Vague Visualizations to Reduce Quantification Bias in Shared Medical Decision Making

Michela Assale, Silvia Bordogna and Federico Cabitza[a]
Universita degli Studi di Milano-Bicocca, Viale Sarca 336, 20126, Milano, Italy

Keywords: Uncertainty Visualization, Medical Decision Making, Visual Metaphor.

Abstract: This paper aims to contribute to the research focusing on how to render properly uncertainty in decision making, especially in regard to classification (like in medical diagnosis) or risk prediction (like in medical prognosis). Information visualizations leverage perception to convey information on data in ways that make their interpretation easier. Unfortunately, many visualizations omit uncertainty or communicate it less than effectively. We devised a novel way, which we call *vague visualization*, to render uncertainty without converting it in any numerical or symbolic form, and tested the usability and task fitness of these alternative solutions in a user study that involved a panel of lay people (as proxies of potential patients). In so doing, we aimed to understand whether our solutions facilitate (or at least do not hinder) communication and understanding of probabilistic estimates in a medical context, and if one solution is more effective than the others. We observed that three different *vague visualizations* convey the right sense of risk with respect to chance (50%) of percentage shown, and inspire an interpretation of the magnitude of the percentages that replicates the typical response of decision making under uncertainty condition. We then claim that these methods are effective because they allow for data interpretations that are uncertain (vague), and yet correct and compatible with appropriate decisions and actions.

1 INTRODUCTION

This paper aims to contribute to the research focusing on how to render and present uncertainty to the decision makers in proper ways, especially in regard to classification (like in medical diagnosis) or risk prediction (like in medical prognosis).

One could rightly observe that defining a "proper way" here is not a trivial task. If decisions can be assessed as being either right or wrong (even *a posteriori*), then a proper way is the way that helps decision makers increase or maintain an acceptable accuracy. However, there many decisions that cannot be traced back to a matter of "right or wrong". For instance, in medicine (that is our reference domain), when uncertainty regards an estimate of the probability of reaching a specific condition after a treatment, one could compare different ways to present the odds and risk of a set of options and focus on their role in suggesting the option that helps get the best health outcome, or helps predict the effects of the intended course of action.

This approach can be denoted as *consequentialist* (as it focuses on the consequences of a decision) and purposely disregards whether the decision makers have "understood" the extent estimate of probability on risk. Although this is probably the soundest approach, it is difficult to pursue. Similarly to other researches (e.g., (Kosara et al., 2001; Finger and Bisantz, 2002; MacEachren et al., 2005; Kinkeldey et al., 2014)), we also equate the "proper way" in terms of the accuracy of those who reads the visualization to reconstruct the quantitative estimates behind it: the more effective is the visualization (of uncertain estimates), the easier to get the uncertain quantities therein represented (Mackinlay, 1986).

However, we would argue that effectiveness should not be optimized, but rather the *embodied understanding* of the uncertainty involved. Failing to achieve this understanding could be the main reason why "many visualization authors choose not to visualize uncertainty" (Hullman, 2019). In this paper we investigate purposely uneffective methods to convey a perception of uncertainty that not necessarily must be translated into quantitative numbers.

This contributes to the research on the role

[a] https://orcid.org/0000-0002-4065-3415

Assale, M., Bordogna, S. and Cabitza, F.
Vague Visualizations to Reduce Quantification Bias in Shared Medical Decision Making.
DOI: 10.5220/0008969802090216
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 3: IVAPP*, pages 209-216
ISBN: 978-989-758-402-2; ISSN: 2184-4321

of quantification in modern society (Porter, 1996), where it is generally critically reflected on the ways in which the hard sciences tend to *quantify* uncertainty, for instance in terms of probabilities or confidence scores, i.e., as real numbers between 0 and 1 (or, equivalently, in terms of percentages). This way to present uncertainty is appealing, especially for its role in making decision makers get an intellectual, abstract, symbolic comprehension of the probabilities involved, but it can also lead to what is called *quantification bias*, to denote both a sort of "overestimation of the importance of quantification in defining the concept of risk" (Boholm, 2019) and an over-reliance on quantified thresholds to make a decision. For instance, if a treatment is associated with a probability of positive outcome of .78, it is considered more likely to be effective than an alternative treatment that is associated with a probability of .75 even if the margin of error is 5%; another common case of quantification bias occurs in statistical analysis, whenever only findings associated to an observed confidence level (or P value) lower than .05 are considered "significant"[1]

However, it is a truism that many aspects in medical observations, which can lead to probability estimates, *cannot* be quantified: quantifying those aspects is convenient and sometimes acts as an effective approximation for any practical aim, but it is still a recognized fallacy in medical reasoning and decision making (O'Mahony, 2017): render signs or symptoms in terms of numbers on ordinal scales, or clear-cut categories, does not make them more objective or free from noise, error and uncertainty.

