

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

3 Louis Williams<sup>1,5</sup>, Kelsey J. Mulder<sup>2,3</sup>, Andrew Charlton-Perez<sup>2</sup>, Matthew Lickiss<sup>4</sup>, Alison  
4 Black<sup>4</sup>, Rachel McCloy<sup>5</sup>, Eugene McSorley<sup>5</sup>, Joe Young<sup>6</sup>

5

6 <sup>1</sup>ICMA Centre, Henley Business School, University of Reading, Whiteknights, PO Box 242,  
7 Reading, RG6 6BA, United Kingdom.

8 <sup>2</sup>Department of Meteorology, Earley Gate, University of Reading, Whiteknights Road, PO  
9 Box 243, Reading, RG6 6BB, United Kingdom.

10 <sup>3</sup>Liberty Specialty Markets, 20 Fenchurch Street, London EC3M 3AW, UK

11 <sup>4</sup>Department of Typography & Graphic Communication, School of Arts, English and  
12 Communication Design, No. 2 Earley Gate, University of Reading, Whiteknights Road, PO  
13 Box 239, Reading RG6 6AU.

14 <sup>5</sup>School of Psychology and Clinical Language Sciences, Earley Gate, University of Reading,  
15 Whiteknights Road, PO Box 238, Reading, RG6 6AL, United Kingdom.

16 <sup>6</sup>Department of Atmospheric Sciences, University of Utah, 115, Salt Lake City, UT 84112,  
17 United States

18

19 Correspondence to: Louis Williams ([louiswilliams@dynamicplanner.com](mailto:louiswilliams@dynamicplanner.com))

20

21 **Abstract.** As the ability to make predictions of uncertainty information representing natural  
22 hazards increases, an important question for those designing and communicating hazard  
23 forecasts is how visualisations of uncertainty influence understanding amongst the intended,  
24 potentially varied, target audiences. End-users have a wide range of differing expertise and  
25 backgrounds, possibly influencing the decision-making process they undertake for a given  
26 forecast presentation. Our previous, linked study (Mulder et al, 2023), examined how the  
27 presentation of uncertainty information influenced end-user decision making. Here, we shift  
28 the focus to examine the decisions and reactions of participants with differing expertise  
29 (Meteorology, Psychology and Graphic Communication students) when presented with  
30 varied hypothetical forecast representations (boxplot, fan plot or spaghetti plot with and  
31 without median lines), using the same eye-tracking methods and experiments. Participants  
32 made decisions about a fictional scenario involving the choices between ships of different  
33 sizes in the face of varying ice thickness forecasts. Eye-movements to the graph area and  
34 key, and how they changed over time (early, intermediate, and later viewing periods), were  
35 examined. More fixations (maintained gaze on one location) and time fixating was spent on  
36 the graph and key during early and intermediate periods of viewing, particularly for boxplots  
37 and fan plots. The inclusion of median lines led to less fixations being made to all graph  
38 types during early and intermediate viewing periods. No difference in eye movement  
39 behaviour was found due to expertise, however those with greater expertise were more  
40 accurate in their decisions, particularly during more difficult scenarios. Where scientific  
41 producers seek to draw users to the central estimate, an anchoring line can significantly  
42 reduce cognitive load leading both experts and non-experts to make more rational decisions.  
43 When asking users to consider extreme scenarios or uncertainty, different prior expertise  
44 can lead to significantly different cognitive load for processing information with an impact on  
45 ability to make appropriate decisions.

46

## 47 **1. Introduction**

48 The importance of understanding the most ideal approach for communicating uncertainty  
49 information is a common across multiple domains in everyday life and across a range of  
50 sciences (Fischhoff, 2012) and is an established problem in geoscience communication  
51 (Stephens et al, 2012). This importance has been highlighted by the current COVID-19  
52 pandemic during which there has been a sharp increase in the use of unfamiliar  
53 visualizations of uncertainty presented to the public in order to explain the basis of decisions  
54 made to justify the response being asked of them to adopt modified and new behaviours in  
55 order to mitigate transmission. As more unfamiliar and detailed information is presented to

56 and interpreted by non-specialists, the decisions made as a result have a significant impact  
57 on health, society and the environment, so careful consideration of communication is  
58 essential (Peters, 2008). It is clear that people have trouble gaining an appropriate  
59 understanding of uncertainty information and how best to use this in order to support optimal  
60 decisions (e.g., Tversky and Kahneman, 1974; Nadav-Greenberg and Joslyn, 2009;  
61 Roulston and Kaplan, 2009; Savelli and Joslyn, 2013). A great deal of research has been  
62 concerned with addressing the most appropriate way to communicate uncertainty to promote  
63 effective decision-making and understanding (Fischhoff, 2012; Milne et al., 2018). Deciding  
64 what uncertainty information should be included, what ought to be emphasized, and the  
65 manner in which it is best conveyed all have an important role to play (Bostrom et al., 2016;  
66 Broad et al, 2012; Morss et al., 2015; Padilla et al., 2015). Furthermore, there is a reluctance  
67 by authors, such as data scientists, journalists, designers and science communicators, to  
68 present visual representations of quantified uncertainty (Hullman 2019). There is a belief that  
69 it will overwhelm the audience and the main purpose of the data, invite criticism and  
70 scepticism, and that it may be erroneously interpreted as incompetence and a lack of  
71 confidence which will encourage a mistrust of the science (Fischhoff, 2012; Gistafson &  
72 Rice, 2019; Hullman, 2019). This research points to the lack of consistent recommendations  
73 and stresses the need for the form of communication being tailored to both the aims and  
74 desired outcomes of the communicator and the needs and abilities of the audience  
75 (Spiegelhalter et al., 2011; Lorenz et al., 2015; Harold et al., 2016; Petropoulos et al., 2022).

