https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

Understanding representations of uncertainty, an eye-tracking study part II: The effect of expertise

Louis Williams[1,4], Kelsey J. Mulder[2], Andrew Charlton-Perez[2], Matthew Lickiss[3], Alison Black[3], Rachel McCloy[4], Eugene McSorley[4], Joe Young[5]

[1]ICMA Centre, Henley Business School, University of Reading, Whiteknights, PO Box 242, Reading, RG6 6BA, United Kingdom.

[2]Department of Meteorology, Earley Gate, University of Reading, Whiteknights Road, PO Box 243, Reading, RG6 6BB, United Kingdom.

[3]Department of Typography & Graphic Communication, School of Arts, English and Communication Design, No. 2 Earley Gate, University of Reading, Whiteknights Road, PO Box 239, Reading RG6 6AU.

[4]School of Psychology and Clinical Language Sciences, Earley Gate, University of Reading, Whiteknights Road, PO Box 238, Reading, RG6 6AL, United Kingdom.

[5]Department of Atmospheric Sciences, University of Utah, 115, Salt Lake City, UT 84112, United States

Correspondence to: Louis Williams (louiswilliams@dynamicplanner.com)

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

Abstract. As the ability to make predictions of uncertainty information representing natural hazards increases, an important question for those designing and communicating hazard forecasts is how visualisations of uncertainty influence understanding amongst the intended, potentially varied, target audiences. End-users have a wide range of differing expertise and backgrounds, possibly influencing the decision-making process they undertake for a given forecast presentation. Our previous, linked study, examined how the presentation of uncertainty information influenced end-user decision making. Here, we shift the focus to examine the decisions and reactions of participants with differing expertise (Meteorology, Psychology and Graphic Communication students) when presented with varied hypothetical forecast representations (boxplot, fan plot or spaghetti plot with and without median lines), using the same eye-tracking methods and experiments. Participants made decisions about a fictional scenario involving the choices between ships of different sizes in the face of varying ice thickness forecasts. Eye-movements to the graph area and key, and how they changed over time (early, intermediate, and later viewing periods), were examined. More fixations (maintained gaze on one location) and time fixating was spent on the graph and key during early and intermediate periods of viewing, particularly for boxplots and fan plots. The inclusion of median lines led to less fixations being made to all graph types during early and intermediate viewing periods. No difference in eye movement behaviour was found due to expertise, however those with greater expertise were more accurate in their decisions, particularly during more difficult scenarios. Where scientific producers seek to draw users to the central estimate, an anchoring line can significantly reduce cognitive load leading both experts and non-experts to make more rational decisions. When asking users to consider extreme scenarios or uncertainty, different prior expertise can lead to significantly different cognitive load for processing information with an impact on ability to make appropriate decisions.

1. Introduction

The importance of understanding the most ideal approach for communicating uncertainty information, an established problem in geoscience communication, has been further highlighted by the current COVID-19 pandemic. As more detailed information is presented to and interpreted by more non-specialists, the decisions made as a result have a significant impact on health, society and the environment, so careful consideration of communication is essential. Within the environmental sciences, making forecasts of natural hazards useful to end-users depends critically on communicating in a concise and informative way. Particularly as end-users have a wide range of differing expertise, spanning a spectrum between geo-

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

physical scientists to those with no formal scientific training. Therefore, the way in which information is displayed is very important for avoiding misperceptions and ensuring appropriate steps are taken by end-users, especially when perceptions of natural hazards can differ between experts and non-experts (Fuchs et al., 2009; Goldberg & Helfman, 2010).

Part I of this study, which from here will be called "companion paper" (Mulder et al., forthcoming), shows that, for all groups, great care is needed in designing graphical representations of uncertain forecasts. This is especially so when attention needs to be given to critical information, and the presentation of the data makes this more difficult. In particular, well known anchoring effects associated with mean or median lines can draw attention away from extreme values for particular presentation types (Broad et al., 2007; Nadav-Greenberg et al. 2008; Mulder et al., 2020). The availability of easy-to-use tools that make the development of complex graphical representations of forecasts quick and cheap to produce, poses new challenges for the geo-scientists. Here, we compare the response of three different groups of end-users with different levels of scientific expertise to the same series of forecast presentations to explore how more and less complex presentations influence decision making and perception.