According to this approach to uncertainty, data visualizations usually render the quantities related to probabilistic estimates in visual forms that allow their users to derive numerical values quite straightforwardly; and purposely so: for instance, in terms of position along graduated scales; or segments of specific length in a Cartesian plane; or, less frequently, as angles, areas and color shading. In fact, a oft-cited paper (Cleveland and McGill, 1984) ranks these different methods according to the accuracy of the comparisons that they enable, suggesting to avoid the latter encodings (like areas and shading) as *too vague* for the human perceptual systems.

In this study we purposely pursue vagueness as a way to provide decision makers with a less cognitive and more immediate, concrete feeling of the uncertainty that affects a specific condition, or prospect. To this aim, we devised some alternative *vague visualizations* (VV). A VV is *a pictorial image to which a visual effect is applied proportionally to the amount of uncertainty that we want to convey with the VV* (see Figure 2 for some examples).

We consider VV as an "extreme" way to render uncertainty, to some extent even more extreme and difficult to decode than areas and color shading; their purpose is to represent uncertainty by means of a *perceptual analogous* that could *hinder*, rather than facilitate, the comprehension of the underlying numerical estimation, as this is how uncertainty works!

In this paper we report about the first user study on the use of VV. In particular, we simulated the use of VV in *shared decision making* (Barry and Edgman-Levitan, 2012), that is a situation in which both doctors and patients have to consider the odds of a series of treatments to understand which one to undertake.

This study is a preliminary step before we can suggest the adoption of VV in the real world and their integration in computational decision support systems. We wanted to test their capability to challenge the users without mislead them, and assess the extent our different methods could trigger the right interpretation of the underlying probabilities. In particular, in this user study we considered two related research questions, which we will consider in Section 3: in short whether VV allow to ascertain relative differences between quantities, and their absolute magnitude, at least coarsely. We achieved significant findings on these aspects, which we will discuss in Section 5.

2 RELATED WORK

2.1 Visualization of Uncertainty

Visualization can be an important means to help people understand the complex concept of uncertainty. Although a wide variety of visualizations have been created to communicate uncertainty, there is limited empirical and widespread evidence of how alternative formats affect human understanding and decision making and why one solution is better than anothers. There are several reasons why it is not yet well established which are the best formats to use, but both hue and color have been observed to be the preferred channel to convey a sense of uncertainty in readers (Hullman et al., 2018).

A comparative study (MacEachren et al., 2012) has assessed the best visual variables to communicate uncertainty through a system of symbols involving

[1] In more sophisticated settings, also the confidence intervals can be computed and presented, but this does not change the main evaluation process: if the lower band of an interval estimate is higher than the upper band of the other one, then the two quantities are considered *different* to any practical aim.

color value, color hue, size, color saturation, location, orientation, grain, arrangement, shape, fuzziness and transparency (Figure 1). Among these means, fuzziness and location were associated with the best results in terms of intuitiveness, followed by color value, arrangement, size and transparency. On the contrary, saturation, which is commonly considered to be related to uncertainty, was ranked low. MacEachren and

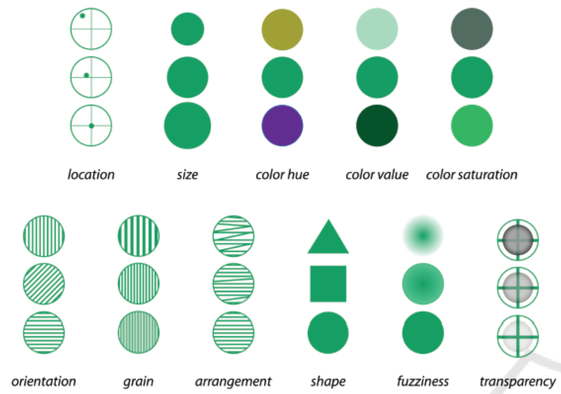

Figure 1: Visual variables taken from (MacEachren et al., 2012).

