DOI: 10.1111/cgf.15094

Eurographics Conference on Visualization (EuroVis) 2024
W. Aigner, D. Archambault, and R. Bujack
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 43 (2024), Number 3

Transparent Risks: The Impact of the Specificity and Visual Encoding of Uncertainty on Decision Making

L. E. Matzen, B. C. Howell, M. Tuft, and K. M. Divis

Sandia National Laboratories, USA

Abstract

People frequently make decisions based on uncertain information. Prior research has shown that visualizations of uncertainty can help to support better decision making. However, research has also shown that different representations of the same information can lead to different patterns of decision making. It is crucial for researchers to develop a better scientific understanding of when, why and how different representations of uncertainty lead viewers to make different decisions. This paper seeks to address this need by comparing geospatial visualizations of wildfire risk to verbal descriptions of the same risk. In three experiments, we manipulated the specificity of the uncertain information as well as the visual cues used to encode risk in the visualizations. All three experiments found that participants were more likely to evacuate in response to a hypothetical wildfire if the risk information was presented verbally. When the risk was presented visually, participants were less likely to evacuate, particularly when transparency was used to encode the risk information. Experiment 1 showed that evacuation rates were lower for transparency maps than for other types of visualizations. Experiments 2 and 3 sought to replicate this effect and to test how it related to other factors. Experiment 2 varied the hue used for the transparency maps and Experiment 3 manipulated the salience of the borders between the different risk levels. These experiments showed lower evacuation rates in response to transparency maps regardless of hue. The effect was partially, but not entirely, mitigated by adding salient borders to the transparency maps. Taken together, these experiments show that using transparency to encode information about risk can lead to very different patterns of decision making than other encodings of the same information.

CCS Concepts

• ***Human-centered computing*** → ***Empirical studies in visualization;***

1. Introduction

Data visualizations are frequently used to convey complex information in support of human decision making. Visualization can be a very useful tool for helping viewers to understand important information and make appropriate decisions. However, there is a growing body of research indicating that visualizing the same information in different ways can impact viewers' comprehension of the information and their subsequent decisions. Much of the research on this topic has focused on comparing visualizations that represent uncertainty in different ways. Decision making under uncertainty is particularly challenging. It is difficult to understand and reason about uncertain information [HQC*18, PKH23], and even the "right" decision can lead to a negative outcome [CBK*16]. On top of these difficulties, visualizations of uncertain information can produce visual-spatial and cognitive biases that impact viewers' decisions [FPS*21, MHTD23, PCRHS18, PRCR17]. These biases can compound the problem, increasing the likelihood of making a decision that has an unanticipated or negative outcome. We do not yet have a systematic understanding of when and how different types of visual cues lead to different patterns of decision making [HQC*18, MHTD23] and additional research is needed to develop cognition-based guidelines for minimizing the decision making biases that can result from visualizations of uncertainty.

Several studies have done head-to-head comparisons of different methods of encoding uncertainty to test how those encodings impact decision making tasks. These studies have consistently found that different encodings can lead to different patterns of decisions, even though the underlying information is the same [Heg11]. Some studies have also compared visualizations of uncertainty to verbal or numerical representations of uncertainty [BCJ*11, BMM05, CBK*16, MHTD23]. These studies have indicated that visualizing uncertainty may make the uncertain outcome seem more concrete or deterministic to viewers, which can lead to errors in interpretation [JL13] or more risk-averse patterns of decision making [BMM05, BAF08, MHTD23, SS05]. For example, [MHTD23] used a wildfire evacuation task in which participants saw the probability that a house would be in the burn zone of a fire represented visually, as an icon array, or numerically, as a natural frequency. They found higher evacuation rates (a more risk-averse pattern of decisions) in response to the visualizations.

In contrast, other studies have observed more risk-averse decisions for text rather than visualizations. An experiment by Cheong and colleagues [CBK*16], which also used a wildfire evacuation task, found higher evacuation rates for text representations than for map-based visualizations of the fire risk (when participants were not under time pressure).

In the present study, we sought to investigate this discrepancy: why do visualizations of uncertainty lead to more risk-averse patterns of decision making in some circumstances, but not others? One of the reasons for the different patterns observed in prior research may be the specificity of the uncertain information. Research by Bisantz and colleagues [BMM05] manipulated the specificity of uncertain information in a stock purchasing task by changing the width of the probability bands shown to participants. They found that when the specificity was low (i.e., the probability bands were wide), participants made more risk-averse decisions when shown visual representations of probability. When the specificity of the information increased, the participants in the visualization condition began to act more similarly to the participants in the numerical condition, making bolder stock purchases. Another key difference may be the nature of the visual representation. The icon arrays and natural frequencies used in [MHTD23] provided a point estimate of probability, whereas the probability bands used in [BMM05] and [CBK*16] provided a range of probabilities. Furthermore, [CBK*16] used map visualizations that provided contextual information about the fire and the relative location of the house that was not available in the text representations. In this case, participants may have made more nuanced evacuation decisions when shown the visualizations, leading to lower overall evacuation rates for the maps relative to the text.

In this study, we manipulated the specificity and encoding of uncertain information in geospatial visualizations in order to investigate how those factors impact decision making. Our experiments used a wildfire evacuation task similar to [CBK*16]. In Experiment 1, we compared text representations and map-based visualizations with three different levels of specificity. We hypothesized that we would see different patterns of decision making for different representations of uncertainty, as has been observed in prior studies. We also expected to see an interaction between the representation type and the level of specificity. For the comparison between the visual and text representations, we expected to see one of two patterns:

1. Text representations of uncertainty would lead to higher evacuation rates than visual representations of uncertainty, consistent with the results observed in Experiment 1 of [CBK*16]. This would indicate that participants were using the contextual information provided by the visualizations, such as the location of the house relative to the edges of the probability band, to modulate their evacuation decisions. We expected that the difference between the text and visualization conditions would decrease as the specificity of the information increased, as in [BMM05]. In the high-specificity condition where the probability bands were very narrow, the relative location of the house within a probability band would no longer support inferences about risk beyond those that could be made from the text representations.
2. Visual representations of uncertainty would lead to higher evacuation rates than text representations of uncertainty, consistent with the results observed in [BMM05, MHTD23, SS05]. These studies found that visual representations led to more risk-averse decision making than text or numerical representations. Once again, we expected that the difference between the text and visualization conditions would decrease as the specificity of the information increased.

Experiment 1 also manipulated how uncertainty was encoded in the visualizations. We created four variants of the stimuli, using two types of color hue maps, a transparency map, and an isarithmic map. The rationale for the selection of each visualization condition is outlined in Section 3.

