated with a 10% probable outcome) and probabilistic treatment of verbal probabilities (e.g., *certain* associated with a 90% probable outcome) indicates that a decision

based on such uncertainty expression may be ill informed and requires risk communication intervention.

Risk communication guidelines should take into account the actual usage of verbal probabilities, their pragmatic meanings, in addition to their probabilistic meanings, or at least try to disentangle the two. Results of Study 4 suggest that numerical probabilities are given a more literal interpretation than corresponding verbal expressions, indicating that ambiguity can be reduced when verbal probabilities are reported in conjunction with numerical ones. This combination has been shown to effectively reduce the interpersonal variability in the probabilistic meaning of verbal probabilities (Budescu, Broomell & Por, 2009; Budescu, Por & Broomell, 2011). The benefits of the numerical and verbal association might stem from the fact that numbers dampen the pragmatic interpretation and reinforce the probabilistic one.

If verbal probabilities are reported one should perhaps avoid the terms *certain* and *possible* in a context of magnitudes. *Certain* is problematic because it may be unclear whether the associated outcome value should be given an “exact” or an “at least” reading, as discussed above. *Possible* is problematic because, in principle, it can be used about all values with a non-zero probability, whereas in practice it will be used primarily about the highest one, regardless of frequency. It is accordingly not very informative about the probabilities involved. However, when both of these terms are used together (like in Experiment 1) they may be useful for capturing a speaker’s ideas about the range of outcomes as an alternative to confidence intervals, or worst case—best case estimates.

References

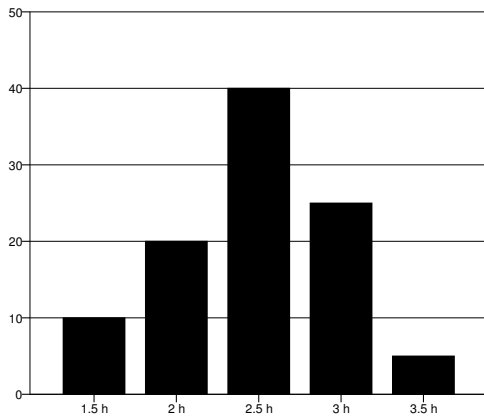
- Bonnefon, J. F., & Villejoubert, G. (2006). Tactful, or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, *17*, 747–751. <http://dx.doi.org/10.1111/j.1467-9280.2006.01776.x>
- Brun, W., & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, *41*, 390–404. [http://dx.doi.org/10.1016/0749-5978\(88\)90036-2](http://dx.doi.org/10.1016/0749-5978(88)90036-2)
- Budescu, D. V., Broomell, S. B., & Por, H. H. (2009). Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science*, *20*, 299–308. <http://dx.doi.org/10.1111/j.1467-9280.2009.02284.x>
- Budescu, D. V., Karelitz, T. M., & Wallsten, T. S. (2003). Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making*, *16*, 159–180. <http://dx.doi.org/10.1002/bdm.440>
- Budescu, D. V., Por, H. H., & Broomell, S. B. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic change*, *113*, 181–200. <http://dx.doi.org/10.1007/s10584-011-0330-3>
- Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *The Psychology of Learning and Motivation*, *32*, 275–318.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon’s Mechanical Turk. *Perspectives on Psychological Science*, *6*, 3–5. <http://dx.doi.org/10.1177/1745691610393980>
- Clarke, V. A., Ruffin, C. L., Hill, D. J., & Beamen, A. L. (1992). Ratings of orally presented verbal expressions of probability by a heterogeneous sample. *Journal of Applied Social Psychology*, *22*, 638–656. <http://dx.doi.org/10.1111/j.1559-1816.1992.tb00995.x>
- Dhami, M. K., & Wallsten, T. S. (2005). Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities. *Memory and Cognition*, *33*, 1057–1068. <http://dx.doi.org/10.3758/BF03193213>
- Ditto, H., & Jemmott, J. B. (1989). From rarity to evaluative extremity: Effects of prevalence information on evaluations of positive and negative characteristics. *Journal of Personality and Social Psychology*, *57*, 16–26. <http://dx.doi.org/10.1037/0022-3514.57.1.16>
- Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*, 889–906. <http://dx.doi.org/10.1037/0022-3514.38.6.889>
- Geurts, B., & Nouwen, R. (2007). At least et al.: The semantics of scalar modifiers. *Language*, *83*, 533–559. <http://dx.doi.org/10.1353/lan.2007.0115>
- Gourdon, A., & Beck, S., R. (2012). Overcoming the framing effect when making decisions based on verbal probabilities: Having more time is helpful but not enough. Manuscript submitted for publication.
- Harris, A. J. L., & Corner, A. (2011). Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1571–1578. <http://dx.doi.org/10.1037/a0024195>
- Honda, H., & Yamagishi, K. (2006). Directional verbal probabilities: Inconsistencies between preferential judgments and numerical meanings. *Experimental Psychology*, *53*, 161–170. <http://dx.doi.org/10.1027/1618-3169.53.3.161>
- Jou, J., Shanteau, J., & Harris, R. J. (1996). An information processing view of framing effects: The role of causal schemas in decision making. *Memory*