colleagues confirmed that abstract sign-vehicles can lead to quicker, though possibly less accurate, judgments. An important concept to consider is the role of the *visual metaphor*, by which fog and blur are metaphors for the lack of clear vision and therefore could be interpreted more quickly as signs of uncertainty (Kinkeldey et al., 2014). Resolution variation was tested in a study (Finger and Bisantz, 2002) in which the authors use resolution to represent uncertainty (with coarser resolution associated with higher uncertainty) and found that users can adequately understand the message.

Some researchers (Slocum et al., 2003) evaluated the effectiveness of intrinsic vs. extrinsic visualization techniques. They compared colour coding, transparency and line glyphs. In their user study, they found that those with a scientific background preferred glyphs, while less experienced users preferred color coding and transparency. This confirms that the most appropriate technique also depends on the purpose and capabilities of the target audience.

3 METHOD

In this section, we illustrate the empirical study that we aimed at the evaluation of the use of *Vague Visualizations* (VVs) to understand probability estimates and risk scores. In particular, we assessed the *representational effectiveness* of three image effects: *blur*, *transparency* and *noise* in rendering a probability value (in what follows we use this expression as equivalent to risk score). With *representational effectiveness* we mean the extent a solution allows for accurate answers, that is it does not mislead the user in the interpretation of a probability value. In fact, as we wrote in Section 1, the proposal of VVs in medical interfaces is aimed at reducing the users' over-reliance on quantitative representations and having them avoid the quantification fallacy and bias. However, VV must not hinder comprehension to the point of misleading interpreters of medical facts, like, e.g., if a dichotomous outcome is more likely than mere chance, or the odds for recovery are more than 1 (i.e., in both cases, above the 50% probability threshold).

Thus, we can formulate the following two research questions that we will address in the rest of the study.

1. Do VVs support (or hinder) the comprehension of probability estimations?[2] We relate this question to an *effectiveness* analysis of the VVs involved.
2. In case the former question has a positive answer, is there any effect that is better (i.e., more efficient/more effective/less misleading) than the others? We relate this question to a comparative analysis among alternative VVs.

Finding a positive answer for the former question would allow to demonstrate the feasibility of our proposal. Finding a positive answer for the second question would give recommendations for the adoption of a specific effect in medical applications.

3.1 Context

The domain of this application is the medical one, in particular the interpretation of medical results coming from decision support systems and probabilistic models. Many statistical models can yield a probability score associated with a given classification or the likelihood of a certain event to happen. Instead of representing the uncertainty related to this confidence or estimated likelihood with a parameter and related confidence interval we want to express probabilistic uncertainty by means of visual uncertainty (i.e., vagueness, indistinctness, fuzziness) to make physicians interpret the results not on the basis of a defined number but rather on a visual impression.

The quantification of a probability in a very specific number seems counterproductive in that it can make us overconfident about machine predictions,

[2] A related question would be: do VVs support the comprehension of the uncertainty that is intrinsic to probability estimation? We do not address this further question here.

blind to anomalies in data, or wary of intuition, which nevertheless may have human decision makers consider clues that the algorithm could not include (Cukier and Mayer-Schönberger, 2013).

Our experiment is based on the conjecture that visual metaphors make the interpretation of the message more intuitive, if not more accurate, and that intuition can be more important of accuracy in many instances of naturalistic decision making (Klein, 2008). In particular, we focused on three metaphors, or effects, with which to create a VV: blur, noise and transparency, defined as follows:

1. *Blur*: the effect that makes contours less defined creating an out of focus effect;
2. *Noise*: the random substitution of image pixels with blank pixel;
3. *Transparency*: the overlapping of the image with the background color;

3.2 User Test Design

To test our research questions, we developed a simple Web-based tool to create VVs [3]. This tool accepts any raster picture as input, all together with a probability value (in percent terms) as parameter; as output it yields the same picture affected by one of the above image effects proportionally to the percentage indicated: 100% was associated with the original image (no effect: 100% purity[4]); while 0% was associated with a highly distorting effect (see Figure 2) and maximum uncertainty. In doing so, we generated a set of 6 VVs, corresponding to different effect percentages, that is almost nil, first quartile, close to 50%, third quartile, and almost 100%, respectively: 10, 25, 40, 60, 75, 90.