Experiment 1 produced unexpected results in several ways. Contrary to our expectations, we observed *larger* differences in performance between conditions as the specificity of the information increased. This was particularly apparent for the transparency condition, which had substantially lower evacuation rates than any other condition. In Experiments 2 and 3, we sought to replicate and explain this unexpected finding. In Experiment 2, we used isarithmic and transparency maps and manipulated the hue (orange or blue) used to encode risk. We replicated our finding of lower evacuation rates for transparency maps, regardless of the hue used. Based on these findings, we speculated that the salience of the visual boundaries in the maps might be driving performance. Prior studies have found that visual boundaries can have a substantial impact on how people reason about uncertainty visualizations [PCRHS18, PQMCR16, Tve13]. In Experiment 3, we tested this by comparing isarithmic maps with and without contour lines to transparency maps with or without salient lines marking the boundaries of each probability band. Although the addition of the boundary lines to the transparency maps increased the evacuation rates, they were still lower than in the isarithmic map conditions.

Taken together, these experiments reinforce the fact that seemingly trivial choices about how to encode uncertain information can have a dramatic impact on human decisions. Our results also demonstrate that more work remains to be done in order to build a deeper understanding of how different types of encodings impact human cognition.

2. Experiment Structure and Materials

The data collection for each study took place on Amazon Mechanical Turk. The participants were required to have the "masters" qualification, to be located in the United States, and to have an approval rate of greater than 95 percent for prior tasks. All of the participants completed a pre-test containing individual differences measures testing numeracy [FZFU*07, Sch97], graph literacy [OJGW19], and general risk-taking propensity [DFH*11, SCRP12]. Participants were paid \$1.50 USD for completing the pre-test. If they followed the instructions and responded appropriately to all questions, they were given a qualification that allowed them to participate in one of the wildfire evacuation experiments. The wildfire experiments took 15-20 minutes to complete and the participants were paid \$4.75 plus a bonus payment based on their task performance. Across all of the experiments, the bonus payments ranged from \$0.00 to \$2.39, with an average bonus of \$1.59.

2.1. Wildfire Evacuation Task

At the beginning of the evacuation task, the participants were shown a picture of a cabin in the woods and were asked to imagine that they lived there. They were told that there was a wildfire in the area, and that on each trial they would be shown the probability of their house being in the burn zone of the fire. Based on that information, they had to decide whether to stay in their house or to evacuate. After making their decision, they were shown whether or not their house had ended up in the burn zone.

As in prior studies [CBK*16, MHTD23], we incentivized the task by linking participants' decisions to bonus payments. The participants had 5 cents added to their bonus for each correct decision. If they chose to evacuate, they had to pay 2 cents from their bonus, representing the real-world costs of evacuating from one's home. If they chose to stay in their house and it ended up burning down, they lost 10 cents, reflecting the higher cost of failing to evacuate from a dangerous situation. Based on this cost/benefit structure, the participants would maximize their bonus if they chose to evacuate whenever the probability of their house being in the burn zone was 40% or higher. The participants' cumulative bonus amount was shown on the screen at all times during the experiment, but they were not informed of the optimal strategy.

These experiments cannot replicate the real-world consequences and complexity of making decisions about evacuation in the face of a fire or other disaster. However, the bonus payments incentivized the participants to make the best possible decisions with the uncertain information that they were given. These types of tightly-controlled experiments are a necessary first step for understanding how visualizations of uncertainty can influence human decision making. Our aim is to identify fundamental patterns in human reasoning that are likely to translate to real-world decision making.

2.2. Stimulus Design

On each trial in the experiments, participants saw the probability that their house would be in the burn zone of a wild fire. This information could be conveyed as text (e.g. "there is a 40-50% chance that your house will be in the burn zone) or in a visualization. For the visualization conditions, we presented 800x600 pixel maps that were generated in python using the Static Map library [Lin22] with the CARTO "light, no labels" style [Vel22,car]. An X indicating the location of the house and a unique overlay indicating the probability of different regions being in the burn zone of the fire was added to each map. Examples of the stimuli are shown in Figure 1.

Across visualization conditions, we manipulated the visual cues used to represent different levels of risk. We also manipulated the specificity of the information by changing the number (and therefore the width) of probability bands. In the lowest level of specificity, there were three probability bands that represented the risk in 30% increments (e.g., a 10-40% chance of being in the burn zone). The medium-specificity condition had six probability bands that represented the risk in 15% increments and the high-specificity condition had nine probability bands that represented the risk in 10% increments. The specificity of the text stimuli was manipulated in the same way, using the same probability bands.

The X indicating the house location for each map was placed using the nine-band map to ensure that the X would not touch the boundaries between the risk bands. This location was held constant for the three- and six-band versions of the same map. This produced sets of stimuli that had the same background map, house location, and the same underlying probability of that house being in the burn zone. In the example in Figure 1, the underlying risk of the house burning down is 40-50%. Although the participants saw wider probability bands in the three- and six-level conditions, they might make their own inferences about the risk to the house based on its location within the wider bands. This type of inference was not possible for the text stimuli, where participants had no information about the relative location of the house.

In each of the experiments, the maps were rotated through four different visual encoding conditions through the use of four counterbalanced lists. Each list had 135 trials, including 15 trials for each of the nine underlying probability levels (10-20%, 20-30%, etc.). For each of these underlying probability levels, there were 12 map stimuli (one for each visualization condition at each level of specificity) and three text stimuli (one for each level of specificity). With this structure, every version of every map was shown in one of the lists, each map appeared equally often in every specificity and encoding condition, and within each list there were equal numbers of stimuli in each condition. The counterbalancing ensured that any differences in the participants' evacuation decisions were due to the experimental manipulations rather than any uncontrolled differences between the map stimuli.

The structure of the experiment lists also ensured that the probabilities presented to the participants were accurate. For the 15 stimuli in a given list with a 10-20% underlying probability of being in the burn zone, the house burned down in 2 of the 15 trials (13% of the time). The outcomes were pre-assigned and were balanced across specificity and visualization conditions so that no one condition would have more negative outcomes than the others.

Note that the participants saw the less specific text stimuli multiple times. For example, the text stimulus stating "Your house is located in the 10-40% burn likelihood zone" was presented three times. For one of these instances, the actual underlying burn probability was 10-20%, for another it was 20-30%, and for another it was 30-40%. These underlying probabilities were not known to the participants and could not be inferred. From their perspective, they saw the same text stimulus three times and it did not always lead to the same outcome. The participants were instructed that all of the trials were independent of one another, so the outcome of a prior trial had no influence on the outcome of later trials.

Each participant completed one of the experimental lists. The stimuli within the lists were presented in a different random order for each participant. At the end of the experiment, they completed a short questionnaire asking their opinions about the different encoding conditions.

