Note: This is not the final version of the manuscript (but almost!). To see the final version: *Journal of Risk Research*, Nov. 2014; doi: 10.1080/13669877.2014.983945

How to improve people's interpretation of probabilities of precipitation?

Marie Juanchich and Miroslav Sirota

Authors' affiliations:

Marie Juanchich: Department of Management, Kingston Business School, Kingston University.

Miroslav Sirota: Department of Primary Care & Public Health Sciences, King's College London

Authors' mail address:

Marie Juanchich, Kingston Business School, Kingston University, Kingston Hill, KT2 7LB, UK.

Miroslav Sirota, is now at Department of Psychology, Kingston University, Penrhyn Road, Kingston upon Thames, Surrey KT1 2EE, United Kingdom. Email: miroslav.sirota@kingston.ac.uk, Phone: +44 (0)20 8417 2896

Correspondence concerning this article should be addressed to Marie Juanchich, Department of Management, Kingston Business School, Kingston University, Kingston Hill, Kingston upon Thames, KT2 7LB. E-mail: M.Juanchich@kingston.ac.uk. Tel: (+44) 20 84 17 59 34.

Acknowledgment: We thank Professor Karl Halvor Teigen and Professor Yannis Georgellis for comments on an earlier draft.

Note: This is not the final version of the manuscript (but almost!). To see the final version: *Journal of Risk Research*, Nov. 2014; doi: 10.1080/13669877.2014.983945

Abstract

Most research into uncertainty focuses on how people estimate probability *magnitude*. By contrast, this paper focuses on *how* people interpret the concept of probability and *why* they often misinterpret it. In a weather forecast context, we hypothesized that the absence of an explicit reference class and the polysemy of the percentage format are causing incorrect probability interpretations and test two interventions to help people make better probability interpretation. In two studies (*N* = 1337), we demonstrate that most people from the United Kingdom and the United States do not interpret probabilities of precipitation correctly. The explicit mention of the reference class helped people to interpret probabilities of precipitation better when the target area was explicit; but this was not the case when it was not specified. Furthermore, the polysemy of the percentage format is not likely to cause these misinterpretations, since a non-polysemous format (e.g., verbal probability) did not facilitate a correct probability interpretation in our studies. A Bayes factor analysis supported both of these conclusions. We discuss theoretical and applied implications of our findings.

Keywords: Weather forecast; probability of precipitation; probabilistic format, probabilistic format preference; Bayes Factor analysis.

Note: This is not the final version of the manuscript (but almost!). To see the final version: *Journal of Risk Research*, Nov. 2014; doi: 10.1080/13669877.2014.983945

Why people do not understand probabilities of precipitation: Effects of forecast formats.

INTRODUCTION

Prior research has focused mainly on probability magnitude perceptions. For instance, how people's characteristics (e.g., white male effect; Olofsson & Rashid, 2011), the personality of the speaker (Juanchich, Sirota, & Butler; 2012, Sirota & Juanchich, 2012), or different presentation or assessment formats (e.g., Smerecnik, Mesters, Kessels, Ruiter, De Vries, & De Vries, 2010; Riege & Teigen, 2013) affect probability perception magnitude. This trend of research has been very fruitful and has shown, for example, that white males perceive negative outcomes to be less probable than white females or people from ethnic minorities (Olofsson & Rashid, 2011), or that people perceive a tactful speaker to convey a higher probability of a negative outcome than a plain speaker (Juanchich et al., 2012), or that graphic representations of statistical information have a positive effect on probability perception accuracy (Smerecnik, et al., 2010).

This strong emphasis on probability magnitude perception contrasts sharply with a lack of investigation into whether people actually understand probabilities correctly. Indeed, although people may say that an outcome is likely, findings indicate that they have difficulty in assessing what is actually quantified by a probability, especially for single event probabilities (e.g., Gigerenzer, Hertwig, van den Broek, Fasolo, & Katsikopoulos, 2005). This misinterpretation of probabilities may lead to risk misperception and ill-informed decision-making. Therefore, the focus of the present paper is on *how* people understand probability, and on the possible factors that could drive probability misinterpretation. More specifically, because it is one of the most common probabilistic messages, we investigate how people understand probability of precipitation.

Probability of precipitation (PoP)

A PoP is the probability that measurable precipitation (more than 0.005 mm) will occur at a specific point (i.e., a rain gauge) in a specific period of time (Murphy & al., 1980; Rogell, 1972). Probabilities of precipitation are computed, based on the proportion of days like tomorrow where a measurable precipitation is observed from a sample of *days like tomorrow* (Gigerenzer et al., 2005; Joslyn, Nadav-Greenberg, Nichols, 2009). Thus the correct reference class of a PoP is: "days like tomorrow"(Gigerenzer et al., 2005; Juslyn et al., 2009). This interpretation will be thereafter labelled as the *Days' interpretation.* A 20% probability of rain is thus derived from a sample of days like tomorrow in which 20% of the days feature a measurable precipitation. Note that there exists a debate about how to best define probability of precipitation, and that even when meteorologists agree they may use different wordings. We chose to use here the definition that appears to generate the strongest consensus (Gigerenzer et al., 2005).

(Mis-)interpretation of probabilities of precipitation

The financial benefits of probabilistic weather forecasts (e.g., there is a 70% chance that it will rain) as opposed to categorical forecasts (e.g., it will rain tomorrow) are now commonly accepted (National Research Council, 2003, 2006). Because of their precise nature, numerical probabilities are considered the best format to express degrees of certainty (Murphy, Lichtenstein, Fischoff, & Winkler, 1980; Winkler, 1990) and their introduction is, thus, largely recommended (American Meteorological Society, 2008; National Research Council, 2003, 2006). And yet, previous research has consistently shown that PoPs are often misunderstood by weather forecast users (Murphy et al., 1980; Sink, 1995; Gigerenzer et al., 2005; Joslyn, et al., 2009; Morss, Demuth, & Lazo, 2008; Morss, Lazo, & Demuth, 2010). When asked to select an interpretation of a PoP from a list of possible interpretations, more than 50% of people surveyed believed that it referred to a proportion of time or region in which it would rain. Most people believed, for example, that a 20% probability of rain meant that it would rain 20% of the time (i.e., Time interpretation) or in 20% of the region (i.e., Region interpretation; Morss, et al, 2008). Note that, based on the formula of computation of PoP, it could be said that the Region interpretation (e.g., “it will rain in 30% of the region”) is not always a poor proposition, as it could be correct in a specific situation. Indeed, the Region interpretation is true in the case where the probability of rain in an area is 100%.