We then developed an online questionnaire that could display the above VVs to a number of respondents, mostly bachelor students and acquaintances whom we invited during class and by email. Respondents participated voluntarily, with no incentives. In this questionnaire, participants were invited to associate each of six different VVs (for each image effect, 2 VVs randomly chosen from the above generated ones), in two tasks of increasing difficulty. For both tasks, we showed a three-VV reference set that indicated a 0%, 50% and 100% value, respectively.

In the first task, the respondents were invited to select whether the VV, with respect to the reference set, represented a value either *clearly higher*, *perhaps higher*, *perhaps lower* or *clearly lower* than 50% (i.e, the threshold for purely random decisions). We called this the *relative accuracy* (RA) task (in that it regards accuracy with respect to the random threshold).

In the second task, the respondents were invited to indicate the exact underlying probability value that the VV was expressing, by means of a slider ranging from 0 to 100. We denoted the second task as the *absolute accuracy* (AA) task.

Thus, each respondent had to perform two RA tasks and two AA tasks for each effect, for a total number of 12 tasks. In particular, for the RA tasks, we defined 2 measures of *accuracy* (or VV *effectiveness*): the rate of *adequately accurate* responses (*adequate accuracy*); and the rate of *approximately accurate* responses (*approximate accuracy*). The former accuracy was defined for the different percent values differently: the ratio between the total number of responses and the number of the respondents who answered *clearly higher*, *perhaps higher*, *perhaps lower* and *clearly lower* for, respectively, the 90%-, 60%-, 40%- and 10%-VV, and any type of *higher-than-50%* (and respectively, *lower*-than-50%) attribute for the 75%-VV (and 25%-VV). For these two latter VVs we did not define the *approximate accuracy*, which was defined for the 90%-, 60%-VV (and 40%- and 10%-VV) in terms of the number of *higher-than-50%* (and respectively, *lower*-than-50%) responses.

We expected two possible sources of bias that could affect our analysis: the order of the questions (i.e., order bias), and the value of percentage shown (i.e., sampling bias). In order to mitigate the former kind of bias, the online questionnaire was implemented to present the 3 different effects to the respondents in random order.

4 RESULTS

More than 100 respondents participated in the user study, in the age range 19-30. Since the sample encompassed only bachelor or master students aged between 19 and 30, we did not stratify the respondents on the basis of age or education level. We decided to remove from the sample the respondents who did not conclude the questionnaire, considering this evidence of low commitment in task execution. The size of the final sample, after cleaning it from partial an-

[3] This simple tool, and its code, are available at the following addresses: https://github.com/PinkLaura/pixel-e-percentuali and https://pinklaura.github.io/pixel-e-percentuali/, respectively.

[4] In the pilot test we observed as the respondents found more natural to cope with the concept of image purity rather than with the (complementary) concept of image fuzziness (as a proxy for uncertainty). Thus, we decided that the purer an image could be perceived, the lower the associated level of uncertainty, and the higher the probability or risk score that the user should try to guess by looking at the image.

Figure 2: Effects applied on image to render 25%, 50% and 75% risk. T, B and N are, respectively the effect of Transparency, Blur and Noise.

swers, was 88 users. The number of tasks concluded for each combination of image effect, task type and percentage value was between 25 and 32.

As said in the previous section, the goal of the study is to understand if VVs are a proper means to convey probability and risk. To this aim, we tested the accuracy of the answers through statistical tests. Accuracy was measured in various ways: for the RA tasks in terms of adequate and approximate accuracy (see Section 3): a good performance for a type of VV is defined according to the extent the corresponding accuracy rates are significantly above the 50% threshold of chance effect; when assessing the AA task, accuracy was determined in terms of the difference between the estimated value and the real values: a good performance was then associated with a hypothesis test failing to reject the null hypothesis that the above average difference is null.

We have decided to rely both on confidence intervals and non-parametric hypothesis tests (the Chi-Square and the Binomial tests): these latter tests were adopted either because the data were ordinal in nature, or as in case of the AA task, their variance was skewed. This choice makes our conclusions more conservative and therefore more reliable in terms of reproducibility in real-world settings but also more prone to fail to detect small effects, also due to the relatively small size of the sample.