3. Experiment 1

A total of 96 people participated in Experiment 1, with 24 participants in each of the four counterbalancing lists. The four visualization conditions used in Experiment 1 included a hue map and a transparency map based on two of the conditions used in

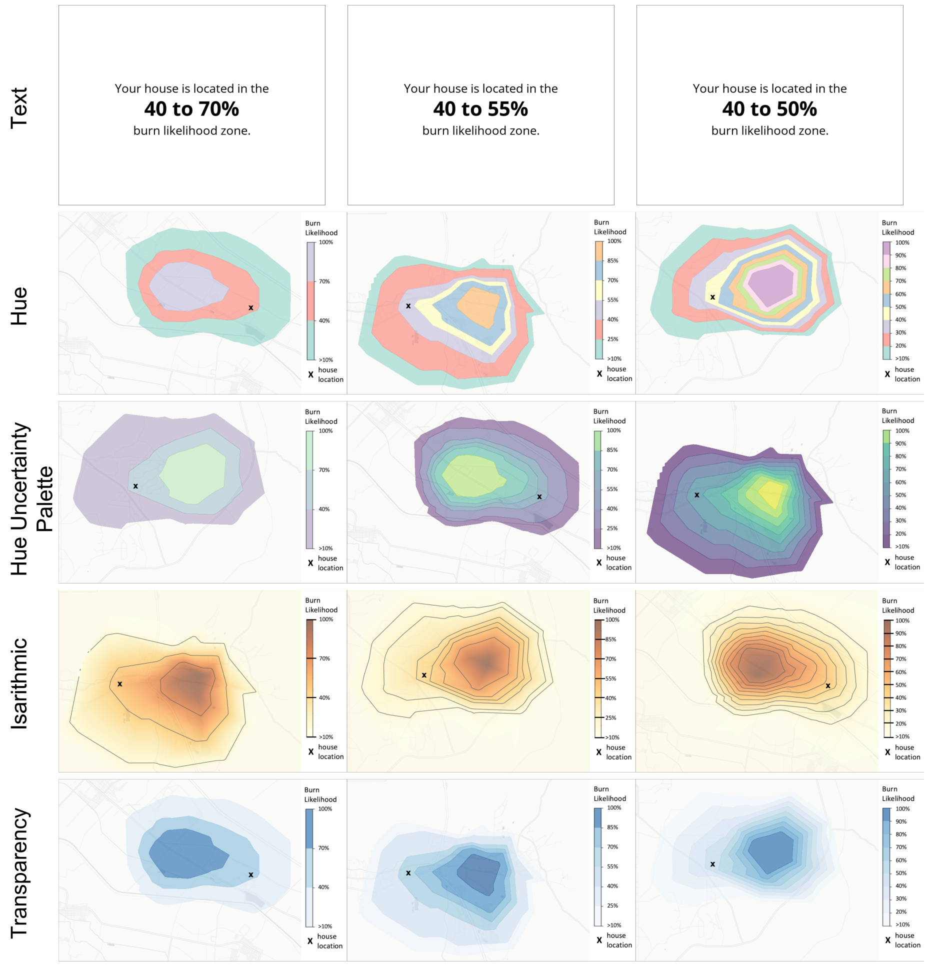

Figure 1: *Examples of the stimuli used in Experiment 1. The columns show the low, medium, and high specificity conditions, from left to right. The rows show the text condition followed by the hue, hue uncertainty palette, isarithmic, and transparency conditions. The main diagonals show examples of the same maps being rotated through visualization conditions.*

[CBK*16], a second hue map using the value-suppressing uncertainty palette [CMH18], and an isarithmic map based on [TLG15]. The hue and transparency conditions were selected because these are common methods of conveying uncertainty in maps, and can have been shown to support higher accuracy in tasks involving geospatial data than other encoding techniques [KMS14]. With respect to judgments of risk, [CBK*16] consistently found higher evacuation rates for transparency maps than for hue maps. We predicted that we would observe a similar pattern. Our hue maps used the Set3 color map from Matplotlib [Hun07], which was chosen because it provided similar hues to those used in [CBK*16]. While [CBK*16] used gray for their transparency maps, we used blue in order to make the overlay more distinct from our background maps.

The second hue map condition was added to test whether a color palette designed for uncertainty visualization would lead to different patterns of decisions than a standard hue map. The value-suppressing uncertainty palette uses more saturated colors for more specific information and less saturated colors for less specific information, providing a visual analog of the increasing uncertainty. We predicted that this additional visual cue might lead to more cautious decisions for the value-suppressing palette relative to the standard hue map at lower levels of specificity.

Finally, we included isarithmic maps because prior research has shown that viewers prefer this type of risk map over others for some types of hazards [TLG15]. The isarithmic maps in our study showed a continuous gradient of color value, with the color becoming lighter as the risk decreased. This gradient was overlaid with boundary lines to delineate regions with similar values. It also lends itself well to the manipulation of specificity, since the underlying gradient was the same for all levels of specificity, but the number of contours changed. We predicted that at high levels of specificity, performance would be similar for the isarithmic maps and the hue maps. At lower levels of specificity, the color gradient might emphasize the difference in risk between the inner and outer edges of the bands, leading to a bigger difference in evacuation rates for houses in different places within the same risk band.

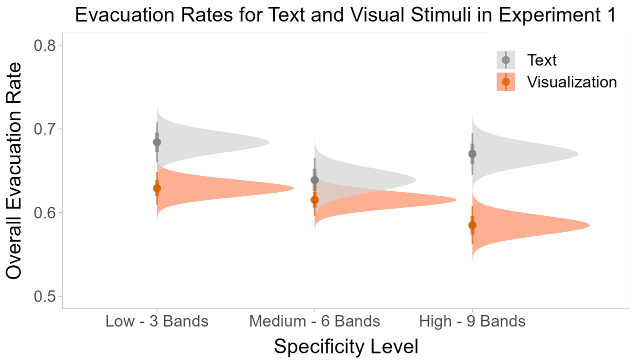

Figure 2: *Average evacuation rates for the text and visualization stimuli in the three different specificity conditions.*

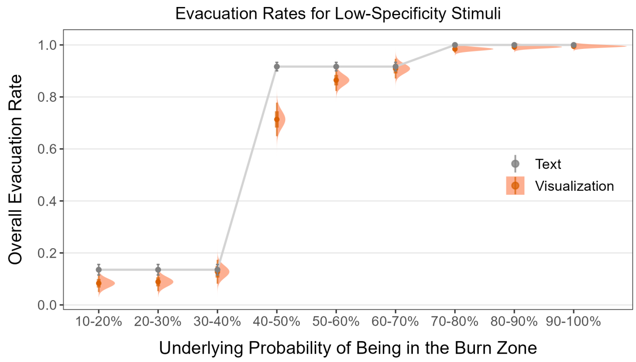

Figure 3: *Average evacuation rates for the text and visualization stimuli in the low-specificity condition for each underlying risk level.*

3.1. Results and Discussion

Two analyses were conducted: one to compare the participants' responses to the text stimuli relative to the visualization stimuli across the three levels of specificity, and one to compare the participants' responses to the four different visualization conditions.