76 Visualizing uncertainty in geoscience forecasts needs to balance robustness, richness, and  
77 saliency (Stephens, et al. 2012). Recently, numerous examples of this have focussed on  
78 creative ways to achieve this (Lorenz et al., 2015; Harold et al., 2016; Petropoulos et al.,  
79 2022). Communication of uncertainty can take the forms of words, but this can lead to issues  
80 of ambiguity caused by the language used and the variation in user interpretation (Wallsten  
81 et al, 1986; Skubisz et al., 2009). However, there is clearly strength to this approach when it  
82 is needed. For example, taking a storyline approach has been shown to be a powerful  
83 technique for communicating risk when less focus is needed on probabilistic information and  
84 more emphasis is needed on plausible future events (Shepherd et al., 2018; Sillmann et al.,  
85 2021). To overcome issues of ambiguity of words, numbers are often used to present  
86 uncertainty as probabilities in the form of fractions (1/100), natural frequencies (1 in 100), or  
87 percentages (1%), but these forms can lead to ratio bias or denominator neglect (Morss et  
88 al., 2008; Kurz-Milcke et al., 2008; Reyna and Brainerd, 2008; Denes-Raj and Epstein, 1994;  
89 Garcia et al., 2010), and the most effective form to use to aid understanding can depend on  
90 the context (Gigerenzer & Hoffrage, 1995; Joslyn & Nichols, [2009](#)). Similarly presenting  
91 uncertainty graphically can take many forms which means they have the advantage of

92 flexibility of presentation, can be tailored for specific audiences, can help with differing levels  
93 of numeracy and can help people focus on the important gist of the information when using  
94 uncertainty to help reach a decision (Feldman-Stewart et al., 2007; Peters et al, 2007; Lipkus  
95 and Holland , 1999). As with the use of words, the choice of graphic to employ is dependent  
96 on the audience and intended message outcome (Spiegelhalter, 2017) and can lead to the  
97 overestimation of risk and negative consequences depending on the framing of the  
98 information (Vischers et al, et al, 2009). Pie charts are good for presenting proportions and  
99 part-to-whole comparisons and benefit from being intuitive and familiar to the public, but  
100 interpretation can sometimes be difficult (Nelson et al., 2009). Bar charts are useful for  
101 communicating magnitude and allowing comparisons (Lipkus, 2007) while line graphs are  
102 helpful in conveying trend information about the change in uncertainty over time. Icons can  
103 also be very useful, especially so for people with low numeracy and have been found to be  
104 effective when supplemented by a tree diagram (Galesic et al., 2009; Gigerenzer et al, 2007;  
105 Kurz-Milcke et al., 2008). These types of graphical communication can also include  
106 information about the range of uncertainty (such as a “cone of uncertainty”, Morss et al.,  
107 2016).

108 Previous research has shown that including uncertainty information can aid users to make  
109 more rational decisions (Nadav-Greenberg et al., 2008; Nadav-Greenberg and Joslyn, 2009;  
110 Roulston and Kaplan, 2009; Savelli and Joslyn, 2013 St John et al., 2000). One way in which  
111 this is achieved is by use of heuristics (Tversky and Kahneman, 1974). If selected wisely  
112 then these can help simplify probabilistic information to bolster and speed decisions promote  
113 optimal interpretation of data. However, poor selection can hinder and encourage suboptimal  
114 decisions (Mulder et al., 2020). For example providing an anchor value alongside data can  
115 help users interpret the data more efficiently by focussing them on that particular value (for  
116 example, focussing people on precipitation level on days like this as a start point to  
117 estimating rainfall) but if chosen poorly can encourage a more extreme and suboptimal  
118 interpretation (focussing on the maximum precipitation level on days like this would  
119 encourage higher estimates of rainfall). In terms of graphical visualization of uncertainty,  
120 providing a central line showing a likely hurricane track has been reported to distract users  
121 from possible hurricane tracks given by the cone of uncertainty. Equally, however, the cone  
122 of uncertainty has been sometimes misinterpreted as showing the extent of the storm (Broad  
123 et al., 2007). Beyond heuristics, other design choices have also been found to affect optimal  
124 and efficient decision-making (Speier, 2006; Kelton et al., 2010; Wickens et al., 2021).  
125 Different designs of boxplots and graphs showing the same information affect decisions and  
126 interpretations (Correll and Gleicher, 2014; Bosetti et al., 2017; Tak et al., 2013, 2015).  
127 Forecasting maximum values from graphs was found to depend on graph type (Mulder et al.,

128 2020). Giving tornado warnings with probabilistic information about where a tornado may  
129 strike increased response in those areas compared with deterministic information (Ash et al.,  
130 2014).

131 Part I of this study, which from here will be called “companion paper” (Mulder et al., 2023),  
132 shows that, for all groups, great care is needed in designing graphical representations of  
133 uncertain forecasts. This is especially so when attention needs to be given to critical  
134 information, and the presentation of the data makes this more difficult. In particular, well  
135 known anchoring effects associated with mean or median lines can draw attention away  
136 from extreme values for particular presentation types (Broad et al., 2007; Nadav-Greenberg  
137 et al. 2008; Mulder et al., 2020). The availability of easy-to-use tools that make the  
138 development of complex graphical representations of forecasts quick and cheap to produce,  
139 poses new challenges for the geo-scientists. Within the environmental sciences, making  
140 forecasts of natural hazards (such as landfall of hurricanes, flooding, seismic risk and the  
141 changing climate) useful to end-users depends critically on communicating in a concise and  
142 informative way. Particularly as end-users have a wide range of differing expertise, spanning  
143 a spectrum between geo-physical scientists to those with no formal scientific training.  
144 Therefore, the way in which information is displayed is very important for avoiding  
145 misperceptions and ensuring appropriate steps are taken by end-users, especially when  
146 perceptions of natural hazards can differ between experts and non-experts (Fuchs et al.,  
147 2009; Goldberg & Helfman, 2010). Here, we compare the response of three different groups  
148 of end-users with different levels of scientific expertise to the same series of forecast  
149 presentations to explore how more and less complex presentations influence decision  
150 making and perception.