Expertise differences may be due to greater familiarity with the ways in which hazard information is made available. This enables experts to make more economically rational decisions and to interpret uncertainty information more effectively (Mulder et al., 2020). However, the role of expertise remains unclear with some studies showing no differences in decision-making tasks with both experts and non-experts able to process and use forecast information to make decisions, with the inclusion of uncertainty information found to be useful for both experts and non-experts (Nadav-Greenberg et al., 2008; Kirschenbaum et al., 2014; Wu et al., 2014). Furthermore, it is unclear whether presentation of uncertainty information in visual formats results in benefits over using verbal and numerical expressions. For instance, uncertainty presented as graphical representations may help with understanding and interpretation (Susac et al., 2017). Additionally, research is required to examine differences in expertise, particularly as deterministic construal errors can be made as observers are often unaware that uncertainty is being depicted within visualisations (Joslyn & Savelli, 2021). Inappropriate information that captures attention is also often relied on, which can distort judgements (Fundel et al., 2019).

Experts are better at directing attention (through eye movements) to the important information required for making a decision. For example, in judgments of flight failures, expert pilots were found to make faster and more correct decisions, making more eye movements to the cues related to failures than non-experts (Schriver et al, 2008). Kang and

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

Landry (2014) also found non-experts to improve after they were trained with the eye movement scan paths of experts; training led non-experts to make fewer errors (false alarms) on aircraft conflict detection tasks. However, there is little research examining eye movements when experts and non-experts are required to make decisions using graphical and numerical forecast information. It is not clear which aspects of forecast information are being examined and when, and equally which, are being ignored.

More generally, research has shown that when viewing images, more fixations are made to informative regions and areas of interest (Unema et al., 2005). The times at which these fixations are made has been found to vary depending on task, decision type and expertise. Antes (1974) found that early fixations, in the first few seconds of viewing pictures, were towards informative areas. Goldberg and Helfman (2010) also showed that important regions of interest were fixated early during observation of different graphs. Experts have been shown to identify and fixate informative aspects of visual information more quickly and more often than non-experts (Maturi & Sheridan 2020; Charness, Reingold, Pomplun, & Stampe, 2001; Kundel, Nodine, Krupinski, & Mello-Thoms, 2008). As well as informative parts of a scene or image, Shimojo et al. (2003) reported that the likelihood that fixation would be made to the item preferred, increased over time, particularly in the final second before selection (see also Glaholt & Reingold, 2009; Simion & Shimojo, 2006; Williams et al., 2018). These results show that informative and preferred areas of images are selectively fixated early on, more often and for longer. As viewing evolves, fixations start to reflect final choices and preferences. The temporal development of this is task-dependent and influenced by expertise.

In our companion paper, we specifically examined how uncertainty information influenced interpretations and viewing behaviour. Regardless of expertise, participants were found to fixate towards median lines and less so to extreme values, with the type of graph and respective keys further influencing gaze and judgements. Here, we explore eye movement behaviour to similar hypothetical scenarios but with particular interest on differences due to participant expertise/background, following the research discussed, of gaze to graph areas and keys over different time periods of the decision-making process. As in our companion paper, we examine gaze patterns when faced with the task of making decisions about a fictional scenario involving the choices between ships of different sizes in the face of varying ice thickness forecasts (30%,50%,70%), when presented in different formats (boxplot, fan plot or spaghetti plot, with and without median lines).

We use eye-tracking techniques and exploration of the accuracy of decision tasks across expertise to address the following questions:

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

1. Does the presence of a median line and expertise affect gaze over the course of the decision-making process?
2. Does expertise affect gaze to the key over the course of the decision-making process?
3. Does expertise affect accuracy of decisions?