Juslyn et al. (2009) investigated formats that would enable a better interpretation of PoPs. Their results showed that an iconic representation of PoPs (e.g., a pie chart representing rain in 30% of the chart) slightly improved PoP interpretation but not to a statistically significant degree (Experiments 1 and 2). A format that did help participants was the mention of both the positive and the negative framings of the probability (e.g., there is a 30% chance that it will rain; there is a 70% chance that it will not rain). In this format, only 36% of participants made a reference class error, whereas 64% made the error with a classical PoP (e.g., there is a 30% probability that it will rain). The mention of the two complementary probabilities highlights that there is also a probability that it will *not* rain, which is inconsistent with the Region and Time interpretations where rain is certain. Knowing that some formats help people to identify the correct reference class of PoP is very useful, but the research conducted so far has not informed the reason why PoP is so commonly misunderstood.

Why do weather forecast users not understand probability of precipitation?

We propose to explore two non-exclusive explanations that could account for the misinterpretations of PoP (namely, Time and Region misinterpretations). The first explanation emphases the ambiguity of the reference class of single event probabilities in general, whereas the second explanation focuses on the polysemy of a particular format of percentages that triggers the ambiguity of the reference class.

The ambiguous reference class of single event probabilities. Gigerenzer et al. (2005) suggested that PoP misinterpretation arises from the ambiguity of the reference class of single event probabilities. In contrast to probabilities of reproducible events, (e.g., die roll), probabilities of single events (e.g., probability of rain), are deemed to refer to an ambiguous reference class (Gigerenzer et al., 2005). The reference class of a probability refers to the class of event that is sampled to produce the probability. Therefore. the perceived reference class of a probability determines the perceived meaning and implications of a probability. For example, when being told that Prozac has a 20% chance of causing a sexual problem, patients need to identify the class of events to which the probability refers (Gigerenzer, 2002). Some patients identify the reference class incorrectly by, for example, thinking that the probability refers to the number of sexual encounters. This wrong interpretation leads patients to believe that they are very likely to experience sexual problems since they think that they will have a sexual problem in 1 in 5 of the times they have sexual intercourse. Other patients identify the reference class correctly as the group of people taking Prozac. This correct interpretation leads patients to believe that they have a small chance of having a sexual problem, because they recognise that they have only a 1 chance in 5 of experiencing any sexual problems because of taking Prozac.

In the case of PoPs, they are labelled as ambiguous, because people reading a PoP believe that its reference class can be a proportion of time (Time misinterpretation), a proportion of space (Region misinterpretation) or a number of days like tomorrow (Days – correct - interpretation).

Importantly, explaining the difficulty for people to understand PoPs correctly by the ambiguity of their reference class, means that PoPs in general will be misinterpreted, whatever the probabilistic format used. This means that whether a PoP is numerical (e.g., 30%, 3 in 10) or verbal (e.g., unlikely, likely), should not change the ability of people to identify the reference class of a probability.

The polysemy of percentage. In addition to the hypothesis that all single event probabilities imply an ambiguous reference class, we suggest that PoP misinterpretations arise because one of the percentage formats used to convey the single event probability is polysemous (i.e., the words have different but related meanings). Indeed, a percentage such as "30%" can be commonly used to quantify probabilities, proportions or frequencies. For example, one could say, "there is a *30%* chance that it will rain", or “tomorrow it will rain in *30%* of the area” or even “it will rain for *30%* of the day”. Note that the polysemous hypothesis does not apply only to percentages. For example, ratios can also be commonly used to describe probabilities and proportions of time or area (e.g., it will rain in 1/3 of the area). Research into how people make sense of polysemous words has shown that the different meanings of the word compete and become available in people’s minds (e.g., Rodd, Gaskell, & Marslen-Wilson, 2002). According to the polysemy hypothesis, the competitive meanings of percentages as a proportion of days, time or region are stored in people’s memories and are all activated when a person reads a percentage. The Days, Time and Region interpretations therefore become available as valid hypothetical answers. The possibility that the existence of the different Days, Time and Region meanings of percentages are all made available when one is reading a PoP and that this triggers a high rate of these interpretations is also in line with literature on the availability bias, showing that information that is more available is perceived to be more likely than information that is less available (Tversky &Kahneman, 1973). The fact that the Days, Time and Region interpretations become available would lead to the perception that they are likely to be correct.

Importantly, the percentage polysemy hypothesis entails that it is specifically the numerical probability format that is causing a high rate of PoP misinterpretation (e.g., percentages). In contrast, the use of another probabilistic format that cannot be used interchangeably to denote a proportion of time or of space should boost the rate of correct interpretations.

How can we improve the interpretation of PoP?

The reference class specification solution. To overcome the ambiguity of reference class in single event probabilities, the recommendation of Gigerenzer et al. was to specify that a 30% probability of rain means that “…3 out of 10 times when meteorologists make this prediction, there will be at least a trace of rain the next day” (Gigerenzer et al., 2005, p. 629). This research was tailored to test the possible benefits to forecast interpretations of such an explanatory procedure, compared with a simple statement of probability of precipitation. It is important to note that an explicit mention of the reference class is also expected to prevent wrong interpretations based on the polysemy of percentages. Yet, if wrongful interpretations are caused by the percentage polysemy, other solutions can be explored.