151 Expertise differences may be due to greater familiarity with the ways in which hazard  
152 information is made available. This enables experts to make more economically rational  
153 decisions and to interpret uncertainty information more effectively (Mulder et al., 2020).  
154 However, the role of expertise remains unclear with some studies showing no differences in  
155 decision-making tasks with both experts and non-experts able to process and use forecast  
156 information to make decisions, with the inclusion of uncertainty information found to be  
157 useful for both experts and non-experts (Nadav-Greenberg et al., 2008; Kirschenbaum et al.,  
158 2014; Wu et al., 2014). Furthermore, it is unclear whether presentation of uncertainty  
159 information in visual formats results in benefits over using verbal and numerical expressions.  
160 For instance, uncertainty presented as pictograph or graphical representations may help with  
161 understanding and interpretation (Zikmund-Fisher et al., 2008; Milne et al., 2015; Susac et  
162 al., 2017). Additionally, research is required to examine differences in expertise, particularly  
163 as deterministic construal errors can be made as observers are often unaware that

164 uncertainty is being depicted within visualisations (Joslyn & Savelli, 2021). Inappropriate  
165 information that captures attention is also often relied on, which can distort judgements  
166 (Fundel et al., 2019).

167 Experts are better at directing attention (through eye movements) to the important  
168 information required for making a decision. For example, in judgments of flight failures,  
169 expert pilots were found to make faster and more correct decisions, making more eye  
170 movements to the cues related to failures than non-experts (Schrivver et al, 2008). Kang and  
171 Landry (2014) also found non-experts to improve after they were trained with the eye  
172 movement scan paths of experts; training led non-experts to make fewer errors (false  
173 alarms) on aircraft conflict detection tasks. However, there is little research examining eye  
174 movements when experts and non-experts are required to make decisions using graphical  
175 and numerical forecast information. It is not clear which aspects of forecast information are  
176 being examined and when, and equally which, are being ignored.

177 More generally, research has shown that when viewing images, more fixations are made to  
178 informative regions and areas of interest (Unema et al., 2005). The times at which these  
179 fixations are made has been found to vary depending on task, decision type and expertise.  
180 Antes (1974) found that early fixations, in the first few seconds of viewing pictures, were  
181 towards informative areas. Goldberg and Helfman (2010) also showed that important regions  
182 of interest were fixated early during observation of different graphs. Experts have been  
183 shown to identify and fixate informative aspects of visual information more quickly and more  
184 often than non-experts (Maturi & Sheridan 2020; Charness, Reingold, Pomplun, &  
185 Stampe, 2001; Kundel, Nodine, Krupinski, & Mello-Thoms, 2008). As well as informative  
186 parts of a scene or image, Shimojo et al. (2003) reported that the likelihood that fixation  
187 would be made to the item preferred, increased over time, particularly in the final second  
188 before selection (see also Glaholt & Reingold, 2009; Simion & Shimojo, 2006; Williams et al.,  
189 2018). These results show that informative and preferred areas of images are selectively  
190 fixated early on, more often and for longer. As viewing evolves, fixations start to reflect final  
191 choices and preferences. The temporal development of this is task-dependent and  
192 influenced by expertise.

193 Here, we explore eye movement behaviour to similar hypothetical scenarios but with  
194 particular interest on differences due to participant expertise/background, following the  
195 research discussed, of gaze to graph areas and keys over different time periods of the  
196 decision-making process. Regardless of expertise, the presence of a median line on graphs  
197 has been found to influence the location of participants gaze fixations moving their  
198 distributions closer to the median line (Mulder et al, 2020). Depending on graph type the

199 presence of a key can lead to errors which may be function of finding that the key is not  
200 directly fixated in those representations (Mulder et al., 2020. Here we explore these  
201 patterns, in particular whether these are a function of expertise. As in our companion paper  
202 (Mulder et al., 2023), we examine gaze patterns when faced with the task of making  
203 decisions about a fictional scenario involving the choices between ships of different sizes in  
204 the face of varying ice thickness forecasts (30%,50%,70%), when presented in different  
205 formats (boxplot, fan plot or spaghetti plot, with and without median lines).

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

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

213

## 214 **2. Methodology**

### 215 **2.1 Participants**

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

228

### 229 **2.2 Design and Procedure**

230 Full methodological details are given in our companion paper, but to restate the core  
231 procedure: A hypothetical scenario of ice thickness forecast for a fictional location was

232 provided to participants. This type of forecast was chosen as is very unlikely to be one that is  
233 familiar to our participants to minimize any effects of preconceived notions of uncertainty.  
234 Participants were informed that they were making shipments across an icy strait and, using  
235 ice-thickness forecasts, had to decide whether to send a small ship or large ship. The small  
236 ship could crush 1-meter thick ice whereas the large ship crushes ice larger than this. There  
237 was a differential cost involved in this decision with small ship costing £1000 to send and the  
238 large ship £5000. They were additionally made aware that if the ice was thicker than 1-meter  
239 and small ship was sent, this would incur a cost penalty of £8000.

240 Ice thickness forecasts were presented in seven different types: deterministic line, box plot,  
241 fan plot and spaghetti plot. Each representation was presented with or without a median line.  
242 Each of these graph types was shown to represent 30%, 50%, and 70% probability of ice  
243 thickness exceeding 1 meter. In this paper we only examined the decision-task question  
244 where participants were asked to select which ship (small or large) to send across an icy  
245 strait 72 hours ahead of time using a 72-hour forecast of ice thickness (see our companion  
246 paper Mulder et al. (2023) for further details on the hypothetical scenarios). While performing  
247 this task, participants wore an Eye link II eye-tracker headset which recorded eye  
248 movements of the right eye as they completed the survey. Head movements were  
249 restrained, and the eye tracker was calibrated to ensure accurate eye movement recording.

### 250 **2.3 Eye tracking apparatus**

251 Participants wore an EyeLink II tracker headset (SR Research Ltd: see [https://www.sr-](https://www.sr-research.com/eyelink-ii/)  
252 [research.com/eyelink-ii/](https://www.sr-research.com/eyelink-ii/) for more details and pictures of the device) which recorded eye  
253 movements of the right eye at a rate of 500Hz as they completed the task. The EyeLink II is  
254 a high-resolution comfortable head-mounted video-based eye tracker with 0.5 deg average  
255 accuracy and 0.01 deg resolution that gives highly accurate spatial and temporal resolution.  
256 Participants gaze was precisely calibrated and re-calibrated throughout the study as  
257 necessary to maintain accurate recording. Each forecast, and task were presented on a 21-  
258 inch colour desktop PC with a monitor refresh rate of 75Hz. Participants were seated at a  
259 distance of 57 cm from the monitor and their head movements were minimized by a chin  
260 rest. Fixation location and its duration were extracted after study completion. Fixation was  
261 defined as times when the eyes were still and not in motion (i.e., no saccades were  
262 detected). These measures were used as proxies of the aspects of the forecasts were being  
263 attended to by participants as they made their decisions. These give a direct insight into the  
264 information and visual features that are salient when participants are attempting to  
265 understand and use uncertainty in forecasting in order to make decisions. For more  
266 information on methods used in eye-tracking studies, see Holmqvist et al. (2011).