2. Methodology

2.1 Participants

Sixty-five participants took part in this study: twenty-two meteorology students, twenty-two psychology students and twenty-one graphic communication students recruited from the University of Reading (38 females, 27 males). Participants were aged 18–32 (M= 21.2) and had completed 0–4 (M=1.0) years of their respective degrees. Meteorology students are considered to have more training in graph reading, scientific data use, and quantitative problem solving as part of their degree and in qualifying for the course, than students on other degree courses which have less of a focus in these areas. Within this study, meteorology students were therefore considered to have greater expertise compared to the psychology and graphic communication students, although psychology students are also likely to have statistical knowledge and experience reading graphs. The research team involved academics who taught on each of these subjects and therefore can substantiate these generalisations.

2.2 Procedure

Full methodological details are given in our companion paper, but to restate the core procedure: A hypothetical scenario of ice thickness forecast was provided to participants. In this paper we only examined the decision-task question where participants were asked to select which ship (small or large) to send across an icy strait 72 hours ahead of time using a 72-hour forecast of ice thickness (see our companion paper for further details on the hypothetical scenarios). Ice thickness forecasts were presented in seven different types: deterministic line, box plot, fan plot and spaghetti plot. Each representation was presented with or without a median line. Each of these graph types was shown to represent 30%, 50%, and 70% probability of ice thickness exceeding 1 meter. While performing this task, participants wore an Eye link II eye-tracker headset which recorded eye movements of the right eye as they completed the survey. Head movements were restrained, and the eye tracker was calibrated to ensure accurate eye movement recording.

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

2.3 Data analysis

Two interest areas were formed from a post hoc classification to address our research questions (graph area and key). Three viewing periods across trials were created (early, intermediate, late). The exact definition of early, intermediate, and late differed by type of graph due to each style evoking slightly different viewing periods. Viewing periods for each specific graph type were of equal bins divided across the average time to complete the question and therefore ranged between 5 to 6 seconds. In this study, we report number of fixations and total fixation duration.

In our companion paper, our analysis of gaze was across all experimental trials and all tasks. However, as we are concerned about the viewing period and want to avoid effects of learning, we examine gaze when participants were faced with each graph type for the first time. Repeated exposure to graph type and the demand to make the same judgement may influence gaze patterns as informative parts of the figures are located more swiftly. Therefore, six trials for each graph type for each participant were examined. We analysed the accuracy of responses to this question (making the safe and cost-effective choice of the two options) and gaze (number and total fixation duration).

2.4 Ethics

The University of Reading Ethics Board approved the study, and the study was conducted in accordance with the standards described in the 1964 Declaration of Helsinki. Participants provided written informed consent. The authors declare that there is no conflict of interest.

3. Results

Based on the results of our companion paper, we further explore the impact of the presence of a median line considering the viewing period, expertise and graph type. We then focus on fixation towards the keys including viewing period, expertise, graph type and the presence of a median line as variables. For both research questions a four-way mixed measures ANOVA was conducted including graph type, presence of a median line and viewing period as within-subject variables, and expertise as a between-subjects variable. Finally, we report the accuracy of responses for the ice ship decision task highlighting any differences due to expertise.

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

3.1 Does the presence of a median line and expertise affect gaze over the course of the decision-making process?

Our companion paper shows how the presence of a median line affects the location of participants fixations; eye movements were closer to the median line. Previous research by Mulder et al. (2020), further shows that the median line influences decisions independent of the type of graph observed. Here, we further examined how the median line influences eye movement behaviour when considering the viewing period from early to late stages, and different levels of expertise, as well as the graph type.

A main effect of presence of a median line was found for number of fixations and total fixation duration made to the graph area, $F(1, 62)= 6.403$, $MSE=32.747$, $p=0.014$, $\eta^2=0.094$; $F(1, 62)= 7.125$, $MSE=2386741.96$, $p=0.01$, $\eta^2=0.103$. More fixations were made, and more time was spent fixating on the graph area of the display when no median line was present (fixation count M=8.74; total duration M=2128.64) compared to when a median line was provided (fixation count M=7.89; total duration M=1887.47).