The use of a non-polysemous format. To overcome the postulated negative effect of the polysemy of percentages, a non polysemous probabilistic format could be used. Verbal probabilities (e.g., there is a chance, it is likely) are linguistic probability quantifiers that cannot be used to describe proportions of time or region. Verbal probabilities are thus not polysemous. For example, one can say “it is unlikely that it will rain tomorrow” but cannot say “it will rain in *it is unlikely* of the time” nor “it will rain in *it is unlikely* of the area”. According to the polysemy hypothesis, a verbal probability forecast would be interpreted more correctly than the traditional percentage PoP. The use of verbal probabilities has been investigated in the past but in studies that suffered from important shortcomings, preventing any decisive conclusions concerning the potential benefits of verbal probabilities. For example, Murphy et al. (1980) asked participants to interpret both a numerical (i.e., 30%) and a verbal forecast (i.e., likely) and found that, with both formats few people chose the correct interpretation (39% and 28%). Yet, Murphy et al. did not test whether this difference was significant. Further, the cross sectional design does not rule out an order effect to explain the differences of performance. Finally, the two probabilities compared vary in more than just format, since the two formats also refer to different probability magnitudes: 30% is a small probability, whereas ‘likely’ reflects a probability of around 70% (e.g., Wallsten & Budescu, 1995).

Based on the polysemy hypothesis, the verbal probability format is expected to enable better interpretation. It is nevertheless acknowledged that this format has some drawbacks. Indeed, verbal probabilities elicit great variability in their probabilistic meaning between individuals and even in the same individual over time (Wallsten & Budescu, 1995).

The dual format solution. Renooij and Witteman (1999) and Witteman, Renooij and Koele (2007) posited that associating verbal probabilities with numerical ones would reduce the variability of the probabilistic meaning of linguistic expression, while preserving the ease of understanding. This method has the advantage of stabilising the probabilistic meaning of the expressions from one context (e.g., prediction of rain) to another (e.g., prediction of storm) and to reduce between-subjects variability. Budescu and colleagues (Budescu, Broomell & Por 2009; Budescu, Por & Broomell 2012; Budescu, Por, Broomel & Smithson 2014) produced evidence of the benefits of this method in climate change risk communication.

Hypotheses. The aim of the present research is to investigate the reason why probabilities of precipitation (PoPs) are so hard to understand. Specifically, we examine whether the absence of the ambiguity of the reference class of single event probabilities or the polysemy of percentages cause incorrect interpretations. To test these two hypotheses, we have compared weather forecast understanding based on a traditional PoP to a PoP associated with a description of the reference class, and to a verbal probability forecast, in three samples, two from the United States and one from the United Kingdom.

Study 1

Method

Participants. The total sample was composed of 953 participants from the US and UK, aged between 18 and 82 (M = 42.52, SD = 15.40); 55.9% were females. Thirty participants did not report their socio-demographic information (16 Americans and 14 British). Given that the level of exposure to probabilities of precipitation had an effect on PoP interpretation in the past (Gigerenzer, et al., 2005), and that the UK and the US introduced PoPs at different times, we present the samples and the data analyses separately.

The sample of Americans was composed of 339 Mechanical Turk workers who completed the 3 minute survey on communication in exchange for 0.10$. Amazon Mechanical Turk is recognised as a reliable source of data for research in Social Sciences with a pool of participants featuring varied socio-demographic characteristics (e.g., Buhrmester, Kwang & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010). Participants were aged between 18 and 82 (Mdn = 28) and 56.0% were females. Most of them were Caucasian (81.4%) and part of the active work force (66.3%); 27.9% were unemployed and 5.9% retired. Only 0.3% did not have a formal education, 51.0% had a high school degree and 48.7 had a higher education degree. At the time of the data collection, people in the US had benefited from probabilistic weather forecasting for 46 years. (1965-2011).

The sample from the UK was composed of 614 individuals contacted by a marketing agency. British participants were rewarded by a voucher. They were aged between 18 and 82 (Mdn = 49) and 55.8% were females. Most of them were Caucasian (80.5%) and part of the active work force (69.8%); 9.2% were unemployed and 21% retired. Only 6.7% had no formal education, 25.5% had a GCSE, 26.3% had a higher school qualification and 41.3% had a higher education degree. At the time of the data collection, people in the UK had benefited from probabilistic weather forecasting for 19 years (1992-2011).

Age, gender, job and ethnicity did not affect weather forecast interpretation; these variables were thus not integrated into further analyses.

Materials and Procedure. Participants read and accepted a brief informed consent form. Then they read one of the four forecasts presented in Table 1 (randomly allocated). The verbal probability was pre-tested to communicate a probability of 30% on average in the two samples studied (US: N = 56, M = 26.78, SD = 27.86; UK: N = 296, M = 31.11, SD = 24.34).

Note: This is not the final version of the manuscript (but almost!). To see the final version: *Journal of Risk Research*, Nov. 2014; doi: 10.1080/13669877.2014.983945

Table 1

List of the formats of probability of precipitation presented to participants.

Formats	Forecast
Numerical probability (NP)	There is a 30% chance that it will rain tomorrow.
Numerical probability with reference class (NP + RefClass)	There is a 30% chance that it will rain tomorrow*. *This means that in 3 out of 10 times, when meteorologists make this prediction, there will be at least a trace of rain the next day.
Verbal probability	It is unlikely that it will rain tomorrow.
Verbal probability and numerical translation	It is unlikely that it will rain tomorrow * *(It is unlikely = 30% chance).

For the numerical probability format with reference class and the dual format, the additional information (the reference class and the numerical probability respectively) was presented as a foot note on the forecast page, with an asterisk as shown in Figure 1.