3.1.1. Visualizations Versus Text

Our first analysis compared the text condition to the combined visualization conditions to test how the two different types of information impacted the participants' evacuation rates. The results are shown in Figure 2. A 2x3 repeated measures ANOVA (information type by level of specificity) showed that there was a significant main effect of information type ($F(1, 95) = 56.15$, $p < 0.001$), a significant main effect of specificity level ($F(2, 190) = 12.96$, $p < 0.001$), and a significant interaction between the two ($F(2, 190) = 11.01$, $p < 0.001$). Bonferroni pairwise comparisons showed that the participants had significantly higher evacuation rates for text than for visualizations in the low-specificity ($p < 0.001$) and high-specificity ($p < 0.001$) conditions but not in the medium-specificity condition ($p = 0.15$). In addition, the evacuation rates for the visualizations were significantly lower in the high-specificity condition than in the low-specificity condition ($p < 0.01$).

These results were consistent with the findings of [CBK*16], which also found higher evacuation rates for text stimuli when participants were not under time pressure. This suggests that participants used the contextual information provided by the visualizations to modulate their evacuation decisions. This effect can be seen in more detail in Figure 3, which plots the evacuation rates for the text and visualization stimuli in the low-specificity condition. Each stimulus had an underlying risk level that could be inferred from the relative location of the house within each risk band in the visualizations. Figure 3 shows that the participants were less likely to evacuate when the house was near the outer edge of the band. When the participants saw text stimuli, they could not make this type of inference. Their evacuation rates in the text condition imply that they based their decisions on the upper end of the probability range presented in the text stimuli.

We expected that the difference between the text and visualization conditions would decrease as the specificity of the information increased. However, this was not the case. Instead, the difference was largest at the highest level of specificity. The analysis of the different visualization conditions, described in the next section, helped to explain this effect.

3.1.2. Comparison of the Visualization Conditions

Our next step was to compare the evacuation rates for each visualization condition at each level of specificity. The results are shown in Figure 4. We analyzed the results via two-way repeated measures ANOVAs testing the effects of condition (text, hue, hue uncertainty palette, isartithmic, and transparency) and risk band for each level of specificity. All three ANOVAs found a significant main effect of condition, a significant main effect of risk band, and a significant interaction between the two. Bonferroni pairwise comparisons were used to test the differences between the conditions for each risk band. The detailed results of the statistical tests are presented in the Supplemental Materials. Here, we highlight the notable differences between the visualization conditions.

For the low-specificity condition, there were no significant differences between conditions in the highest or lowest risk band. The participants almost always chose to stay in the lowest risk condition and to evacuate in the highest risk condition, regardless of how the risk information was presented. However, there were some differences between conditions in the intermediate risk band, where there was a 40-70% chance of the house being in the burn zone. The participants were significantly more likely to evacuate when they saw text than when they saw the uncertainty palette or transparency maps. They were also significantly more likely to evacuate when they saw the hue maps than when they saw the transparency maps. Notably, there were no significant differences between the standard hue maps and the value-suppressing uncertainty hue maps. We had predicted that the desaturation of the colors used in the uncertainty palette maps might emphasize the higher uncertainty in the low-specificity condition, leading to higher evacuation rates. However, the participants' responses were very similar for both types of hue maps.

In the medium-specificity condition, the visualization conditions began to diverge more. The participants had the lowest evacuation rates in the transparency condition. Their evacuation rates for transparency were significantly lower than the text condition in the 25-40% risk band and significantly lower than all other conditions in the 40-55% band. By the time the risk level was above 55%, the participants tended to evacuate no matter how the information was shown. The other conditions did not differ significantly from one another for any of the risk bands.

In the high-specificity condition, we observed even greater divergence between the transparency condition and the others. The participants had significantly lower evacuation rates in the transparency condition than in the text condition beginning in the 20-30% risk band and continuing until the 60-70% risk band. The transparency condition also had significantly lower evacuation rates than all but the isarithmic condition in the 30-40% risk band, and lower evacuation rates than all of the other visualization conditions in the 40-50% and 50-60% risk bands.

Meanwhile, the evacuation rates for the hue, uncertainty palette, and isarithmic conditions were generally very similar across all of the risk bands. The one exception was the isarithmic condition in the 30-40% risk band, which had significantly lower evacuation rates than both the text and hue uncertainty palette conditions. This finding suggests that the color values in the isarithmic maps may have had some influence on the participants' decisions, leading them to treat the outermost risk bands with similarly light values as if they were more similar to one another. Aside from this, there were no indications that the participants responded differently to the isarithmic maps than to the hue maps.

The evacuation rates were numerically highest for the text condition for most of the risk bands, but this difference was only significant in the 40-50% band. These results indicate that the lower evacuation rates for visualizations relative to text in our first analysis were largely driven by the low evacuation rates for the transparency condition. The other visual conditions generally had similar evacuation rates to one another and did not differ significantly from the text condition, with the exception of the 40-50% band.

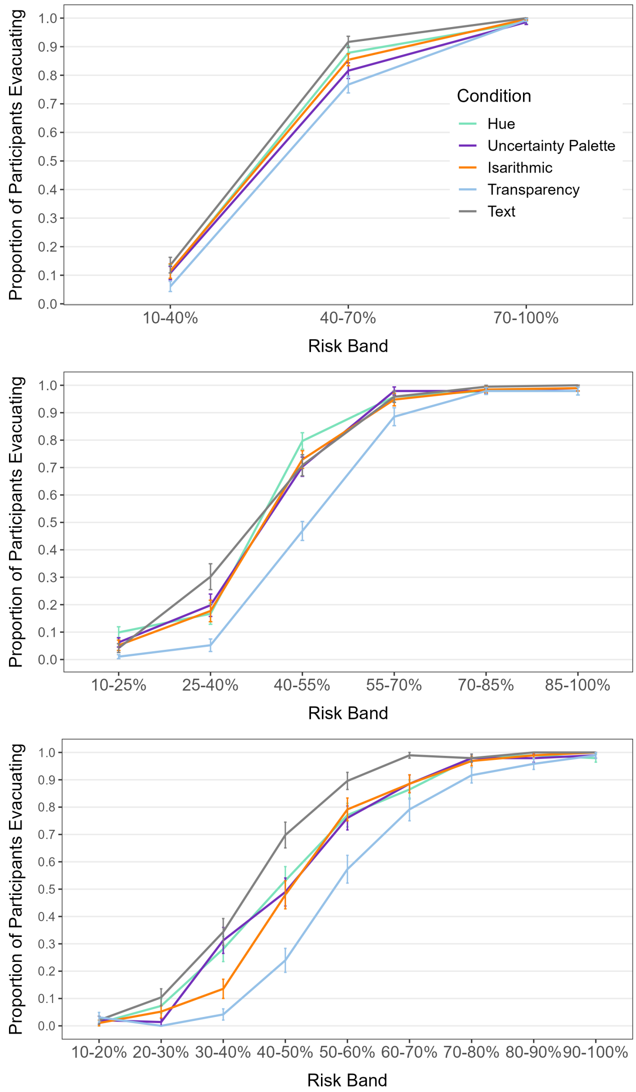

Figure 4: *Evacuation rates for each condition and risk band in Experiment 1.*

3.2. Why Did the Transparency Condition Have Lower Evacuation Rates?

The low evacuation rates for the transparency condition were unexpected. The experiments in [CBK*16] found numerically higher evacuation rates for transparency maps relative to hue maps. Yet the evacuation rates for the transparency maps in our study were considerably lower than in any other condition. Transparency is widely used to convey uncertainty in geospatial data visualizations [KMS14]. If this type of encoding can lead viewers to have different perceptions of risk, and therefore different patterns of decisions relative to other encodings of the same information, that has important implications for the design of geospatial visualizations that convey information about risk.