A main effect of graph type was also found for number of fixations and total fixation duration made to the graph area, $F(2, 124)= 15.098$, $MSE=26.406$, $p<0.001$, $\eta^2=0.196$; $F(2, 124)= 16.810$, $MSE=1635280.256$, $p<0.001$, $\eta^2=0.213$. Boxplots elicited more fixations, and more time was spent fixating on boxplots (fixation count M=9.07; total duration M=2222.21) and fan plots (fixation count M=8.71; total duration M=2091.04) compared to spaghetti plots (fixation count M=7.17; total duration M=1710.92).

There was also a main effect of the viewing period for number of fixations and total fixation duration made to the graph area, $F(2, 124)= 59.608$, $MSE=36.762$, $p<0.001$, $\eta^2=0.488$; $F(2, 124)= 57.417$, $MSE=2294640.505$, $p<0.001$, $\eta^2=0.481$. There was found to be a greater number of fixations with longer dwell times on the graph area during early (fixation count M=9.83; total duration M=2399.96) and intermediate (fixation count M=9.52; total duration M=2284.11) viewing periods compared to later periods (fixation count M=5.60; total duration M=1340.09).

There was no main effect of expertise on gaze behaviour measured by both fixation count and total duration; $F(1, 62)= 0.536$, $MSE=64.185$, $p=0.588$, $\eta^2=0.017$; $F(1, 62)= 1.770$, $MSE=3970562.258$, $p=0.179$, $\eta^2=0.054$, respectively.

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

As well as the main effects of median line, graph type and viewing period, there was an interaction between the median line and viewing period for total fixation duration, $F(2, 124)$= 3.598, *MSE*=1543871.74, p=0.03, η^2=0.055. Less time was spent fixating the graph area during the early and intermediate stages of viewing when a median line was present (Early total duration M= 2174.97; Intermediate total duration M= 2137.79) compared to when no median line was present (Early total duration M= 2624.96; Intermediate total duration M= 2430.43), p<0.001; p=0.05, respectively. However, no differences were found due to the presence (later total duration M= 1349.65) or absence (later total duration M= 1330.54) of a median line during the later stages, p=0.896. No other interactions were found to be significant. These findings support that the median line can reduce cognitive load; impacting the total fixation duration and number of fixations made on the graph area, particularly during early stages of the decision-making process, and adds to results from our companion paper that showed how fixation location was towards the median line when present, regardless of the type of graph.

3.2 Is gaze to the key influenced by expertise and the viewing period during the decision-making process?

Mulder et al. (2020) found that particularly non-experts were misinterpreting data presented in a boxplot and suggest that not referring to the boxplot key led to making such errors. Our companion paper examined eye movements to the graph keys and found that less fixation was made to the spaghetti plot and boxplot keys compared to the fan plot keys. Here, we examine fixation to the key over different periods of the decision-making process. As non-experts can particularly misinterpret data from boxplots, we consider differing levels of expertise.

A main effect of graph type was found for number of fixations and total fixation duration made to the key, $F(2, 124)$= 42.900, *MSE*=8.096, p<0.001, η^2=0.409; $F(2, 124)$= 42.396, *MSE*=574225.040, p<0.001, η^2=0.406. More fixations were made, and more time was spent fixating on fan plot keys (fixation count M=2.45; total duration M=626.79) compared to both boxplot (fixation count M=1.48; total duration M=387.75) and spaghetti plot keys (fixation count M=0.56; total duration M=127.13), and more fixations and time spent on boxplot compared to spaghetti plot keys.

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

There was a main effect of the viewing period on the number of fixations that were made to the key within the display, as well as the total amount of fixation, $F(2, 124)= 17.967$, $MSE=6.593$, $p<0.001$, $\eta^2=0.225$; $F(2, 124)= 21.003$, $MSE=416719.669$, $p<0.001$, $\eta^2=0.253$. More fixations and longer dwell time to the key occurred during the early (fixation count M=1.61; total duration M=407.15) and intermediate (fixation count M=1.99; total duration M=515.33) viewing periods compared to later periods (fixation count M=0.90; total duration M=219.20).