- and Cognition, 24, 1–15. <http://dx.doi.org/10.3758/BF03197268>
- Juanchich, M., Sirota, M., & Butler, C. L. (2012). The perceived functions of linguistic risk quantifiers and their effect on risk, negativity perception and decision making. *Organizational Behaviour and Human Decision Processes*, 118, 72–81. <http://dx.doi.org/10.1016/j.obhdp.2012.01.002>
- Juanchich, M., Teigen, K. H., & Villejoubert, G. (2010). Is guilt ‘likely’ or ‘not certain’? Contrast with previous probabilities determines choice of terms. *Acta Psychologica*, 135, 267–277. <http://dx.doi.org/10.1016/j.actpsy.2010.04.016>
- Juanchich, M., Sirota, M., Karelitz, T. & Villejoubert, G. (2013). Can membership-functions capture the directionality of verbal probabilities?. *Thinking and Reasoning*, 19, 2. <http://dx.doi.org/10.1080/13546783.2013.772538>
- Karelitz, T. M., & Budescu, D. V. (2004). You say probable and I say likely: Improving inter-personal communication with verbal probability phrases. *Journal of Experimental Psychology: Applied*, 10, 25–41. <http://dx.doi.org/10.1037/1076-898X.10.1.25>
- Keren, G., & Teigen, K. H. (2001). Why is $p = .90$ better than $p = .70$? Preference for definitive predictions by lay consumers of probability judgments. *Psychonomic Bulletin & Review*, 8, 191–202. <http://dx.doi.org/10.3758/BF03196156>
- Levinson, S. C. (2000). *Presumptive meanings*. Cambridge, MA: The MIT Press.
- Lynn, M. (1991). Scarcity effects on value: A quantitative review of the commodity theory literature. *Psychology and Marketing*, 8, 43–57. <http://dx.doi.org/10.1002/mar.4220080105>
- Mandel, D. R. (2001). Gain-loss framing and choice: Separating outcome formulations from descriptor formulations. *Organizational Behavior and Human Decision Processes*, 85, 56–76. <http://dx.doi.org/10.1006/obhd.2000.2932>
- Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition*, 93, 1–41. <http://dx.doi.org/10.1016/j.cognition.2003.10.002>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision making*, 5, 411–419.
- Reyna, V. (1981). The language of possibility and probability: Effects of negation on meaning. *Memory & Cognition*, 9, 642–650.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54, 93–105. <http://dx.doi.org/10.1037/0003-066X.54.2.93>
- Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, 101, 467–494. <http://dx.doi.org/10.1016/j.cognition.2005.11.001>
- Sirota, M., & Juanchich, M. (2012). To what extent do politeness expectations shape risk perception? Even numerical probabilities are under the spell! *Acta Psychologica*, 141, 391–399. <http://dx.doi.org/10.1016/j.actpsy.2012.09.004>
- Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, 68, 27–38. [http://dx.doi.org/10.1016/0001-6918\(88\)90043-1](http://dx.doi.org/10.1016/0001-6918(88)90043-1)
- Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica*, 88, 233–258. [http://dx.doi.org/10.1016/0001-6918\(93\)E0071-9](http://dx.doi.org/10.1016/0001-6918(93)E0071-9)
- Teigen, K. H., & Filkuková, P. (2013). Can > will: Predictions of what *can* happen are extreme, but believed to be probable. *Journal of Behavioral Decision Making*, 26, 68–78. <http://dx.doi.org/10.1002/bdm.761>
- Teigen, K. H., Juanchich, M., & Filkuková, P. (in press). Verbal probabilities: An alternative approach. *Quarterly Journal of Experimental Psychology*. <http://dx.doi.org/10.1080/17470218.2013.793731>
- Teigen, K. H., Juanchich, M., & Riege, A. (2013). Improbable outcomes: Infrequent or extraordinary? *Cognition*, 127, 119–139. <http://dx.doi.org/10.1016/j.cognition.2012.12.005>
- Theil, M. (2002). The role of translations of verbal into numerical probability expressions in risk management: A meta-analysis. *Journal of Risk Research*, 5, 177–186. <http://dx.doi.org/10.1080/13669870110038179>
- Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, 39, 176–190. <http://dx.doi.org/10.1287/mnsc.39.2.176>
- Windschitl, P. D. & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology-Applied*, 2, 343–364. <http://dx.doi.org/10.1037/1076-898X.2.4.343>

Appendix A. Vignettes and associated distribution used in Experiment 1

Vignette 1: Computer batteries

A sample of computers of the brand “Comfor” was tested to check how long the batteries last before they need to be recharged. All computers were used by students for lecture notes and similar purposes. The figure below shows how many batteries lasted for how many hours (duration is rounded to the nearest half hour).