## 267 **2.4 Data analysis**

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

275 In our companion paper (Mulder et al., 2023), our analysis of gaze was across all  
276 experimental trials and all tasks. However, as we are concerned about the viewing period  
277 and want to avoid effects of learning, we examine gaze when participants were faced with  
278 each graph type for the first time. Repeated exposure to graph type and the demand to  
279 make the same judgement may influence gaze patterns as informative parts of the figures  
280 are located more swiftly. Therefore, six trials for each graph type for each participant were  
281 examined. We analysed the accuracy of responses to this question (making the safe and  
282 cost-effective choice of the two options) and gaze (number and total fixation duration).

283

## 284 **3. Results**

285 Based on the results of our companion paper (Mulder et al., 2023), we further explore the  
286 impact of the presence of a median line considering the viewing period, expertise and graph  
287 type. We then focus on fixation towards the keys including viewing period, expertise, graph  
288 type and the presence of a median line as variables. For both research questions a four-way  
289 mixed measures ANOVA was conducted including graph type, presence of a median line  
290 and viewing period as within-subject variables, and expertise as a between-subjects  
291 variable. Finally, we report the accuracy of responses for the ice ship decision task  
292 highlighting any differences due to expertise.

293

### 294 **3.1 Does the presence of a median line and expertise affect gaze over the course of** 295 **the decision-making process?**

296 Here, we examined how the presence of the median line influences eye movement  
297 behaviour when considered across the viewing period from early to late stages, and different  
298 levels of expertise, as well as the graph type.

299 A main effect of presence of a median line was found for number of fixations and total  
300 fixation duration made to the graph area,  $F(1, 62)= 6.403$ ,  $MSE=32.747$ ,  $p=0.014$ ,  $\eta^2$   
301  $=0.094$ ;  $F(1, 62)= 7.125$ ,  $MSE=2386741.96$ ,  $p=0.01$ ,  $\eta^2=0.103$ . More fixations were made,  
302 and more time was spent fixating on the graph area of the display when no median line was  
303 present (fixation count  $M=8.74$ ; total duration  $M=2128.64$ ) compared to when a median line  
304 was provided (fixation count  $M=7.89$ ; total duration  $M=1887.47$ ).

305 A main effect of graph type was also found for number of fixations and total fixation duration  
306 made to the graph area,  $F(2, 124)= 15.098$ ,  $MSE=26.406$ ,  $p<0.001$ ,  $\eta^2=0.196$ ;  $F(2, 124)=$   
307  $16.810$ ,  $MSE=1635280.256$ ,  $p<0.001$ ,  $\eta^2=0.213$ . Boxplots elicited more fixations, and more  
308 time was spent fixating on boxplots (fixation count  $M=9.07$ ; total duration  $M=2222.21$ ) and  
309 fan plots (fixation count  $M=8.71$ ; total duration  $M=2091.04$ ) compared to spaghetti plots  
310 (fixation count  $M=7.17$ ; total duration  $M=1710.92$ ).

311 There was also a main effect of the viewing period for number of fixations and total fixation  
312 duration made to the graph area,  $F(2, 124)= 59.608$ ,  $MSE=36.762$ ,  $p<0.001$ ,  $\eta^2=0.488$ ;  $F(2,$   
313  $124)= 57.417$ ,  $MSE=2294640.505$ ,  $p<0.001$ ,  $\eta^2=0.481$ . There was found to be a greater  
314 number of fixations with longer dwell times on the graph area during early (fixation count  
315  $M=9.83$ ; total duration  $M=2399.96$ ) and intermediate (fixation count  $M=9.52$ ; total duration  
316  $M=2284.11$ ) viewing periods compared to later periods (fixation count  $M=5.60$ ; total duration  
317  $M=1340.09$ ).

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

321 As well as the main effects of median line, graph type and viewing period, there was an  
322 interaction between the median line and viewing period for total fixation duration,  $F(2, 124)=$   
323  $3.598$ ,  $MSE=1543871.74$ ,  $p=0.03$ ,  $\eta^2=0.055$ . Less time was spent fixating the graph area  
324 during the early and intermediate stages of viewing when a median line was present (Early  
325 total duration  $M= 2174.97$ ; Intermediate total duration  $M= 2137.79$ ) compared to when no  
326 median line was present (Early total duration  $M= 2624.96$ ; Intermediate total duration  $M=$   
327  $2430.43$ ),  $p<0.001$ ;  $p=0.05$ , respectively. However, no differences were found due to the  
328 presence (later total duration  $M= 1349.65$ ) or absence (later total duration  $M= 1330.54$ ) of a

329 median line during the later stages,  $p=0.896$ . No other interactions were found to be  
330 significant. These findings support that the median line can reduce cognitive load; impacting  
331 the total fixation duration and number of fixations made on the graph area, particularly during  
332 early stages of the decision-making process, and adds to results from our companion paper  
333 that showed how fixation location was towards the median line when present, regardless of  
334 the type of graph.

335

### 336 **3.2 Is gaze to the key influenced by expertise and the viewing period during the** 337 **decision-making process?**

338 In order to examine how gaze parameters on the graph key change throughout the viewing  
339 period prior to the final decision, we extracted the number of fixations made to the key and  
340 their duration.

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

348 There was a main effect of the viewing period on the number of fixations that were made to  
349 the key within the display, as well as the total amount of fixation,  $F(2, 124)= 17.967$ ,  
350  $MSE=6.593$ ,  $p<0.001$ ,  $\eta^2=0.225$ ;  $F(2, 124)= 21.003$ ,  $MSE=416719.669$ ,  $p<0.001$ ,  $\eta^2$   
351  $=0.253$ . More fixations and longer dwell time to the key occurred during the early (fixation  
352 count  $M=1.61$ ; total duration  $M=407.15$ ) and intermediate (fixation count  $M=1.99$ ; total  
353 duration  $M=515.33$ ) viewing periods compared to later periods (fixation count  $M=0.90$ ; total  
354 duration  $M=219.20$ ).