4.1 Relative Effectiveness Tasks

For the RA task, the results are indicated in Figure 3 and Figure 4 for the approximate accuracy and adequately accuracy constructs defined in Section 3, respectively.

4.2 Effect Comparison

In this subsection we want to verify if there is any difference in accuracy comparing the three methods used to create VVs.

To this aim, we performed a chi-square test of goodness of fit to determine whether there was any better effect among those tested. We checked for any difference in each percentage level and then in the cumulative case. Thus, we observed that for all of the cases the distribution of errors for the three effects was not significantly different from the uniform one, and no significant difference was detected among the three effects, neither with respect to single percentage levels nor across all the considered effects.

The greatest difference among the three effects

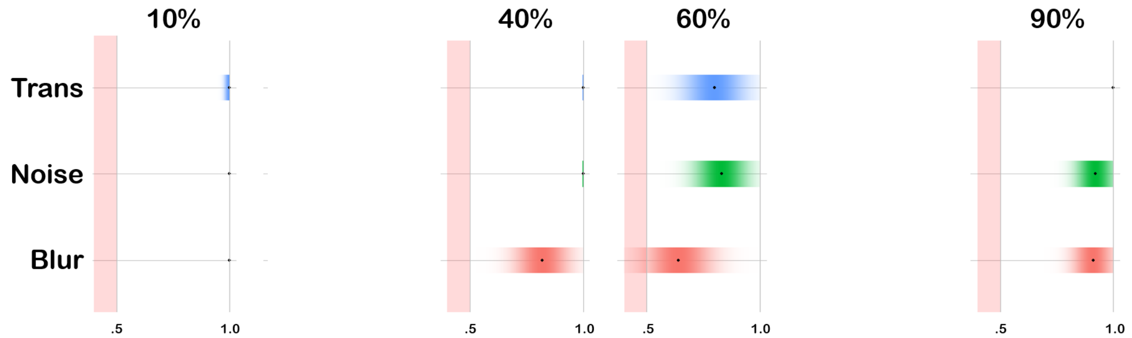

Figure 3: Success rate in regard to approximate accuracy for the probability levels where this construct is defined.

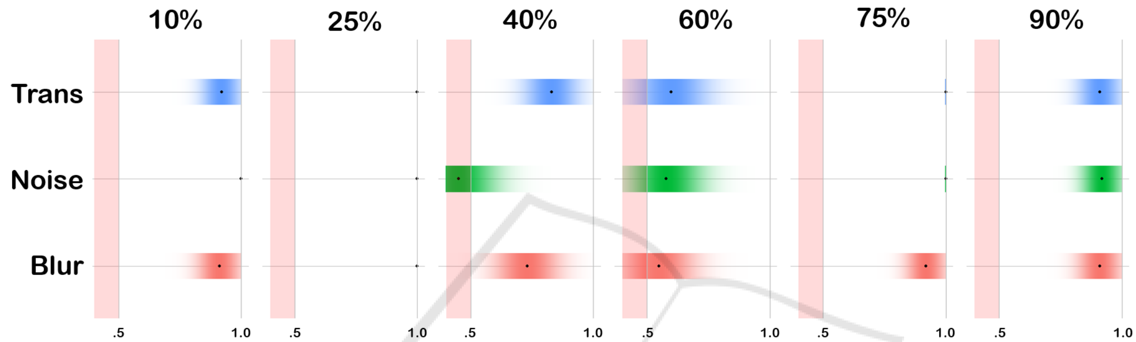

Figure 4: Success rate in regard to adequate accuracy for the probability levels where this construct is defined.

can be observed in the case of the 40% level (adequately correct) which corresponds to a p-value of .19 ($\chi^2(2, N = 68) = 3.31$). In this case, the difference between the Noise effect and the Transparency effect is statistically significant (p=.034), being this latter one significantly more effective ($\chi^2(1, N = 32) = 4.5$). Instead, the case of the smallest difference among the methods occurred in the case of the 75% level (adequately correct), with a corresponding p-value equal to .99 ($\chi^2(2, N = 68) = 0$).