No main effect of the median line on gaze to the key, measured by both fixation count and total duration, was found; $F(1, 62)= 0.175$, $MSE=7.574$, $p=0.677$, $\eta^2=0.003$; $F(1, 62)= 0.061$, $MSE=543399.152$, $p=0.805$, $\eta^2=0.001$, respectively. Nor was there a main effect of expertise on fixation count and total fixation duration; $F(1, 62)= 0.251$, $MSE=10.191$, $p=0.779$, $\eta^2=0.008$; $F(1, 62)= 0.141$, $MSE=730099.249$, $p=0.869$, $\eta^2=0.005$, respectively.

An interaction between the graph type and viewing period for fixation count and total fixation duration was found, $F(4, 248) = 3.578$, $MSE=4.724$, $p=0.007$, $\eta^2=0.055$; $F(4, 248) = 4.260$, $MSE=330504.612$, $p=0.002$, $\eta^2=0.064$., respectively. More fixations were made, and more time was spent fixating the boxplot key during the early (fixation count M= 1.68; total duration M=423.76) and intermediate (fixation count M= 2.06; total duration M=577.11) stages of the viewing period compared to the later stage (fixation count M=0.71; total duration M=162.39 $p<0.005$. Similarly, more fixations were made, and more time was spent fixating the fan plot key during the early (fixation count M= 2.69; total duration M=695.64) and intermediate stages (fixation count M= 3.10; total duration M= 791.37) compared to the later stage (fixation count M=1.55; total duration M=393.37) $p<0.005$. However, no differences were found between viewing periods for spaghetti plots, $p>0.05$. The reason for less fixation being to spaghetti plot keys generally, and no differences overtime, could be due to the intuitiveness of this form of plot and the simplicity of the key.

3.3 Does expertise affect accuracy of decisions?

Mulder et al. (2020) found no significant difference in accuracy of decisions made between the graph types, just in the amount of uncertainty interpreted from them. Here, accuracy responses on the number of times participants correctly identified which ship would be most economically rational to send were measured considering expertise and probability of risk.

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

	Meteorology	Psychology	Graphic Communication
30% probability	74%	66.2%	75.5%
50% probability	87%	70.1%	72.1%
70% probability	95.4%	96.1%	94.6%

Table 1. presents accuracy results for all probabilities of risk for differing expertise. A small ship is the correct ship to send for a 30% risk of ice thickness and a large ship for 50% and 70% risk levels.

Overall, participants were accurate in their choice of ship (Meteorology= 85.5%; Psychology= 77.9%; Graphic communication = 80.7%); however, some differences were apparent due to expertise. A one-way ANOVA shows differences in accuracy when presented with 50% probability of risk, which is the most challenging task, $F(2,64)= 4.029$, $p=0.023$. Multiple comparisons show meteorology students to be significantly more accurate than psychology students in choosing the large ship during these scenarios, $p=0.035$, and more accurate than graphic communication students, although this difference is not significant, $p=0.08$. No differences between expertise were found for the 30% and 70% trials, $p>0.05$.

4. Discussion and Conclusions

As scientific information is increasingly being presented to non-specialists graphically, it is important to consider how this information is delivered. This approach to open science, less dependent on expert interpretation, is a natural development as general scientific literacy increases and is welcomed by both scientific producers and consumers. As this approach develops, it becomes much more important to have a clear understanding of the biases in interpretation that results from different forms of data presentation. While relevant to many fields of science, there is a particular need for this understanding in the environmental sciences as environmental hazards increase and change.

Prior research presents mixed results, with some authors suggesting that when making slight variations to graph representations that display uncertainty, decisions and interpretations differ (Correll & Gleicher, 2014; Tak et al., 2015), whilst others show that despite greater discrepancies in forecast representation, such as between graphic visualisations and written forms, there are no differences (Nadav-Greenberg & Joslyn, 2009). Furthermore, few studies explore how experts and non-experts interpret forecast

https://doi.org/10.5194/egusphere-2022-929
Preprint. Discussion started: 10 November 2022

information from different types of graphical forecast representations (Mulder et al., 2020). The current research examines these areas further by using eye-movement techniques considering expertise, and the viewing period during the decision-making process when observing a range of graph types.