355 No main effect of the median line on gaze to the key, measured by both fixation count and  
356 total duration, was found;  $F(1, 62)= 0.175$ ,  $MSE=7.574$ ,  $p=0.677$ ,  $\eta^2=0.003$ ;  $F(1, 62)=$   
357  $0.061$ ,  $MSE=543399.152$ ,  $p=0.805$ ,  $\eta^2=0.001$ , respectively. Nor was there a main effect of

358 expertise on fixation count and total fixation duration;  $F(1, 62)= 0.251$ ,  $MSE=10.191$ ,  
 359  $p=0.779$ ,  $\eta^2=0.008$ ;  $F(1, 62)= 0.141$ ,  $MSE=730099.249$ ,  $p=0.869$ ,  $\eta^2=0.005$ , respectively.

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

373

### 374 **3.3 Does expertise affect accuracy of decisions?**

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

379

	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%

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

382

383 Overall, participants were accurate in their choice of ship (Meteorology= 85.5%;  
 384 Psychology= 77.9%; Graphic communication = 80.7%); however, some differences were

385 apparent due to expertise. A one-way ANOVA shows differences in accuracy when  
386 presented with 50% probability of risk, which is the most challenging task,  $F(2,64)= 4.029$ ,  
387  $p=0.023$ . Multiple comparisons show meteorology students to be significantly more accurate  
388 than psychology students in choosing the large ship during these scenarios,  $p=0.035$ , and  
389 more accurate than graphic communication students, although this difference is not  
390 significant,  $p=0.08$ . No differences between expertise were found for the 30% and 70% trials,  
391  $p>0.05$ .

392

#### 393 **4. Discussion and Conclusions**

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

402 Prior research presents mixed results, with some authors suggesting that when making  
403 slight variations to graph representations that display uncertainty, decisions and  
404 interpretations differ (Correll & Gleicher, 2014; Tak et al., 2015), whilst others show that  
405 despite greater discrepancies in forecast representation, such as between graphic  
406 visualisations and written forms, there are no differences (Nadav-Greenberg & Joslyn,  
407 2009). Furthermore, few studies explore how experts and non-experts interpret forecast  
408 information from different types of graphical forecast representations (Mulder et al., 2020).  
409 The current research examines these areas further by using eye-movement techniques  
410 considering expertise, and the viewing period during the decision-making process when  
411 observing a range of graph types.

412 More economically rational responses to the ship decision were made by meteorology  
413 students (greater level of expertise) during the most difficult scenarios. We found  
414 participants, regardless of expertise, to spend less time fixating the overall graph when a  
415 median line was presented, particularly during early and intermediate stages of viewing. This  
416 provides more evidence for the anchoring bias suggested in previous papers (Mulder et al.,  
417 2020). Participants focussed on the key for boxplots and fan plots more during early and  
418 intermediate stages compared to later stages. This provides evidence that early stages of

419 viewing are more exploratory and towards informative areas (Buswell, 1935; Yarbush, 1967;  
420 Antes, 1974; Nodine et al, 1993; Locher, 2006; Locher et al, 2007; Locher, 2015; Goldberg &  
421 Helfman, 2010). However, considering the results and the differences found due to graph  
422 type, spaghetti plots appear to be simpler to interpret, potentially reducing cognitive load  
423 (Walter and Bex, 2021), corroborating the findings in Mulder et al. (2020) that the spaghetti  
424 plot helped users interpret extreme values.

425 Overall, this study, together with the analysis in our companion paper (Mulder et al., 2023),  
426 demonstrates that there are many challenges when presenting natural hazard data to both  
427 experts and non-experts, the way that information is portrayed can impact interpretations  
428 and decisions. It is important to note that the graph area and key discussed here are specific  
429 to the particular tasks presented in this study and are used as indicators of the impact of  
430 expertise, graph type and the viewing period. Furthermore, course of study within higher  
431 education was used as a proxy for expertise, with meteorology students being regarded to  
432 have higher levels. However, future research would benefit from examining behaviour and  
433 decisions of academics and forecasters who would be considered as experts.

434 Responses to the ship decision (small or large) based on economic rationality supports the  
435 importance of expertise as accuracy reduces dependent on the probability of ice thickness,  
436 with those with greater expertise being more accurate during more uncertain situations.  
437 While their accuracy was as low as others for 30% probability conditions, with a little less  
438 uncertainty (50% probability of risk) accuracy improved more so than the other groups. This  
439 suggests that they were able to use their expertise to understand the forecasts to inform  
440 their decisions more effectively than the other groups. However, expertise appears to have  
441 little impact on eye movement behaviour within our study. Differences between experts and  
442 non-experts on decisions and interpretations of best-guess forecasts and their inference of  
443 uncertainty have been reported previously (Mulder et al., 2020). However, Doyle et al.  
444 (2014) found no differences in the use of probabilistic information for forecasts of volcanic  
445 eruptions. Other contradictory evidence has also been reported testing numeracy as a  
446 predictor for making economically rational decisions (Roulston and Kaplan, 2009; Tak et al.,  
447 2015). Differences may be due to what “expert” means in these circumstances. As pointed  
448 out, our sample used years of study as the expertise proxy and while showing some effect  
449 may not reflect the decision-making and behaviour of those with many years of experience.  
450 Thus, it may well be the case that those with greater expertise would show a more effective  
451 use of forecast information provided both in terms of accuracy and more effective  
452 information extract shown through eye movement differences not found in our sample.

453 The results show how median lines can reduce cognitive load drawing users to the central  
454 estimate regardless of expertise. A median line reduces the perceived uncertainty in a  
455 graphic, even when explicitly presented (Mulder et al. 2020), so use of a median line should  
456 be used when the amount of uncertainty in the estimate is less critical to understand. Use of  
457 the key within graphical representations can also impact interpretations of data. For forecast  
458 providers this suggests that standard information design principles which seek to reduce  
459 visual noise in data presentation and draw the user to the critical parts can have major  
460 benefits for their ability to effectively communicate with both expert and non-expert end-  
461 users.