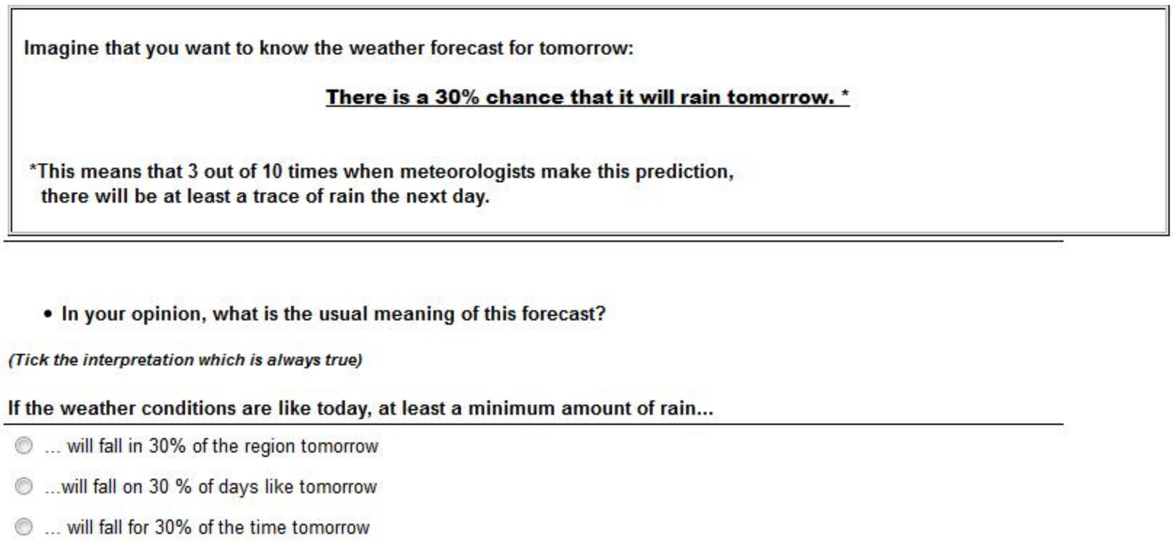

Figure 1.

Example of material used in Study 1 to provide a probability of precipitation and to assess its interpretation (experimental condition numerical probability with reference class).

Participants were asked to imagine that they were wondering about the weather for the next day and were then provided with a forecast. After reading the forecast, participants read a list of three interpretations taken from Gigerenzer et al. (2005). Participants were instructed to select the interpretation of the forecast that was always true from among three possibilities as shown in Figure 1. The correct interpretation is the Days' interpretation (second option in Figure 1). The mention 'always true' was introduced in the original design from Gigerenzer and was kept in our design because the Region forecast can be correct when the probability of rain is equal to 100%. The three options were presented in a randomised order for each participant. The proposed PoP interpretations were worded exactly as in Gigerenzer et al. (2005), which was different from the wording of Murphy et al. (1980). For example, in the former, the correct interpretation was phrased as "the occurrence of precipitation at a particular point in the forecast area". Finally, participants reported socio-demographic information.

RESULTS

More than half of the participants interpreted the PoP incorrectly as either a proportion of time (i.e., it will rain 30% of the time) or of region (i.e., it will rain in 30% of the region) – see Table 2. The Days' interpretation was nevertheless the most common (i.e., it will rain in 30% of days like tomorrow) as described in the left panel of Figure 2, which shows the proportion of correct Days' interpretations according to the weather forecast format and in the two samples.

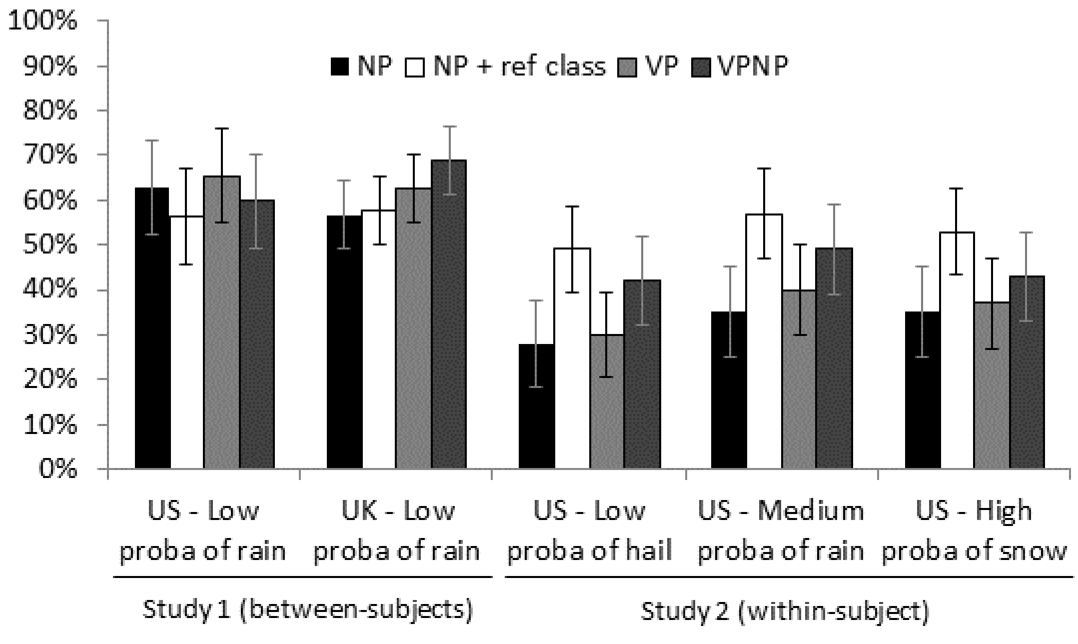

Figure 2. Proportion of correct interpretation of weather forecasts as a function of the precipitation forecast format in Studies 1 and 2.

Note: Error bars represent 95% confidence intervals.

The format of presentation of the probability did not affect the PoP interpretation chosen by participants in the American sample nor in the British sample, χ^2 (6, *N* = 339) = 3.13, *p* = .793, φ = .10, χ^2 (6, *N* = 614) = 5.34, *p* = .501, φ = .09. Overall (when combining the data from the two samples), the explicit mention of the reference class did not improve the proportion of correct interpretations of the probability of rain, but acted to the contrary (- 2% compared to the simple numerical condition; see last row of the top panel of Table 2).

Further, in comparison with the forecasts that included a numerical probability, verbal probabilistic forecasts elicited a slightly higher rate of correct interpretation (+ 0.5%). Finally, the presence of both numerical and verbal formats was associated with the highest rate of correct interpretation (+ 4.6 compared to the simple numerical condition). However, the effect of the format of presentation of the probability did not have a statistically significant effect, χ^2 (6, *N* = 953) = 6.51, *p* = .368, Cramer's *V* = .08.

When we compare the findings of Murphy et al. (1980) with the present ones, we can conclude that in 30 years the rate of correct interpretation in America has increased substantially: + 6% in the numerical probability condition (39% to 45%) and + 30% in a verbal probability condition (28% to 58%). However, the difference in the rate of correct interpretation could also be due to methodological (e.g., the use of a different question formulation, the use of a different set of response options) or statistical (e.g., high sampling errors given a small sample size, *N* = 78) reasons. The method we used was the same as Gigerenzer et al. (2005) who collected their data in 2002 in both Europe and the US. By contrast with their findings, the present findings also indicate an increase in the rate of correct interpretations of 8% in the last 12 years (37% to 45%).