To better understand the reasons for our unexpected finding in Experiment 1, we designed Experiments 2 and 3. We speculated that the low evacuation rates for the transparency condition could have been caused by two factors: the color used in the transparency maps and/or the visual salience of the boundaries between the risk bands in that condition.

The transparency maps in Experiment 1 varied the transparency of a shade of blue. This hue was chosen somewhat arbitrarily, but was similar to color schemes that have been used in several prior studies of decision making with uncertainty visualizations [PRCR17, MHBG16]. However, blue may be a poor choice for conveying fire risk. Prior work on the use of color in visualizations has found that viewers have biases related to mappings between colors and concepts [LFK*13, SSM23]. A mismatch between the two can be detrimental to comprehension and performance. In this case, the mismatch between the blue hue and the type of risk being conveyed may have made the risk seem lower to the participants, biasing their decisions. Experiment 2 was designed to test the impact of hue (a fiery hue like orange or a non-fiery hue like blue) on the participants' decisions.

A second difference between the transparency maps and the other visualization conditions was the salience of the boundaries between the risk bands. Prior work on human comprehension of data visualizations has demonstrated that people tend to treat things inside of a visual boundary as being categorically different from things outside of the boundary, even when that is not actually the case [PRCR17, NS12, PCH21]. In the hue and isarithmic maps, the boundaries between risk bands were more visually salient than those in the transparency maps, particularly at higher levels of specificity. This may have helped to emphasize which risk band the house was in. Without these salient boundaries, the viewers may have treated the different risk bands as if they were more similar to one another. In addition, the lack of salient boundaries may have made the outer risk bands in the nine-band transparency maps difficult for participants to see. They may have had a harder time determining which risk band their house was in, and thus underestimated the risk. Experiment 3 was designed to test this possibility, by comparing transparency and isarithmic maps with and without salient visual boundaries between the risk bands.

4. Experiment 2: Testing the Influence of Hue

Experiment 2 used the same structure and procedure as Experiment 1. There were four visualization conditions: orange or blue isarithmic maps and orange or blue transparency maps. As before, the maps had three levels of specificity. The orange isarithmic and blue transparency maps were identical to those used in Experiment 1 (Figure 1). Examples of the blue isarithmic and orange transparency maps created for Experiment 2 are shown in Figure 5. There were 72 participants in Experiment 2, with 18 in each of the four counterbalancing lists.

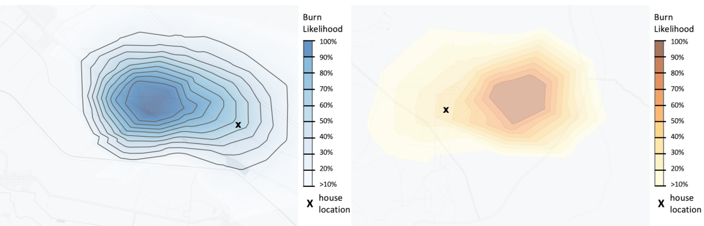

Figure 5: *Examples of the nine-band blue isarithmic and orange transparency maps used in Experiment 2. Additional examples can be seen in the Supplemental Materials.*

If the mismatch between the blue hue and the fire risk being conveyed was the primary cause of the low evacuation rates seen for the transparency condition in Experiment 1, we would expect to see lower evacuation rates for the two conditions using blue color schemes relative to the two conditions using orange color schemes. However, if the primary cause was the use of transparency to encode risk, we would expect to see lower evacuation rates for the transparency maps relative to the isarithmic maps, regardless of hue.

4.1. Results

Our first analysis compared the text and visualization conditions and generally replicated the findings of Experiment 1: there were significant main effects of information type and specificity as well as a significant interaction. Once again, the average evacuation rates for the visualizations decreased as the specificity increased. In the high-specificity condition, the evacuation rates for visualizations were significantly lower than for text. The details of this analysis can be found in the Supplemental Materials.

As in Experiment 1, our next step was to compare the evacuation rates for each visualization condition at each level of specificity. The results are shown in Figure 6. We used two-way repeated measures ANOVAs to test the effects of condition (text, isarithmic blue, isarithmic orange, transparency blue, and transparency orange) and risk band for each level of specificity. A summary of the results follows; the detailed results of the statistical tests are presented in the Supplemental Materials.

For the low-specificity condition, there was a significant main effect of risk band, but there was not a significant main effect of condition nor a significant interaction.

For the medium-specificity condition, there were significant main effects of condition and risk band as well as a significant interaction. Post-hoc tests with Bonferroni correction were used to compare the evacuation rates for each condition in each risk band. The only significant differences were in the 40-55% band. Both transparency conditions had significantly lower evacuation rates than the text condition, and the orange transparency maps had significantly lower evacuation rates than the orange isarithmic maps.

For the high-specificity condition, there were also significant main effects of condition and risk band, as well as a significant interaction. The evacuation rates for the two transparency conditions were very similar and were never significantly different from one another. Similarly, there were no significant differences in the evacuation rates between the two isarithmic conditions, although the evacuation rates were numerically lower for the blue isarithmic maps in the intermediate risk bands. Overall, the color scheme used had relatively little impact on the participants' decisions. In contrast, the use of transparency to encode risk had a substantial impact on the participants' decisions. The transparency maps generally had lower evacuation rates than the other conditions. Both transparency conditions had significantly lower evacuation rates than the text and isarithmic orange conditions beginning with the 40-50% risk band and continuing through the 60-70% risk band.

5. Experiment 3: Visual Boundaries

The results of Experiment 2 indicate that the choice of hue was not the driving factor behind the surprising results for the transparency conditions. The evacuation rates remained lower for the transparency maps even when a hue that was more semantically congruous with the fire risk was used. Changing the hue used in the isarithmic maps to a non-fiery color led to a small decrease in evacuation rates for some conditions, but this decrease was not statistically significant.