462 More broadly, taken together the results reported here and those reported by Mulder et al  
463 (2023) suggest that incorporating eye-tracking and other techniques from cognitive science  
464 into the process of the design of forecast communication tools could be extremely fruitful.  
465 These techniques are now well-established with technology that makes them relatively  
466 cheap to set up and use. Graphical presentation of geo-scientific forecasts can happen with  
467 a range of breadth and longevity of communication in mind. While eye-tracking and related  
468 techniques would not be appropriate for all purposes, where graphics are being developed  
469 for routine and wide use, for example routine weather forecasts, this kind of approach would  
470 be a very valuable addition to end-user engagement. One obvious extension to the work in  
471 the two parts of this study is applying the same techniques to well-known and widely used  
472 geo-scientific forecast graphics.

473

## 474 **5. Author contributions**

475 Louis Williams: Conceptualization, Investigation, Formal analysis, Writing – original draft  
476 preparation

477 Kelsey Mulder: Writing – review & editing

478 Andrew Charlton-Perez: Funding acquisition, Writing – review & editing

479 Matthew Lickiss: Writing – review & editing

480 Alison Black: Funding acquisition, Writing – review & editing

481 Rachel McCloy: Funding acquisition, Writing – review & editing

482 Eugene McSorley: Conceptualization, Resources, Writing – review & editing

483 Joe Young: Funding acquisition

484 *Acknowledgments.* We thank our eye-tracking study participants. This research is  
485 funded by the Natural Environment Research Council (NERC) under the Probability,  
486 Uncertainty and Risk in the Environment (PURE) Programme (NE/J017221/1). Data created

487 during the research reported in this article are openly available from the University of  
488 Reading Research Data Archive at <http://dx.doi.org/10.17864/1947.110>

489

490 The authors declare that they have no conflict of interest.

491

## 492 **Ethical Statement**

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

496

## 497 **References**

498 Ash, K. D., Schumann III, R. L., and Bowser, G. C.: Tornado warning trade-offs: Evaluating  
499 choices for visually communicating risk, *Weather, climate, and society*, 6, 104–118, 2014.

500

501 Antes, J.: The time course of picture viewing. *Journal of Experimental Psychology*,  
502 103(1), 62–70, 1974, <http://doi:10.1037/h0036799>

503

504 Bosetti, V., Weber, E., Berger, L., Budescu, D. V., Liu, N., and Tavoni, M.: COP21 climate  
505 negotiators' responses to climate model forecasts, *Nature Climate Change*, 7, 185–190,  
506 2017.

507

508 Bostrom, A., Morss, R. E., Lazo, J. K., Demuth, J. L., Lazrus, H. and Hudson, R.: A Mental  
509 Models Study of Hurricane Forecast and Warning Production, Communication, and  
510 Decision-Making. *Weather, Climate and Society*, 8, 111–129, 2016,  
511 <https://doi.org/10.1175/WCAS-D-15-0033.1>.

512

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

516



517 Broad, K., Demuth, J. L., Morss, R. E., Hearn-Morrow, B, and Lazo, J. L.: Creation and  
518 communication of hurricane risk information. *Bulletin of the American Meteorological*  
519 *Society*, 93, 1133–1145, 2012, doi:10.1175/ BAMS-D-11-00150.1.

520

521 Charness, N., Reingold, E. M., Pomplun, M., and Stampe, D. M.: The perceptual aspect of  
522 skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, 29(8),  
523 1146–1152, 2001. <https://doi.org/10.3758/BF03206384>

524

525 Correll, M., and Gleicher, M.: Error bars considered harmful: Exploring alternate encodings  
526 for mean and error. *IEEE transactions on visualization and computer graphics*, 20(12), 2142-  
527 2151, 2014. <http://doi:10.1109/TVCG.2014.2346298>

528

529 Denes-Raj, V. and Epstein, S.: Conflict between intuitive and rational processing: when  
530 people behave against their better judgment. *Journal of personality and social*  
531 *psychology*, 66, p.819, 1994.

532

533 Doyle, E.E., McClure, J., Johnston, D.M. and Paton, D: Communicating likelihoods and  
534 probabilities in forecasts of volcanic eruptions. *Journal of Volcanology and Geothermal*  
535 *Research*, 272, pp.1-15, 2014.

536

537 Feldman-Stewart, D., Brundage, M. D., and Zotov, V.: Further insight into the perception of  
538 quantitative information: judgments of gist in treatment decisions. *Medical Decision Making*,  
539 27: 34–43, 2007.

540

541 Fischhoff, B.: *Communicating Risks and Benefits: An Evidence-Based User's Guide*.  
542 Government Printing Office, 2012

543

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

547

548 Fundel, V. J., Fleischhut, N., Herzog, S. M., Göber, M., and Hagedorn, R.: Promoting the  
549 use of probabilistic weather forecasts through a dialogue between scientists, developers and  
550 end-users. *Quarterly Journal of the Royal Meteorological Society*, 145, 210-231, 2019,  
551 <https://doi.org/10.1002/qj.3482>  
552

553 Galesic, M., Garcia-Retamero, R. and Gigerenzer, G.: Using icon arrays to communicate  
554 medical risks: overcoming low numeracy. *Health psychology*, 28, 210, 2009.  
555

556 Garcia-Retamero, R., Galesic, M. and Gigerenzer, G.: Do icon arrays help reduce  
557 denominator neglect? *Medical Decision Making*, 30, 672-684, 2010.  
558

559 Gigerenzer, G., and Hoffrage, U.: How to improve Bayesian reasoning without instruction:  
560 Frequency formats. *Psychological Review*, 102, 684–704, 1995,  
561 <https://doi.org/10.1037/0033-295X.102.4.684>  
562

563 Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L.M. and Woloshin, S.: Helping  
564 doctors and patients make sense of health statistics. *Psychological science in the public*  
565 *interest*, 8, 53-96, 2007.  
566

567 Gustafson, A., and Rice, R. E.: The Effects of Uncertainty Frames in Three Science  
568 Communication Topics. *Science Communication*, 41(6), 679–706, 2019,  
569 <doi.org/10.1177/1075547019870811>  
570