Note: This is not the final version of the manuscript (but almost!). To see the final version: *Journal of Risk Research*, Nov. 2014; doi: 10.1080/13669877.2014.983945

Table 2.

Probability of precipitation interpretations as a function of the format of the probability in Studies 1 and 2. The formats were the following: numerical probability - NP, numerical probability with explicit mention to the reference class- NP + RefClass, verbal probability – VP and verbal probability with numerical translation - VP + NP.

Study 1	NP	NP + RefClass	VP	VP + NP	Total
UK sample (*N* = 614)					
Time	27.7%	34.7%	28.9%	31.9%	30.9%
Region	29.7%	26.6%	32.2%	22.7%	27.9%
Days	*44.7%*	*42.7%*	*45.2%*	*49.3%*	*45.3%*
US sample (*N* = 339)					
Time	29.5%	33.7%	26.9%	29.3%	30.1%
Region	21.8%	16.8%	15.4%	14.6%	17.1%
Days	*48.7%*	*49.5%*	*57.7%*	*56.1%*	*52.8%*
Total (*N* = 953)					
Time	28.3%	34.3%	28.3%	30.9%	30.6%
Region	27.0%	23.0%	26.5%	19.7%	24.0%
Days	*44.7%*	*42.7%*	*45.2%*	*49.3%*	*45.3%*
Study 2 - UK (N = 384)	NP	NP + RefClass	VP	VP + NP	Total
Rain					
Time	32.7%	26.5%	34.4%	24.5%	29.4%
Region	13.3%	10.2%	17.8%	15.3%	14.1%
Days	*34.7%*	*57.1%*	*40.0%*	*49.0%*	*45.3%*
Other	19.4%	6.1%	7.8%	11.2%	11.2%
Hail					
Time	40.8%	29.6%	36.7%	30.6%	34.4%
Region	13.3%	15.3%	27.8%	14.3%	17.4%
Days	*27.6%*	*49.0%*	*30.0%*	*41.8%*	*37.2%*
Other	18.4%	6.1%	5.6%	13.3%	10.9%
Snow					
Time	37.8%	29.6%	35.6%	35.7%	34.6%
Region	8.2%	10.2%	15.6%	10.2%	10.9%
Days	*34.7%*	*53.1%*	*36.7%*	*42.9%*	*41.9%*
Other	19.4%	7.1%	12.2%	11.2%	12.5%

Note: The proportions of correct weather forecast interpretation are in italics. In Study 2 the weather forecast events were presented in within-subjects.

The traditional statistical analyses as conducted here do not assess the extent to which the data supports the null effect of format on PoP interpretations. Therefore, to quantify support for the null effect hypothesis, we conducted a Bayes Factor analysis. This type of analysis is now recommended in psychology in general, especially when establishing support for null hypotheses (Rouder, Speckman, Sun, Morey & Iverson, 2009; Wagenmakers, Wetzels, Borsboom & van der Maas, 2011). In our analysis, the Bayes factor provides the ratio of marginal likelihood of the data, given that there is no effect of format, (H_0) to the probability of the data, given that formats affect the interpretation of PoPs, (H_1). A Bayes factor greater than 1 ($BF_{01} > 1$) indicates supporting evidence for H_0, whereas a Bayes factor lower than 1 ($BF_{01} < 1$) indicates supporting evidence in favour of H_1. Further, a greater departure from 1 indicates stronger evidence (Wetzels, Matzke, Lee, Rouder, Iverson, & Wagenmakers, 2011). For example, a Bayes Factor ranging from 1/100 to 1/30 indicates the existence of very strong evidence in favour of H_1 whereas a Bayes Factor ranging from 30 to 100 indicates very strong evidence in favour of H_0 (cf. Wetzels et al, 2011).

The Bayes factor analysis for the proportion of correct PoP interpretation yielded substantial evidence supporting the null hypothesis that the formats manipulated here did not affect the PoP interpretation, $BF_{01} = 96.9$ (assuming a uniform distribution prior; see Albert, 2009). This result means that the data are almost 97 times more likely under H_0 (assuming a null effect of formats) than H_1 (assuming any effect of formats). Such evidence is considered to provide *very strong* support for the null hypothesis (Wetzels et al, 2011).

Study 2

This study aimed to test the robustness of the findings of Study 1 in an improved experimental design and to extend previous findings to two new precipitation events (snow and hail) and two new probability magnitudes (medium and high). An additional goal was to test which format participants would prefer to receive when seeking weather forecast information.

The design of Study 2 overcomes two methodological limitations of Study 1: the force choice setting and the fact that in Study 1 the area for which the forecast was formulated was not described (e.g., it is likely that it will rain – but where?).

Indeed, in Study 1, participants had to choose between three options which might or might not have matched their subjective interpretations of the forecast. The forced choice may have contributed to our findings that the different formats did not affect participants' interpretations of a probability of precipitation. Perhaps participants had in mind a correct interpretation but did not recognise it in one of the three provided options. In turn, some participants may have chosen an interpretation that they did not really believe in, creating some variability. In study 2, we gave participants the possibility of providing their own personal interpretation. This should provide a better setting to test the possible benefits of our interventions (providing an explicit reference class and using a verbal probability), as it should reduce this source of variability and increase the data validity.

A second limitation of study 1 was that participants received a forecast without a clear definition of the area to which the forecast applied (i.e., target area). Therefore, participants had to imagine the area that the forecast focused on, which could have created variability in data. For example, a participant could have thought about her village, another about her State. This variability may have had an impact on people's interpretations, by, for example, affecting the rate of selection of the Region interpretation, more relevant if one thinks about a big area like a state than a very small one like a village. Specifying the target area of the forecast should decrease this source of variability and increase the validity of the data.