There are two incomplete sentences below. Please, complete each sentence below with the number that seems most natural in this context. You may also add one of the suggested words (at least, around, at most or up to) if you feel this could improve the sentence To complete the sentence, drag and drop the items that you would like to use from the list on the left into the box on the right.

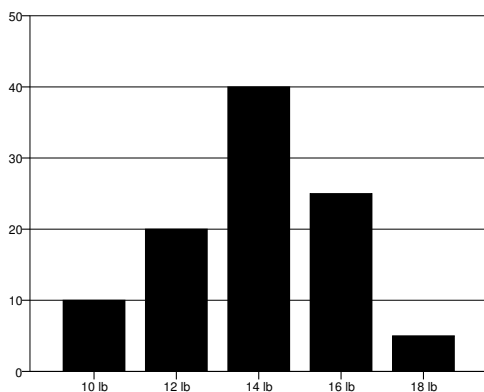
It is certain [possible] that the battery will last

_____.

- at least
- exactly
- around
- up to
- 1.5 hours
- 2 hours
- 2.5 hours
- 3 hours
- 3.5 hours

Vignette 2: Weight reduction

A weight reduction product "Taremare" based on seaweed shows the following weight loss (in pounds) for a sample of men and women adhering to the diet over a period of three months.



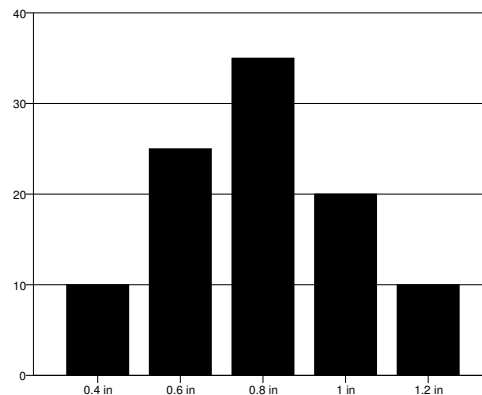
There are two incomplete sentences below. Please, complete each sentence below with the number that seems most natural in this context. You may also add one of the suggested words (at least, around, at most or up to) if you feel this could improve the sentence To complete the sentence, drag and drop the items that you would like to use from the list on the left into the box on the right.

By adhering to the Taremare program it is certain [possible] that one will lose _____

- _____.
- at least
- exactly
- around
- up to
- 10 pounds
- 12 pounds
- 14 pounds
- 16 pounds
- 18 pounds

Vignette 3: Jeans

A sample of jeans of the brand "Kenvelo" was machine washed in the regular way and tested for shrinkage. The figure below shows how much the jeans shrunk in length (shrinkage is rounded to the nearest tenth of inch).



There are two incomplete sentences below. Please, complete each sentence below with the number that seems most natural in this context. You may also add one of the suggested words (at least, around, at most or up to) if you feel this could improve the sentence To complete the sentence, drag and drop the items that you would like to use from the list on the left into the box on the right.

It is certain [possible] that the Kenvelo jeans will shrink _____.

- at least
- exactly
- around

up to
0.4 inch
0.6 inch
0.8 inch
1 inch
1.2 inches

Appendix B. Vignettes and associated distribution used in Experiments 2 and 3

Vignette 1: Computer batteries

“A sample of computers of the brand “Comfor” was tested to check how long the batteries last before they need to be recharged. All computers were used by students for lecture notes and similar purposes. The figure below shows how many batteries lasted for how many hours (duration is rounded to the nearest half hour)”.

Experiment 1: Unimodal bar graph with five bars accompanying the vignette showed a monotonically increasing or decreasing distribution of durations, from 1.5 hours (5% or 40%) to 3.5 hours (40% or 5%), with intermediate values: 2 hours (10% or 25%), 2.5 hours (20%), 3 hours (25% or 10%)

Experiment 2: Bimodal bar graph with five bars accompanying the vignette showed a U shaped distribution of durations, with either a mode on the left or on the right. For a left mode distribution, participants read the following values: 1.5 hours (35%), 2 hours (10%), 2.5 hours (10%), 3 hours (10%) and 3.5 hours (45%).