571 Glaholt, M. G., and Reingold, E. M.: The time course of gaze bias in visual decision  
572 tasks. *Visual Cognition*, 17(8), 1228-1243, 2009,  
573 <http://dx.doi.org/10.1080/13506280802362962>  
574

575 Goldberg, J. H., and Helfman, J. I.: Comparing information graphics: a critical look at eye  
576 tracking. In *Proceedings of the 3rd BELIV'10 Workshop: Beyond time and errors: novel*  
577 *evaluation methods for Information Visualization*, 71-78, 2010, ACM. [http://](http://doi:10.1145/2110192.2110203)  
578 [doi:10.1145/2110192.2110203](http://doi:10.1145/2110192.2110203)

579

580 Harold, J., Lorenzoni, I., Shipley, T. F., and Coventry, K. R.: Cognitive and psychological  
581 science insights to improve climate change data visualization, *Nature Climate Change*, 6,  
582 1080–1089, 2016.

583

584 Hullman, J.: Why Authors Don't Visualize Uncertainty, *IEEE Transactions on Visualization*  
585 *and Computer Graphics*, 26, 130-139, 2020, doi: 10.1109/TVCG.2019.2934287.

586

587 Joslyn, S.L. and Nichols, R.M.: Probability or frequency? Expressing forecast uncertainty in  
588 public weather forecasts. *Meteorological Applications*, 16, 309-314,  
589 2009, <https://doi.org/10.1002/met.121>

590

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

594

595 Kang, Z., and Landry, S. J.: Using scanpaths as a learning method for a conflict detection  
596 task of multiple target tracking. *Human Factors: The Journal of the Human Factors and*  
597 *Ergonomics Society*, 56, 6, 1150-1162, 2014, 0018720814523066.  
598 <https://doi.org/10.1177/0018720814523066>

599

600 Kelton, A. S., Pennington, R. R., and Tuttle, B. M.: The effects of information presentation  
601 format on judgment and decision making: A review of the information systems research,  
602 *Journal of Information Systems*, 24, 79–105, 2010.

603

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

607

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

612

613 Kurz-Milcke, E., Gigerenzer, G., and Martignon, L.: Transparency in risk communication:  
614 graphical and analog tools. *Annals of the New York Academy Sciences*, 1128:18-28, 2008,  
615 doi: 10.1196/annals.1399.004. PMID: 18469211.

616

617 Lipkus, I.M.: Numeric, verbal, and visual formats of conveying health risks: suggested best  
618 practices and future recommendations. *Medical decision making*, 27, pp.696-713, 2007.

619

620 Lipkus, I.M. and Hollands, J.G.: The visual communication of risk. *JNCI monographs*, 1999,  
621 149-163, 1999.

622

623 Lorenz, S., Dessai, S., Forster, P. M., and Paavola, J.: Tailoring the visual communication of  
624 climate projections for local adaptation practitioners in Germany and the UK, *Philosophical  
625 Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373,  
626 2015.

627

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

631

632 Milne, A. E., Glendining, M. J. , Lark, R. M., Perryman, S. A., Gordon, T., and Whitmore, A.  
633 P.: Communicating the uncertainty in estimated greenhouse gas emissions from agriculture.  
634 *Journal of Environmental Management*, 160, 139-53, 2015. doi:  
635 10.1016/j.jenvman.2015.05.034.

636

637 Morss, R., Demuth, J.L., and Lazo, J. K.: Communicating uncertainty in weather forecasts: A  
638 survey of the U.S. public. *Weather Forecasting*, 23, 974–991, 2008,  
639 doi:10.1175/2008WAF2007088.1.

640

641 Morss, R. E., Demuth, J. L., Bostrom, A., Lazo, J. K., and Lazrus, H.: Flash flood risks and  
642 warning decisions in Boulder, Colorado: A mental models study of forecasters, public

643 officials, and media broadcasters in Boulder, Colorado. *Risk Analysis*, 35(11), 2009-28,  
644 2015. doi: 10.1111/risa.12403.

645

646 Mulder, K. J., Lickiss, M., Black, A., Charlton-Perez, A. J., McCloy, R., and Young, J. S.:  
647 Designing environmental uncertainty information for experts and non-experts: Does data  
648 presentation affect users' decisions and interpretations? *Meteorological Applications*, 27,  
649 e1821, 2020.

650

651 Mulder, K., Williams, L., Lickiss, M., Black, A., Charlton-Perez, A., McCloy, R., McSorley, E.  
652 and Young, J., 2023. Understanding representations of uncertainty, an eye-tracking study  
653 part II: The effect of expertise. *EGUsphere*, pp.1-15.

654

655 Nadav-Greenberg, L. and Joslyn, S. L.: Uncertainty forecasts improve decision making  
656 among nonexperts, *Journal of Cognitive Engineering and Decision Making*, 3, 209–227,  
657 2009.

658

659 Nadav-Greenberg, L., Joslyn, S. L., and Taing, M. U.: The effect of uncertainty visualizations  
660 on decision making in weather forecasting, *Journal of Cognitive Engineering and Decision*  
661 *Making*, 2, 24–47, 2008.

662

663 Nelson, D.E., Hesse, B.W., and Croyle, R.T.: *Making Data Talk: The Science and Practice of*  
664 *Translating Public Health Research and Surveillance Findings to Policy Makers, the Public,*  
665 *and the Press.* Oxford University Press, 2009.

666

667 Padilla, L., Hansen, G., Ruginski, I. T., Kramer, H. S., Thompson, W. B., and Creem-Regehr,  
668 S. H.: The influence of different graphical displays on nonexpert decision making under  
669 uncertainty. *Journal of Experimental Psychology: Applied*, 21, 37–46, 2015. doi:  
670 10.1037/xap0000037

671

672 Peters, E.: Numeracy and the Perception and Communication of Risk. *Annals of the New*  
673 *York Academy of Sciences*, 1128, 1-7, 2008, <https://doi.org/10.1196/annals.1399.001>

674

675 Peters, E., Hibbard, J., Slovic, P., and Dieckmann, N.: Numeracy skill and the  
676 communication, comprehension, and use of risk-benefit information. *Health affairs*, 26, 741-  
677 748, 2007.