Method

Participants. The final sample consisted of 384 Amazon Mechanical Turk workers who completed the 3-minute survey on weather forecast in exchange for 0.30$. The sample size was determined in two steps. First, based on a power calculation, we determined an initial sample of 341 participants would be needed to detect a small effect $w = 0.2$, assuming $df = 6$, $\alpha = 0.5$ and power $(1-\beta) = 0.8$. Second, we increased the initial sample size by + 10% to account for potential participant attrition to reach the final sample size of 384. Participants were aged between 18 and 72 (*Mdn* = 28) and 33.6% were females. Most of them were Caucasian (77.3%) and part of the active work force (80.2%); 17.7% were unemployed and 2.1% retired. Only 0.5% did not have a formal education, 32.3% had a high school degree and 67.2% had at least a College degree. At the time of the data collection in the US, people had benefited from probabilistic weather forecasting for 49 years at the time of the data collection (1965-2014).

Design, Procedure and Materials. The format was manipulated in a between-subjects design in three vignettes describing different precipitation forecasts: a low probability of snow, a medium probability of rain and a high probability of hail. Participants therefore read the three weather forecast vignettes and, for each, selected the interpretation that was correct. The interpretation selection task featured the same three interpretations as in Study 1 (Time, Region and Days) which were presented in a randomised order for each participant. Along with those three interpretations, participants had the option to specify a personal interpretation in case none of the above-mentioned appeared satisfactory. Then participants indicated their format preferences when looking for three probabilities of precipitation events (i.e., rain, hail and snow). For each PoP event, participants could select among five format options: numerical, numerical with explanation, verbal, verbal-numerical or other. Each format was provided with an example featuring a low probability of occurrence of the precipitation in question. For example: "*Numerical format. E.g., there is a 30% chance that it will rain"*. The option 'I prefer another format' gave the possibility for participants to specify their own preference in a text box. Each question appeared on a different page and the order of the different forecast interpretations and the order of the format preference questions were randomly presented for each participant. Finally, participants reported basic socio-demographic characteristics.

Results

Overall

Participants interpreted correctly only 41% of the precipitation forecasts across the different precipitation conditions (snow, rain and hail). Overall, the most frequent answer was the Days interpretation (41%), followed by the Time interpretation (33%) and the Region interpretation (14%). The Other interpretation was the least common answer (11%). Most personal interpretations consisted of a reformulation of the weather forecast. There were different types of personal interpretations. The first type of personal interpretation consisted in reformulating the forecast very closely but specifying the quantity of precipitation expected ("30% chance it will snow *at all*"; "70% chance it will hail *at all"*). The second type of interpretation consisted in reformulating the probability quantifier (i.e., replacing the term 'chance' by 'probability' or 'likelihood' – e.g., "30% likelihood that it will snow in Wales tomorrow, if the conditions stay the same as today"). The third type of personal interpretation consisted in swapping the forecasted precipitation for its alternative outcome (e.g., reframing a low probability of snow by a high probability of no snow; e.g., "it will most likely not snow") or complementing the forecasted precipitation with its alternative outcome

(There is a 30% chance that snow will fall in Wales tomorrow, 70% chance it will not). In a few instances, the personal interpretation was a mix of several types of interpretation: for example, a mix of quantifier adaptation and framing (e.g., likelihood is 50% chance of rain, 50% chance of no rain at all"; "I assume it to mean there is a 70% likelihood that it will hail, while there is a 30% chance of none").

Participants interpreted the different precipitation forecasts similarly as there was no significantly statistical differences between scenarios, Friedman's χ^2 (2, N = 383) = 1.09, p = .580.

Effect of the format

For the three precipitation forecasts, participants were more likely to select the correct interpretation when they read the forecast in a numerical probability format with an explicit mention to the reference class (See Table 2). The format also seemed to affect the tendency to select a personal interpretation. Participants were twice as likely to propose their own interpretation when they received a numerical probability forecast. The format of presentation of the probability of a precipitation had a statistically significant effect on the interpretation of the forecast for the Rain and Hail vignettes, χ^2 (9, N = 384) = 19.66, p = .020, Cramer's V = .13 and χ^2 (9, N = 384) = 27.49, p = .001, Cramer's V = .27, respectively. In the Snow vignette, the effect of format was not statistically significant, χ^2 (9, N = 384) = 14.43, p = .108, Cramer's V = .11.

Similarly as in Study 1, we have conducted a Bayes factor analysis for the proportion of correct PoP interpretations for all three precipitation conditions (assuming a uniform distribution prior, Albert, 2009). The Bayes Factor analysis provided strong evidence supporting the hypothesis postulating the effect of the format on the PoP interpretation in two out of the three precipitation conditions. The Bayes factors indicated strong evidence for the effect of the format in the rain and in the hail scenario, respectively, BF01 = 0.25 and BF01 = 0.16 and anecdotal support for the null hypothesis (i.e., format does not affect interpretation) in the snow scenario, BF01 = 1.42.

Format preference

Most participants reported that they preferred the simple numerical probability format (see Table 3). Participants' second favourite was the numerical probability featuring an explanation, whereas the two verbal probability formats earned less than 10% of the preference. The type of precipitation forecasted did not have an effect on preference, Friedman's χ^2 (2, N = 383) = 1.09, p = .580. Two participants consistently preferred providing their own format preference. The two formats they described were similar to one of the options presented but with the information presented in a different order. One described a numerical probability ("chance of rain: 30%") and the second a dual format introducing the numerical probability first (" 30% chance, unlikely it will rain").

Table 3. Preferred probability of precipitation format.

	NP	NP + RefClass	VP	VP + NP	Other
Snow	54.4%	31.3%	3.9%	9.9%	0.5%
Rain	57.0%	30.5%	4.7%	7.3%	0.5%
Hail	50.8%	33.3	5.7%	9.4%	0.5%

The effect of the format was different between studies 1 and 2. In study 1, participants interpreted the weather forecast most frequently when provided with a verbal probability format and least often when provided with a numerical format with a reference class explanation. However, this trend was not statistically significant. In study 2, participants interpreted the weather forecast correctly most often when provided with a numerical probability and an explicit mention of the reference class. Findings of study 2 are in line with the expectations of Gigerenzer et al. (2005) who suggested that this intervention would help people identify correctly the reference class of probabilities of precipitation.
Results between studies 1 and 2 may differ because in study 1 the forecasted area was not defined, whereas it was defined in Study 2 (Wales, Surrey and Scotland for the low probability of rain, medium probability of snow and high probability of hail respectively). Together, findings of studies 1 and 2 indicate that the presence or absence of the target area in a forecast could interact with the format of the probability to determine the interpretation of the forecast.