More economically rational responses to the ship decision were made by meteorology students (greater level of expertise) during the most difficult scenarios. We found participants, regardless of expertise, to spend less time fixating the overall graph when a median line was presented, particularly during early and intermediate stages of viewing. This provides more evidence for the anchoring bias suggested in previous papers (Mulder et al., 2020), and in our companion paper. Participants focussed on the key for boxplots and fan plots more during early and intermediate stages compared to later stages. This provides evidence that early stages of viewing are more exploratory and towards informative areas (Antes, 1974; Goldberg & Helfman, 2010). However, considering the results and the differences found due to graph type, spaghetti plots appear to be simpler to interpret, potentially reducing cognitive load, corroborating the findings in Mulder et al. (2020) that the spaghetti plot helped users interpret extreme values.

Overall, this study, together with the analysis in our companion paper, demonstrates that there are many challenges when presenting natural hazard data to both experts and non-experts, the way that information is portrayed can impact interpretations and decisions. It is important to note that the graph area and key discussed here are specific to the particular tasks presented in this study and are used as indicators of the impact of expertise, graph type and the viewing period. Furthermore, course of study within higher education was used as a proxy for expertise, with meteorology students being regarded to have higher levels. However, future research would benefit from examining behaviour and decisions of academics and forecasters who would be considered as experts.

Responses to which ship participants opt for due to the risk of ice thickness (small or large) supports the importance of expertise as accuracy reduces dependent on the probability of ice thickness, with those with greater expertise being more accurate during more uncertain situations. However, expertise appears to have little impact on eye movement behaviour within our study. The results show how median lines can reduce cognitive load drawing users to the central estimate regardless of expertise. However, it is important to note that a median line reduces the perceived uncertainty in a graphic, even when explicitly presented (Mulder et al. 2020), so use of a median line should be used when the amount of uncertainty in the estimate is less critical to understand. Use of the key within graphical representations can also impact interpretations of data. For forecast providers this suggests that standard information design principles which seek to reduce visual noise in data presentation and draw the user to the critical parts can have major benefits for their ability to effectively communicate with both expert and non-expert end-users.

More broadly, taken together the two parts of the study suggest that incorporating eye-tracking and other techniques from cognitive science into the process of the design of forecast communication tools could be extremely fruitful. These techniques are now well-established with technology that makes them relatively cheap to set up and use. Graphical presentation of geo-scientific forecasts can happen with a range of breadth and longevity of communication in mind. While eye-tracking and related techniques would not be appropriate for all purposes, where graphics are being developed for routine and wide use, for example routine weather forecasts, this kind of approach would be a very valuable addition to end-user engagement. One obvious extension to the work in the two parts of this study is applying the same techniques to well-known and widely used geo-scientific forecast graphics.

5. Author contributions

Louis Williams: Conceptualization, Investigation, Formal analysis, Writing – original draft preparation
Kelsey Mulder: Writing – review & editing
Andrew Charlton-Perez: Funding acquisition, Writing – review & editing
Matthew Lickiss: Writing – review & editing
Alison Black: Funding acquisition, Writing – review & editing
Rachel McCloy: Funding acquisition, Writing – review & editing
Eugene McSorley: Conceptualization, Resources, Writing – review & editing
Joe Young: Funding acquisition

Acknowledgments. We thank our eye-tracking study participants. This research is funded by the Natural Environment Research Council (NERC) under the Probability, Uncertainty and Risk in the Environment (PURE) Programme (NE/J017221/1). Data created during the research reported in this article are openly available from the University of Reading Research Data Archive at http://dx.doi.org/10.17864/1947.110

The authors declare that they have no conflict of interest.

References

Antes, J. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, 103(1), 62–70. http://doi:10.1037/h0036799

Broad, K., A. Leiserowitz, J. Weinkle, and M. Steketee, 2007: Misinterpretations of the "cone of uncertainty" in florida during the 2004 hurricane season. *Bulletin of the American Meteorological Society, 88* (5), 651–668. https://doi.org/10.1175/BAMS-88-5-651

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, *29*(8), 1146–1152. https://doi.org/10.3758/BF03206384

Correll, M., & Gleicher, M. (2014). Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics*, *20*(12), 2142-2151. http://doi:10.1109/TVCG.2014.2346298