678

679 Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S.,  
680 Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L.,  
681 Cirillo, P., Clements, M. P., Cordeiro, C., Oliveira, F. L. C., De Baets, S., Dokumentov,  
682 A., Ellison, J., Fiszeder, P., Franses, P. H., Frazier, D. T., Gilliland, M., Gönül, M. S.,  
683 Goodwin, P., Grossi, L., Grushka-Cockayne, Y., Guidolin, M., Gunter, U., Guo,  
684 X., Guseo, R., Harvey, N., Hendry, D. F., Hollyman, R., Januschowski, T., Jeon, J., Jose, V.  
685 R. R., Kang, Y., Koehler, Anne B. Kolassa, S., Kourentzes, N., Leva, S., Li, F., Litsiou, K.,  
686 Makridakis, S., Martin, G. M., Martinez, A. B., Meeran, S., Modis, T., Nikolopoulos, K.,  
687 Önköl, D., Paccagnini, A., Panagiotelis, A., Panapakidis, I., Pavía, J. M., Pedio, M.,  
688 Pedregal, D. J., Pinson, P., Ramos, P., Rapach, D. E., Reade, J. J., Rostami-Tabar, B.,  
689 Rubaszek, M., Sermpinis, G., Shang, H. L., Spiliotis, E., Syntetos, A. A., Talagala, P. D.,  
690 Talagala, T. S., Tashman, L., Thomakos, D., Thorarinsdottir, T., Todini, E., Arenas, J. R. T.,  
691 Wang, X., Winkler, R. L., Yusupova, A., and Ziel, F.: Forecasting: theory and practice,  
692 *International Journal of Forecasting*, 38, 705–871, 2022. Roulston, M. S. and Kaplan, T. R.:  
693 A laboratory-based study of understanding of uncertainty in 5-day site-specific temperature  
694 forecasts, *Meteorological Applications: A journal of forecasting, practical applications,*  
695 *training techniques and modelling*, 16, 237–244, 2009.

696

697 Reyna, V.F. and Brainerd, C.J.: Numeracy, ratio bias, and denominator neglect in judgments  
698 of risk and probability. *Learning and individual differences*, 18, 89-107, 2008.

699

700 Roulston, M.S. and Kaplan, T.R.: A laboratory-based study of understanding of uncertainty  
701 in 5-day site-specific temperature forecasts. *Meteorological Applications*, 16, 237–244, 2009,  
702 <https://doi.org/10.1002/met.113>.

703

704 Savelli, S. and Joslyn, S.: The advantages of predictive interval forecasts for non-expert  
705 users and the impact of visualizations, *Applied Cognitive Psychology*, 27, 527–541, 2013.

706

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

710

711 Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M.,  
712 Fowler, H. J., James, R., Maraun, D., Martius, O., and Senior, C. A.: Storylines: an  
713 alternative approach to representing uncertainty in physical aspects of climate change,  
714 *Climatic change*, 151, 555–571, 2018.

715

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

718

719 Sillmann, J., Shepherd, T. G., van den Hurk, B., Hazeleger, W., Martius, O., Slingo, J., and  
720 Zscheischler, J.: Event-based storylines to address climate risk, *Earth's Future*, 9,  
721 e2020EF001 783, 2021.

722

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

726

727 Skubisz, C., Reimer, T., and Hoffrage, U.: Communicating Quantitative Risk Information,  
728 *Annals of the International Communication Association*, 33:1, 177-211, 2009, DOI:  
729 10.1080/23808985.2009.11679087

730

731 Speier, C.: The influence of information presentation formats on complex task decision-  
732 making performance, *International journal of human computer studies*, 64, 1115–1131,  
733 2006.

734

735 Spiegelhalter, D.: Risk and uncertainty communication. *Annual Review of Statistics and Its*  
736 *Application* 4, 31-60, 2017.

737

738 Spiegelhalter, D., Pearson, M., and Short, I.: Visualizing uncertainty about the future,  
739 Science, 333, 1393–1400, 2011.

740

741 St John, M., Callan, J., Proctor, S., and Holste, S.: Tactical decision-making under  
742 uncertainty: Experiments I and II, Tech. rep., PACIFIC 375 SCIENCES AND ENGINEERING  
743 GROUP INC SAN DIEGO CA, 2000.

744

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

749

750 Tak, S., Toet, A., and van Erp, J.: The perception of visual uncertainty representation by  
751 non-experts, *IEEE transactions on visualization and computer graphics*, 20, 935–943, 2013.

752

753 Tak, S., Toet, A., & Van Erp, J.: Public understanding of visual representations of uncertainty  
754 in temperature forecasts. *Journal of cognitive engineering and decision making*, 9, 3, 241-  
755 262, 2015, <https://doi.org/10.1177/1555343415591275>

756

757 Tversky, A. and Kahneman, D.: Judgment under uncertainty: Heuristics and biases, *science*,  
758 185, 1124–1131, 1974.

759

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

764

765 Wallsten T. S., Budescu D. V., Rapoport A., Zwick R., and Forsyth B.: Measuring the vague  
766 meaning of probabilistic terms. *Journal of Experimental Psychology: General*, 155, 348-365,  
767 1986.

768

769 Walter, K., and Bex, P.: Cognitive load influences oculomotor behavior in natural scenes.  
770 *Scientific Reports*, 11, 12405, 2021, <https://doi.org/10.1038/s41598-021-91845-5>



771

772 Wickens, C. D., Helton, W. S., Hollands, J. G., and Banbury, S.: Engineering psychology and  
773 human performance, Routledge, 2021.

774

775 Williams, L., McSorley, E., and McCloy, R.: The relationship between aesthetic and drawing  
776 preferences. *Psychology of Aesthetics, Creativity, and the Arts*, 12, 3, 259, 2018.

777 <https://doi.org/10.1037/aca0000188>

778

779 Wu, H. C., Lindell, M. K., Prater, C. S., and Samuelson, C. D.: Effects of track and threat  
780 information on judgments of hurricane strike probability. *Risk analysis*, 34, 6, 1025-1039,

781 2014, <https://doi.org/10.1111/risa.12128>

782