DISCUSSION

The present research investigated the interpretation of probabilities of precipitation (PoPs) as a function of its format of presentation. Overall, less than half of the 1337 participants correctly understood a PoP. Taken together with previous research, this finding indicates that the interpretation of PoPs has not improved significantly in the last 30 years.

Overall, in the numerical probability conditions, 38% of the participants gave a correct interpretation, in 2011 and 2014, and 39% did so in 1980 (Murphy et al.) and on average around 37% in 2002 (Gigerenzer et al., 2005). This result clearly calls for a greater attention to examining how people understand probabilities instead of just focusing on people's subjective probability magnitude. Probability interpretation and probability perception are entangled and it is thus necessary to learn how people actually understand probabilities to comprehend the meaning of their subjective probability perceptions. For example, a person reporting a 30% chance of rain, may in fact be sure that it will rain – the uncertainty being when or where. The consequences of probability misinterpretations should be further investigated in the weather forecast area but also in other contexts where the misunderstanding of probabilities can have critical consequences, such as in legal and medical contexts.

We tested the suggestion of Gigerenzer et al.'s (2005) that the mention of the reference class would improve the interpretation of probabilities of precipitation. We used the reference class suggested by Gigerenzer et al.: "… in 3 out of 10 times when meteorologists make this prediction, there will be at least a trace of rain the next day" (pp. 629). Findings indicate that, under specific circumstances, weather forecasts featuring this explanation can indeed improve the correct understanding of precipitation forecasts. This was the case when the forecasted area was explicitly described and when participants were given the possibility of providing their own personal interpretation (Study 2), whereas it was not the case when the target area was not specified and in a forced choice setting (Study 1). Importantly, findings show that even with an explicit reference class, single event probabilities remained difficult to understand, with only 53% of the participants selecting the correct interpretation of a precipitation forecast featuring an explicit mention to the reference class. The explicit mention to the reference class in PoP, as suggested by Gigerenzer et al. (2005), should therefore be further scrutinised and possibly adapted before being used as a tool to enhance people's understanding of PoP.

Moreover, our results also show that the verbal probability forecast – preventing the polysemous ambiguity – did not provide a statistically significant better understanding than the traditional numerical probability forecast.

Taken together, the effect of formats on single event probability's interpretations indicate that the polysemy of percentage is not a driver of erroneous interpretations and that the reference class of single event probabilities might well be inherently hard to identify, whatever the format of presentation of the probability (e.g., numerical or verbal). To better understand why it is so hard for people to identify the reference class of PoPs, further research could focus on why the presence of the double framing of numerical probabilities improved this matter (Juslyn et al., 2009).

The measuring of the weather forecast interpretations by a presentation of three options followed that of Gigerenzer et al. (2005). In study 2, participants were given the opportunity to provide their personal interpretations of the forecast. However, this gave little insight into the reference class that people identify, as the personal interpretations mainly consisted in a reformulation of the weather forecast. This was also the case in previous research as in Murphy et al. (1980) who asked what a numerical probability of rain meant. Further, it is possible that people do not understand the interpretation question as intended by scientists. Indeed, when they gave their personal interpretations, participants did not seek to identify the reference class of the probabilities, but rather to clarify the nature of uncertainty (e.g., likelihood? probability?) or the alternative outcomes that might or might not occur (chance of rain and chance of no rain). Subsequent research should investigate the perceived reference class of probabilities of precipitation by asking participants more explicitly about the source of information that is used to form the forecast. Future investigations could also extend the number of interpretations provided as possible answers or could use a free response format, as recommended by Morss et al. (2008).

The pattern of the format preference found in Study 2 replicated the preference for numerical probabilities observed in the past, whereas verbal probabilities received few votes (Wallsten, Budescu, Zwick, & Kemp, 1993) and so confirmed the occurrence of this preference in a weather forecast context. Findings on format preference illuminate two important and novel facts. First, participants preferred the simple numerical probability format over the numerical format including an explicit mention to the reference class. This means that participants do not recognise this format as helping them understand the forecast better. Perhaps participants preferred the numerical probability format, because it was simpler and shorter than the one including reference class and it required less effort to process. Second, only a few participants expressed a preference for the dual (verbal/numerical) format. Considering that verbal probabilities fit better with how people think about uncertainty whilst nevertheless preferring to receive numerical probabilities, it has been suggested that the dual format could represent the best of both worlds – a verbal probability providing a ‘natural’ feeling of uncertainty and a numerical one providing a precise estimate of uncertainty (Budescu, et al., 2009; Budescu, et al., 2012; Budescu, Por, Broomel & Smithson, 2014). However, our data indicate that less than one in ten people wish to receive a probabilistic forecast in this format. It is interesting to consider that the format that was best understood was not the favourite one. One participant stated that a better format would be a dual numerical/verbal format instead of the verbal/numerical one, suggesting swapping the order of the two elements of the dual format (e.g., there is a 30% chance – it is unlikely).

At a more general level, the present manuscript, featuring non-significant effects (Study 1), represents an endeavor to decrease the publication bias observed in psychology where non-significant results are left "in the drawer" (Francis, 2012). By reporting both statistically significant and non-statistically significant findings, we contribute to the validity of meta-analyses and the estimation of overall size effects (van Assen, van Aert, Nuijten, & Wicherts, 2014). We strongly believe that null results may have a high informative value if researchers harness appropriate research methods and conduct relevant data analyses (e.g., well powered study and Bayes factor analysis).

Results indicate that specifying the reference class can improve the interpretation of probabilities of precipitation when the target area is specified. Findings indicate that using a verbal probability format (assumed to be non-polysemous) does not improve the interpretation of probabilities of precipitation.

REFERENCES

Ambury, S. (1982). On the economic value of probability of precipitation forecasts in Canada. *Journal of Applied Meteorology, 21*, 495-498.