Based on these results, what is natural to say? Complete the two sentences below in a way that seems natural in this context:

It is certain that the battery in a Comfor computer will last for hours

It is possible that the battery in a Comfor computer will last for hours

Vignette 2: Weight reduction

“A weight reduction product “Taremare” based on seaweed shows the following results for ten men and women following the diet over a period of three months”.

Experiment 1: The results of ten individual dieters were listed in a table with three columns: Weight before, weight after and weight reduction (calculated as the difference between the first two values). For an increasing distribution, weight loss ranged from 3 kg (one person) to 9-10 kg (four persons), with 7 kg as the median value; for a decreasing distribution four persons lost 3-4 kg and only one lost 10 kg.

Experiment 2: Unimodal bar graph with four bars accompanying the vignette showed a U shaped distribution of durations, with either a mode on the left or on the right. For a left mode distribution, participants read the following values: 10 lb (30%), 12lb (15%), 14lb (15%), and 16lb (40%).

The sentences to be completed were:

It is certain that with the Taremare program you will lose lb

It is possible that with the Taremare program you will lose lb

Vignette 3: Jeans

A sample of jeans of the brand “Kenvelo” was machine washed in the regular way and tested for shrinkage. The figure below shows how much the jeans shrunk in length (shrinkages are rounded to the nearest half cm - 0.2 inch).

Experiment 1: An unimodal bar graph with five bars showed a distribution of shrinkage from 1cm (10%) to 3 cm (35%), with as intermediate values: 1.5 cm (10%), 2 cm (20%), 2.5 cm (25%). (Percentages in parentheses apply to increasing condition)

Experiment 2: An unimodal bar graph with five bars showed a distribution of shrinkage from 0.4 inch (25% or 45% as a function of the location of the mode) to 1.2 inch (45% or 25% as a function of the location of the mode), with as intermediate values: 0.6 inch (10%), 0.8 inch (10%), 1 inch (10%).

The sentences to be completed were:

It is certain that Kenvelo jeans will shrink _____ in after washing.

It is possible that Kenvelo jeans will shrink _____ in after washing.

Vignette 4: Mail

(Experiment 1 only)

The postal services investigate how long it takes to send a letter from Norway to various addresses in the US. Ten letters are mailed on a regular Monday at 3 pm.

1 (4) letters arrives on Wednesday

2 (3) letters arrive on Thursday

3 (2) letters arrive on Friday

4 (1) letters arrive next Monday

The sentences to be completed were:

It is certain that a letter from Norway to USA will take _____ days.

It is possible that a letter from Norway to USA will take _____ days.

Vignette 5: Shmulp

(Experiment 2 only)

Please read the following vignette; don't worry if you do not understand the meaning of some words. Try to complete the statement with the expression that sounds most appropriate given the context.

Shmulps have a different number of glomps (glps). The graph below shows to what extent a sample of Shmulps is glomping. A bimodal bar graph with four bars showed a distribution of shmulp from 1 glp (45% or 35%, as a function of the mode location) to 4 glps (35% or 45%), with the intermediate values: 2 glps (10%), and 3 glps (10%).

The sentences to be completed were:

It is certain that a shmulp will have ____ glps.

It is possible that a shmulp will have ____ glps.

Shmulp

Please read the following vignette; no worry if you do not understand the meaning of some words. Try to complete the statement with the expression that sounds most appropriate given the context.

There exist different types of shmulps. 10% are gelering, 20% are laurding and 70% are glinpsing.

You wonder which kind of shmulp you will receive and someone responds to you: "It is possible [entirely possible; not certain] that you will receive a ____ shmulp".

Appendix C. Vignettes used in Experiment 4

Teacher

Imagine that Lea is a new teacher who will be the tutor of a student this year. In Lea's school, students are attending three different programmes: 10% of the students attend programme A, 20% attend programme B and 70% attend programme C. Lea asks a colleague which programme her tutee will attend, and the colleague responds:

"It is possible [entirely possible; not certain] that the student will attend programme ____."

Transport

Jane either goes to work by car, bus, or by bike. Her husband says: "Jane goes to work by car 60% of the time, by bus 30% and by bike 10%, so it is possible [entirely possible; not certain] that tomorrow she will go to work by _____."

Car

Jason won a car on an internet lottery. He is wondering about the colour of the car he won. Asked about that, the secretary of the lottery company says: "5% of the cars are white, 10% are red, 20% are grey and 65% are black, so it is possible [entirely possible; not certain] that you will have a ____ car".