In Experiment 3, we tested an alternative hypothesis: that the

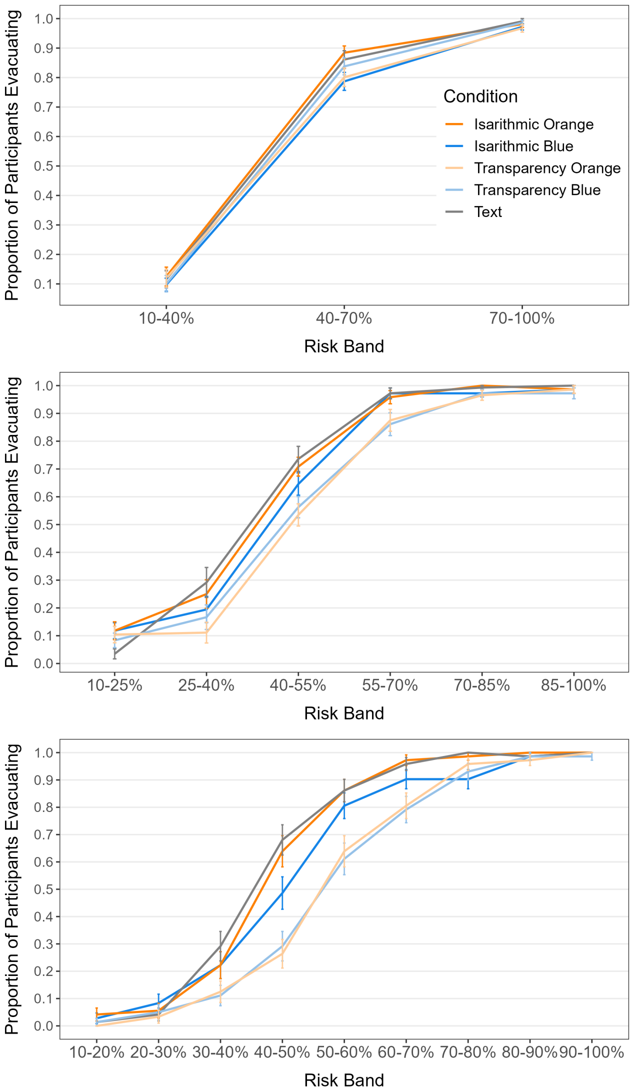

Figure 6: *Evacuation rates for each condition and risk band in Experiment 2.*

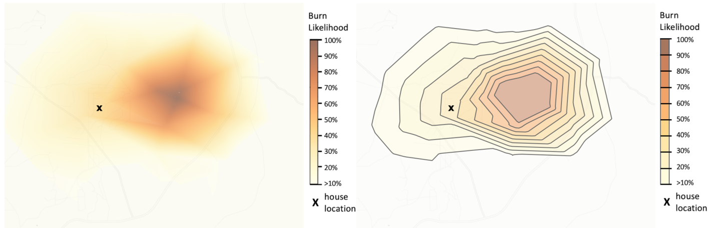

Figure 7: *Examples of the new stimuli created for Experiment 3: isarithmic maps with no contour lines (i.e., gradient maps) and transparency maps with borders around each band.*

key driver of the low evacuation rates for the transparency conditions was the lack of salient visual boundaries between the risk bands. Experiment 3 had the same structure as the prior experiments and used four visualization conditions. All four conditions used an orange color scheme. The conditions consisted of isarithmic maps (identical to those in Experiment 1), transparency maps (identical to those in Experiment 2), isarithmic maps without contour lines, and transparency maps with black borders added to the edge of each risk band. Examples of the stimuli are shown in Figure 7. Technically, the isarithmic maps without contour lines were no longer isarithmic maps; they were simply a gradient of color value. The gradient maps were identical across the different specificity conditions, with the exception that the legend was divided into a different number of bands. Examples of the full range of stimuli are shown in the Supplemental Materials.

A total of 80 participants completed Experiment 3, with 20 in each counterbalancing list.

5.1. Results

As in the earlier experiments, our first analysis compared the text stimuli to the combined visualization conditions. Once again, we found significant main effects of information type and specificity and a significant interaction. The results replicated the patterns observed in the prior experiments. The participants had higher evacuation rates for text than for visualizations and the evacuation rates for the visualizations were significantly lower in the high-specificity condition than in the low-specificity condition. The details of this analysis can be found in the Supplemental Materials.

The comparisons between the conditions across each level of risk and specificity are shown in Figure 8. Two-way repeated measures ANOVAs were used to test the effects of condition (text, isarithmic maps, isarithmic maps without lines, transparency maps, and transparency maps with lines) and risk band for each level of specificity. For all three specificity conditions, there were significant main effects of condition and risk band as well as a significant interaction between the two. Post-hoc tests with Bonferroni correction were used to compare the evacuation rates for each condition in each risk band. A summary of notable results follows; the detailed results of the statistical tests are presented in the Supplemental Materials.

In the low-specificity condition, the evacuation rates for the 10-40% risk band were significantly lower for both transparency conditions than for the text condition. In the 40-70% risk band, the evacuation rates for the isarithmic maps without contours were significantly lower than the standard isarithmic maps and the text stimuli. There were no other significant differences.

For the medium-specificity condition, the evacuation rates for the text stimuli and the two types of isarithmic maps did not differ significantly in any risk band. Similarly there were not significant differences between the two types of transparency maps in any risk band. However, the evacuation rates for the transparency maps without lines were significantly lower than the text and both isarithmic conditions in the 25-40% and 40-55% risk bands. In those same risk bands, the evacuation rates for the transparency maps *with* boundary lines were significantly lower than the text condition and the standard isarithmic maps, but they did not differ significantly from the isarithmic maps without lines.

In the high-specificity condition, there were no significant differences between the two types of isarithmic maps, although the evacuation rates were numerically lower for the isarithmic maps without lines in the intermediate risk bands. Similarly, there were no significant differences between the two types of transparency maps, although the evacuation rates were numerically lower for the version without boundary lines. The transparency maps without salient boundary lines had significantly lower evacuation rates than the text and the standard isarithmic maps for all of the risk bands from 30-40% to 70-80%. They also had significantly lower evacuation rates than the isarithmic maps without lines for the risk bands from 30-40% to 50-60%. The transparency maps *with* boundary lines had significantly lower evacuation rates than the text stimuli and both types of isarithmic maps in the 30-40% risk band, than the text and standard isarithmic maps in the 40-50% risk band, and than the text in the 50-60% and 60-70% risk bands.

Experiment 3 replicated our finding that transparency maps led to dramatically lower evacuation rates than text or isarithmic maps, especially in the high-specificity condition. Adding salient boundary lines to the transparency maps increased the evacuation rates somewhat, but the increase was not significant after Bonferroni correction. Removing the contour lines from the isarithmic maps decreased the evacuation rates slightly, but this decrease was also not significant. The results suggest that the different patterns of decision making for the isarithmic and transparency maps are not solely due to the salience of the boundaries between the risk bands.