American Meteorological Society (2008). Enhancing weather information with probability forecasts. An information statement of the American Meteorological Society. *Bulletin of the American Meteorological Society, 89.*

Budescu, D. V., Broomell, S. B., & Por, H.-H. (2009). Improving communication of uncertainty in the IPCC reports. *Psychological Science 20*, 299-308.

Budescu, D. V., Broomell, S. B., & Por, H.-H. (2012). Effective communication of uncertainty in the IPCC reports. *Climatic Change, 113,* 181-200. doi: 10.1007/s10584-011-0330-3

Budescu, D. V., Por, H.-H., Broomell, S. B. & Smithson, M. (2014). The interpretation of IPCC probabilistic statements around the world. *Nature Climate Change.* Advanced online publication. doi: 10.1038/NCLIMATE2194.

Buhrmester, M., Kwang, T. & Gosling, S. D. (2011). Amazon's Mechanical Turk. *Perspectives on Psychological Science*, *6,* 3-5. doi: 10.1177/1745691610393980

Demuth, J. L., Morrow, B. H., & Lazo, J. K. (2009). Weather forecast uncertainty information: An exploratory study with broadcast meteorologists. *Bulletin of the American Meteorological Society, 90*, 1614-1618. doi: 10.1175/2009BAMS2787.1

Gigerenzer, G. (2002). *Reckoning with risk: Learning to live with uncertainty.* London: Penguin Books.

Gigerenzer, G., Hertwig, R., van den Broek, E., Fasolo, B. & Katsikopoulos, K. V. (2005). A “30% chance of rain tomorrow": How does the public understand probabilistic weather forecasts? *Risk Analysis, 25*, 623-629.

Joslyn, S., Nadav-Greenberg, L. & Nichols, R. M. (2009). Understanding probability of precipitation: Assessment and enhancement of end-user understanding. *Bulletin of the American Meteorological Society, 90*, 185-193. doi: 10.1175/2008BAMS2509.1

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 Behavior and Human Decision Processes, 118*, 72-81. doi: 10.1016/j.obhdp.2012.01.002

Morss, R. E., Demuth, J. L., & Lazo, J. K. (2008). Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Weather and Forecasting*, *23*, 974-991.

Morss, R. E., Lazo, J. K. & Demuth, J. L. (2010). Examining the use of weather forecasts in decision scenarios: results from a US survey with implications for uncertainty communication. *Meteorological applications, 17*, 149 - 162. doi: 10.1002/met.196

Murphy, A. H., Lichtenstein, S., Fischhoff, B. & Winkler, R. L. (1980). Misinterpretations of precipitation probability forecasts. *Bulletin of the American Meteorological Society, 61*, 695-701. doi: 10.1175/2008WAF2007088.1

National Research Council. (2006). *Completing the forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts.* National Academies Press. Retrieved November 2, 2011, from http://www.nap.edu/openbook.php?record_id=11699&page=R1.

National Research Council. (2003). *Communicating Uncertainties in Weather and Climate Information: A Workshop Summary.* National Academies Press.

National Weather Service (2009). Explaining "Probability of Precipitation". Retrieved November 2, 2011, from http://www.srh.noaa.gov/ffc/?n=pop

Olofsson, A. & Rashid, S. (2011). The white (male) effect and risk perception: Can equality make a difference? *Risk Analysis, 31, 1016-1032*. doi: 10.1111/j.15396924.2010.01566.x

Paolacci, G., Chandler, J. & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411-419.

Wallsten, T. S., Budescu, D. V., Zwick, R. & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, *31*, 135-138.

Renooij, S. & Witteman, C. (1999). Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning, 22*, 169-194.

Riege, A. C. & Teigen, K. H. (2013). Additivity neglect in probability estimates: Effects of numeracy and response format. *Organizational Behavior and Human Decision Processes*. doi: 10.1016/j.obhdp.2012.11.004

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin Review,16*, 225-37.

Rodd, J. Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: semantic competition in lexical access. *Journal of Memory and Language, 46*, 245–266. doi:10.1006/jmla.2001.2810

Rogell, R. H. (1972). Weather terminology and the general public. *Weatherwise, 25*, 126 - 132. doi: 10.1080/00431672.1972.9931588

Sink, S. A. (1995). Determining the public's understanding of precipitation forecasts; results of a survey. *National Weather Digest, 19*, 9-15.

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. doi: 10.1016/j.actpsy.2012.09.004

Smerecnik, C. M. R., Mesters, I., Kessels, L. T. E., Ruiter, R. A. C., Vries, N. K. D. & Vries, H. D. (2010). Understanding the positive effects of graphical risk information on comprehension: Measuring attention directed to written, tabular, and graphical risk information. *Risk Analysis*, *30*, 1387–1398. doi: 10.1111/j.1539-6924.2010.01435.x

Thomas, R. P., Dougherty, M. R., Amber M. Sprenger & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review, 115*, 155–185. doi: 10.1037/0033-295X.115.1.155.

van Assen, M. A. L. M., van Aert, R. C. M., Nuijten, M. B. & Wicherts, J. M. (2014). Why Publishing Everything Is More Effective than Selective Publishing of Statistically Significant Results. *PLoS ONE* , *9*, online advanced publication. doi: 10.1371/journal.pone.0084896

Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HLJ. Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal Of Personality And Social Psychology* 2011;100(3):426-32.Wallsten, T. S. & Budescu, D. V. (1995). A review of human linguistic probability processing : General principles and empirical evidence. *The Knowledge Engineering Review, 10*, 43-62. doi: 10.1017/S0269888900007256.

Weber, E., Böckenholt, U., Hilton, D. J. & Wallace, B. (1993). Determinants of diagnostic hypothesis generation Effects of information base rates and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1151-1164.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t-tests. *Perspectives on Psychological Science, 6*, 291-298.

Witteman, C., Renooij, S. & Koele, P. (2007). Medicine in words and numbers: A cross-sectional survey comparing probability assessment scales. *BMC Medical Informatics and Decision Making, 7*, 13-20. doi: 10.1186/1472-6947-7-13

Winkler, R., L. (1990). Comment: Representing and communicating uncertainty. *Statistical Science, 5*, 23-26.