4.3 Absolute Accuracy Tasks

In this section we explore if the effectiveness of the effects varies with the level of percentage rendered as VV. Through exploratory analysis it seems that small percentage values were overestimated, while the opposite is observed for high values. The effect is detectable, at descriptive level, for each of the three methods.

To test this idea we used a non-parametric test, the Mann Whitney U, to verify whether the true value of median of errors is lower/greater than 0. We obtained a p-value equal $3.12 * 10^{-11}$ for the VVs at risk level 10%: we can then reject the null hypothesis that median of the difference between answers and real value is zero and accept the alternative one. Considering that the confidence interval is between +8.00 and +12.49, we can also say that the inflation of the estimate is statistically significant. With the same test we have evaluated risk level 25%, p-value is $3.48 * 10^{-8}$ and even in this case we reject the null hypothesis. The estimation coming from this method does not have a distribution with median equal 0. With a confidence level of 95% we can say that conveying a risk/probability level through blur, noise or transparency effects, have users overestimate the true value. In 95% of the responses, the real median value is included in the related confidence interval, that, in this case, is between +5.00 and +10.00. P-value at level 40% is instead equal to 0.9. As suggested by the descriptive analysis, in this case it is reasonable to believe that the distribution of the answers has a median in 0. Once crossed the 50% the effect reverses: in fact we notice an underestimation of the real value; p-value at level 60% is equal to 0.01, and, for 75% and 90%, percentage shown p-value is near 0.

Results are reported in Figure 5. We can see that some confidence intervals (indicated by the boxes' notches) overlap, in particular at 60% the median could be very similar to the one at 75%, but it could also be confused with 40%. In the same way 25% and 10% are not clearly different.

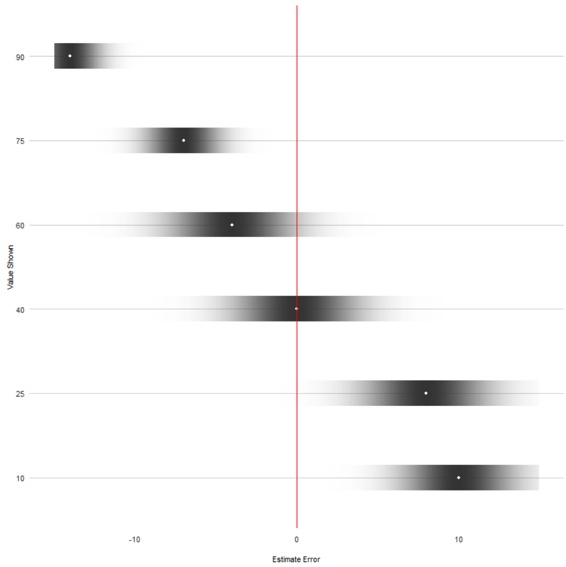

Figure 5: CI of median of distribution of error at confidence level 95%, obtained through Mann-Whitney test. Low percentage values (e.g. 10%) are overestimated while higher values (e.g. 90%) are underestimated.

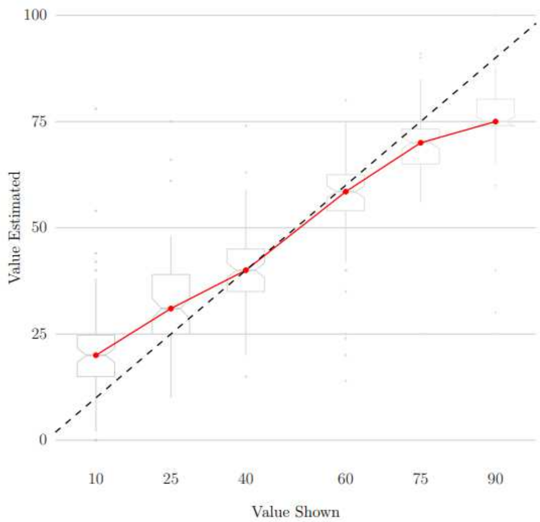

Figure 6: For each of the six percentage values shown, the box-plot of the answers is given. Combining the medians (red dots) we can draw a s-shaped line. The dashed line represents the ideal case in which the value estimated is exactly the same as the value shown.