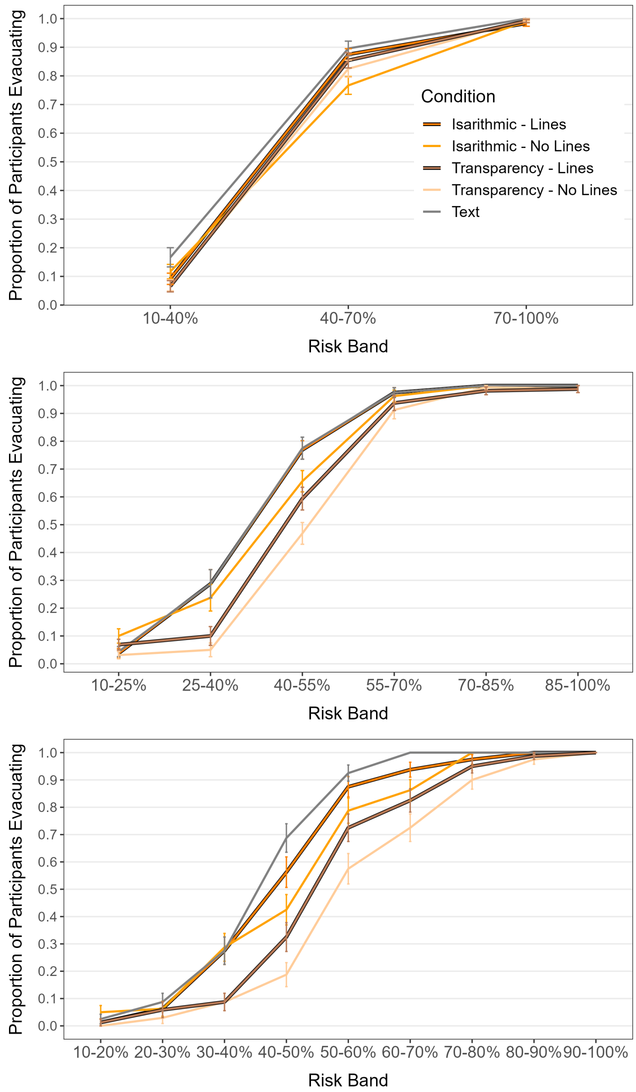

Figure 8: *Evacuation rates for each condition and risk band in Experiment 3.*

6. Discussion

Across three experiments, we found higher evacuation rates for text relative to visualizations of risk in a wildfire evacuation task. This was consistent with the findings of [CBK*16]. The results indicated that the participants used the contextual information provided by the geospatial representations of uncertainty to modulate their evacuation decisions. In the low-specificity conditions where the risk bands were wide, the participants adjusted their decisions based on the relative location of the house within the band. This kind of modulation was not supported by the text stimuli. For text stimuli, the participants appeared to base their decisions on the upper end of the probability range, leading to higher average evacuation rates.

We expected that the differences between the text and the visualizations would decrease as the specificity of the information about risk increased. At the highest level of specificity, the probability bands were quite narrow and the participants could not make additional inferences about the risk to the house based on its position within the band. Contrary to our expectations, the difference in evacuation rates became larger as the specificity increased. This effect was replicated in all three experiments, as shown in Figure 9. As the specificity increased, the average evacuation rates for the visualizations decreased while the average evacuation rates for the text stimuli remained relatively stable.

Our analysis of the different visualization conditions showed that this effect was largely driven by the transparency maps, which had substantially lower evacuation rates in the high-specificity condition than the other visual encodings. Experiment 2 showed that this effect was not an artifact of the hue chosen for the transparency maps. Experiment 3 showed that the salience of the boundaries between the risk bands had some impact, but did not fully account for the low evacuation rates for the transparency maps.

The reason for the lower evacuation rates in the transparency map conditions remains unclear. It is possible that the participants had a harder time determining which risk band their house was in with the transparency maps. At the highest level of specificity, the differences between the gradations of transparency were subtle and the contrast with the background was low. The participants may have struggled to perceive the boundaries between the bands or

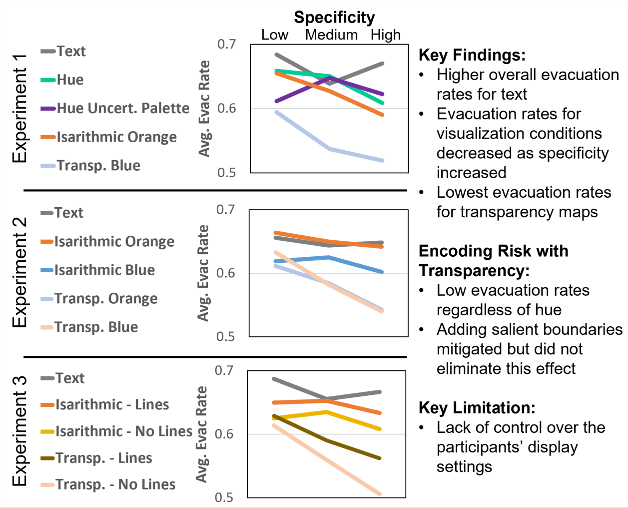

Figure 9: *A summary of the results of all three experiments. The plots show the average overall evacuation rates for each condition.*

struggled to interpret the legend. However, we would expect the addition of border lines to the transparency maps and legends to alleviate this problem. The edges of the bands for the transparency maps with border lines and the isarithmic maps should be equally easy to perceive and the legends equally easy to read. Yet the participants still had lower evacuation rates for the transparency maps.

Prior psychophysics research has found that human perception of transparency is inaccurate, with the visual system using lightness to make an approximate estimate of transparency [Bec85]. This leads to consistent biases in how people perceive transparency in images [SA02]. Furthermore, perception of transparency is influenced by numerous other visual features, including color, contrast, and the nature of the background surface [SA02, WPH17]. Research on color in visualizations has shown that people tend to infer that darker and more opaque colors map to larger quantities [SGS*18]. In cartography, contrast is effective for encoding quantitative differences [Bre94]. In our study, the lightness, transparency, and lack of contrast in the lower risk bands could have biased participants to underestimate the risk, even in the conditions with salient borders between the risk bands. The influence of these factors on risk perception warrants further research.

6.1. Limitations and Future Directions

One of the key limitations of this study is that the risk was not real. Although there was a small amount of money at stake for each decision, this game-like task is very different from making a decision about evacuation in real life. Despite this limitation, it is important to conduct tightly-controlled studies that systematically manipulate how uncertain information is presented. This will help to advance the scientific understanding of visualization cognition, which can then be extended to real-world contexts.

Another important limitation is that the data was collected online, so we had no control over the brightness and contrast settings on the participants' displays or the lighting of their environment, all of which impact perception [FBFS08]. Variation in screen settings could have disproportionately impacted the transparency stimuli, particularly for the high-specificity condition. Although uncontrolled display settings are realistic when communicating risk to the general public, additional research with in-lab studies and controlled display settings will be needed to better understand the effects observed in this study. Human perception of transparency is complex, yet this complexity has received relatively little attention in data visualization research [Che11, WPH17]. Additional systematic experiments are needed to assess the relationships among transparency, hue, and lightness in visualizations, determining how these cues interact with risk perception and decision making.