Fuchs, S., Spachinger, K., Dorner, W., Rochman, J., & Serrhini, K. (2009). Evaluating cartographic design in flood risk mapping. *Environmental Hazards*, *8*(1), 52-70. http://doi:10.3763/ehaz.2009.0007

Fundel, V. J., Fleischhut, N., Herzog, S. M., Göber, M., & Hagedorn, R. (2019). Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *Quarterly Journal of the Royal Meteorological Society*, *145*, 210-231. https://doi.org/10.1002/qj.3482

Glaholt, M. G., & Reingold, E. M. (2009). The time course of gaze bias in visual decision tasks. *Visual Cognition*, *17*(8), 1228-1243. http://dx.doi.org/10.1080/13506280802362962

Goldberg, J. H., & Helfman, J. I. (2010, April). Comparing information graphics: a critical look at eye tracking. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization* (pp. 71-78). ACM. http://doi:10.1145/2110192.2110203

Joslyn, S., & Savelli, S. (2020). Visualizing Uncertainty for Non-Expert End Users: The Challenge of the Deterministic Construal Error. *Frontiers in Computer Science*, *2*, 58. https://doi.org/10.3389/fcomp.2020.590232

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720814523066. https://doi.org/10.1177/0018720814523066

Kirschenbaum, S. S., Trafton, J. G., Schunn, C. D., & Trickett, S. B. (2014). Visualizing uncertainty: The impact on performance. *Human factors*, *56*(3), 509-520. https://doi.org/10.1177/0018720813498093

Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Mello-Thoms, C. (2008). Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic Radiology*, *15*(7), 881–886. https://doi.org/10.1016/j.acra.2008.01.023

Maturi, K.S., Sheridan, H. Expertise effects on attention and eye-movement control during visual search: Evidence from the domain of music reading. *Atten Percept Psychophys* 82**,** 2201–2208 (2020). https://doi.org/10.3758/s13414-020-01979-3;

Mulder, K. J., Lickiss, M., Black, A., Charlton-Perez, A., McCloy, R., and Young, J.,(2020) Designing environmental uncertainty information for experts and non-experts: Does data presentation affect users' decisions and interpretations?. *Meteorological Applications* 27 (1), e1821. https://doi.org/10.1002/met.1821

Nadav-Greenberg, L., Joslyn, S. L., & Taing, M. U. (2008). The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making*, *2*(1), 24-47. https://doi.org/10.1518/155534308X284354

Nadav-Greenberg, L., & Joslyn, S. L. (2009). Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making*, *3*(3), 209-227. https://doi.org/10.1518/155534309X474460

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors*, *50*(6), 864-878. https://doi.org/10.1518/001872008X374974

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature neuroscience*, *6*(12), 1317-1322. http:// doi:10.1038/nn1150

Simion, C., & Shimojo, S. (2006). Early interactions between orienting, visual sampling and decision making in facial preference. *Vision research*, *46*(20), 3331-3335. https://doi.org/10.1016/j.visres.2006.04.019

Susac, A., Bubic, A., Martinjak, P., Planinic, M., & Palmovic, M. (2017). Graphical representations of data improve student understanding of measurement and uncertainty: An eye-tracking study. *Physical Review Physics Education Research*, *13*(2), 020125. https://doi.org/10.1103/PhysRevPhysEducRes.13.020125

Tak, S., Toet, A., & Van Erp, J. (2015). Public understanding of visual representations of uncertainty in temperature forecasts. *Journal of cognitive engineering and decision making*, *9*(3), 241-262. https://doi.org/10.1177/1555343415591275

Unema, P. J., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual cognition*, *12*(3), 473-494. http://dx.doi.org/10.1080/13506280444000409

Williams, L., McSorley, E., & McCloy, R. (2018). The relationship between aesthetic and drawing preferences. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(3), 259. https://doi.org/10.1037/aca0000188

Wu, H. C., Lindell, M. K., Prater, C. S., & Samuelson, C. D. (2014). Effects of track and threat information on judgments of hurricane strike probability. *Risk analysis*, *34*(6), 1025-1039. https://doi.org/10.1111/risa.12128