5 DISCUSSION

To summarize the results obtained through the two tasks of the user test, we can conclude that:

- The blur, noise and transparency effects are effective in conveying the idea whether probability is different from chance and risk is lower or greater than chance (50%), with error rate almost nil.
- There is an overestimation/underestimation at the boundaries of the value range, which has had an impact on the error rate of the AA task (i.e., the task where users had to guess the accurate risk/probability level).

Depending on the purpose of the visualization, the proposed methods can be of varying effectiveness. If the goal is to convey an idea of the relationship between different estimates (e.g., with respect to the 0%, 50% or 100% levels) the VV method was proved effective; otherwise, if the intention is to communicate probability estimates accurately, a bias similar to regression to the mean will distort the perception of the original percentage.

The distortion in perception of proportions is not a new concept in decision making under uncertainty condition. A similar s-shaped line has already been observed (Tversky and Kahneman, 1992) during the studies to develop prospect theory, which describes how individuals assess in an asymmetric manner their loss and gain perspectives.

Moreover, we could not detect significant differences between the three methods; however, the small size of the sample prevented us to verify if, at a specific percentage levels, estimates are different. For example, the overestimation of lower risk values could be more or less emphasised using one method versus another, but only a greater sample could allow to get significant results, if any.

6 CONCLUSIONS

This work aims to contribute to the research field focusing on novel and effective visualization methods to communicate the uncertainty underlying a numerical and probabilistic estimation of risk, with a particular interest in the domain of doctor-patient communication.

We performed an agile review of the state of the art that allowed us to understand why communicating uncertainty is a crucial step in the process of extracting information from data to make better decisions in real-life situations.

The main aim of the study was to verify whether avoiding to rely on objective percentages or numbers to render uncertainty and, instead, using a "paradoxical" visual metaphor (in particular the blur, transparency and noise effect on pictorial images) can allow to communicate communication effectively. In particular, we tried to verify if *vague visualizations are a valid way to convey probability*, and if *there is any vague visualization method that is better among* *the proposed ones*.

All the three effects we employed to build a vague visualization convey the idea whether risks are greater or lower than chance of the percentages effectively (i.e., a percentage lower than 50% is correctly perceived so, as well as those greater than 50%). We also observed that vague visualizations are a valid means to communicate intermediate values, while we have observed a regression to mean when extreme values are shown. Finally we have not found any method to be better than the others.

We will soon undertake further experiments in controlled, as well in real-world settings, to see whether the decisions made by physicians are different when they are supported by a risk prediction that is rendered in terms of clear-cut quantities, or conveyed through a vague visualization and if they are equally satisfied of their decision support.

ACKNOWLEDGEMENTS

The authors are grateful to Laura Nesossi who, during her bachelor project work, developed the tool used to create the VVs used in this study and for her bachelor thesis helped the authors collect the responses of the online questionnaire.

REFERENCES

Barry, M. J. and Edgman-Levitan, S. (2012). Shared decision making – the pinnacle of patient-centered care. *New England Journal of Medicine*, 366(9):780–781.

Boholm, M. (2019). Risk and quantification: A linguistic study. *Risk Analysis*.

Cleveland, W. S. and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554.

Cukier, K. and Mayer-Schönberger, V. (2013). The dictatorship of data. *MIT Technology Review*.

Finger, R. and Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science*, 3(1):1–25.

Hullman, J. (2019). Why authors don't visualize uncertainty. *IEEE transactions on visualization and computer graphics*.

Hullman, J., Qiao, X., Correll, M., Kale, A., and Kay, M. (2018). In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE transactions on visualization and computer graphics*, 25(1):903–913.

Kinkeldey, C., MacEachren, A. M., and Schiewe, J. (2014). How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386.

Klein, G. (2008). Naturalistic decision making. *Human factors*, 50(3):456–460.

Kosara, R., Miksch, S., Hauser, H., et al. (2001). Semantic depth of field. In *infovis*, volume 1, pages 97–104.

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., and Hetzler, E. (2005). Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160.

MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., and Gahegan, M. (2012). Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141.

O'Mahony, S. (2017). Medicine and the mcnamara fallacy. *The journal of the Royal College of Physicians of Edinburgh*, 47(3):281–287.

Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.

Slocum, T. A., Cliburn, D. C., Feddema, J. J., and Miller, J. R. (2003). Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*, 30(4):299–317.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.