Another limitation of our study is that there is a high correlation between the risk level and the distance from the center of the fire. While this is also realistic, it is possible that some participants simply looked at the distance between the house and the center of the hazard map and ignored the rest of the information provided by the risk maps. We counterbalanced the stimuli so that every map appeared equally often in every visualization condition, so the relationship between distance and risk was the same for all conditions. However, the distance may have impacted the participants' decisions differently in the different visualization conditions. For example, the presence or absence of salient visual boundaries between the house and the center of the fire may have impacted how the participants perceived the distance and/or the risk. In the Supplemental Materials we present a model showing how the distance influenced the participants' decisions. In future work, it would be useful to test stimuli in which the risk level and distance are manipulated independently to further explore these effects.

6.2. Conclusions

Our findings suggest that the participants perceived the wildfire risk to be lower when transparency was used to encode the risk level. This resulted in consistently lower evacuation rates relative to other visual encodings. The reason for this difference remains unclear. However, it is important to develop a fuller understanding of this effect because transparency is widely used to encode uncertain information in geospatial visualizations and is often preferred over other encodings [KMS14, Dre02]. It is crucial for visualization designers to know if this type of encoding can lead to systematic differences in how viewers interpret risk. If this effect is pervasive, the use of transparency could have unintended consequences for risk perception and subsequent decision making.

7. Acknowledgments

The authors gratefully acknowledge Jamie Coram for assistance with the stimulus creation. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

[BAF08] Bostrom A., Anselin L., Farris J.: Visualizing seismic risk and uncertainty: A review of related research. *Annals of the New York Academy of Sciences 1128*, 1 (2008), 29–40. 1

[BCJ*11] Bisantz A., Cao D., Jenkins M., Pennathur P., Farry M., Roth E., Potter S., Pfautz J.: Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making 5*, 3 (2011), 277–293. 1

[Bec85] Beck J.: Perception of transparency in man and machine. *Computer vision, graphics, and image processing 31*, 2 (1985), 127–138. 10

[BMM05] Bisantz A., Marsiglio S., Munch J.: Displaying uncertainty: Investigating the effects of display format and specificity. *Human Factors: The Journal of the Human Factors and Ergonomics Society 47*, 4 (2005), 777–796. 1, 2

[Bre94] Brewer C. A.: Color use guidelines for mapping. *Visualization in modern cartography 1994*, 123-148 (1994), 7. 10

[car] Carto. https://carto.com/. Accessed: 2022-02-21. 3

[CBK*16] Cheong L., Bleisch S., Kealy A., Tolhurst K., Wilkening T., Duckham M.: Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science 30*, 7 (2016), 1377–1404. 1, 2, 3, 4, 5, 6, 9

[Che11] Cheung B. C.-k.: Using transparency in visualization. 10

[CMH18] Correll M., Moritz D., Heer J.: Value-suppressing uncertainty palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, Association for Computing Machinery, p. 1–11. URL: https://doi.org/10.1145/3173574.3174216, doi:10.1145/3173574.3174216. 4

[DFH*11] Dohmen T., Falk A., Huffman D., Sunde U., Schupp J., Wagner G.: Individual risk attitutdes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*, 3 (2011), 522–550. 2

[Dre02] Drecki I.: Visualisation of uncertainty in geographical data. *Spatial data quality* (2002), 140–159. 10

[FBFS08] Fetterly K. A., Blume H. R., Flynn M. J., Samei E.: Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays. *Journal of Digital Imaging 21* (2008), 193–207. 10

[FPS*21] Franconeri S., Padilla L., Shah P., Zacks J., Hullman J.: The science of visual data communication: What works? *Psychological Science in the Public Interest* 22, 3 (2021), 110–161. 1

[FZFU*07] Fagerlin A., Zikmund-Fisher B., Ubel P., Jankovic A., Derry H., Smith D.: Measuring numeracy without a math test: development of the subjective numeracy scale. *Medical Decision Making: An International Journal of the Society for Medical Decision Making 27*, 5 (2007), 672–680. 2

[Heg11] Hegarty M.: The cognitive science of visual-spatial displays: Implications for design. *Topics in cognitive science 3*, 3 (2011), 446–474. 1

[HQC*18] Hullman J., Qiao X., Correll M., Kale A., Kay M.: In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2018), 903–913. 1

[Hun07] Hunter J. D.: Matplotlib: A 2d graphics environment. *Computing in Science & Engineering 9*, 3 (2007), 90–95. doi:10.1109/MCSE.2007.55. 4

[JL13] Joslyn S., LeClerc J.: Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science 22*, 4 (2013), 308–315. 1

[KMS14] Kinkeldey C., MacEachren A. M., Schiewe J.: How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal 51*, 4 (2014), 372–386. 4, 6, 10

[LFK*13] Lin S., Fortuna J., Kulkarni C., Stone M., Heer J.: Selecting semantically-resonant colors for data visualization. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 401–410. 6

[Lin22] Lingg C.: Static map. https://github.com/komoot/staticmap, 2022. 3

[MHBG16] McKenzie G., Hegarty M., Barrett T., Goodchild M.: Assessing the effectiveness of different visualizations for judgments of positional uncertainty. *International Journal of Geographical Information Science 30*, 2 (2016), 221–239. 6

[MHTD23] Matzen L. E., Howell B. C., Trumbo M. C. S., Divis K. M.: Numerical and visual representations of uncertainty lead to different patterns of decision making. *IEEE Computer Graphics and Applications 43*, 5 (2023), 72–82. doi:10.1109/MCG.2023.3299875. 1, 2, 3

[NS12] Newman G. E., Scholl B. J.: Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias. *Psychonomic Bulletin and Review 19*, 4 (Aug. 2012), 601–607. 7

[OJGW19] Okan Y., Janssen E., Galesic M., Waters E.: Using the short graph literacy scale to predict precursors of health behavior change. *Medical Decision Making 39*, 3 (2019), 183–195. 2

[PCH21] Padilla L., Castro S. C., Hosseinpour H.: A review of uncertainty visualization errors: Working memory as an explanatory theory. *Psychology of Learning and Motivation 74* (2021), 275–315. 7

[PCRHS18] Padilla L. M., Creem-Regehr S. H., Hegarty M., Stefanucci J. K.: Decision making with visualizations: a cognitive framework across disciplines. *Cognitive research: principles and implications 3*, 1 (2018), 1–25. 1, 2

[PKH23] Padilla L., Kay M., Hullman J.: Uncertainty visualization. In *Wiley StatsRef: Statistics Reference Online*, Balakrishnan N., Colton T., Everitt B., Piegorsch W., Ruggeri F., Teugels) J., (Eds.). 2023. URL: https://doi.org/10.1002/9781118445112.stat08296, doi:10.1002/9781118445112.stat08296. 1

[PQMCR16] Padilla L., Quinan P. S., Meyer M., Creem-Regehr S. H.: Evaluating the impact of binning 2d scalar fields. *IEEE transactions on visualization and computer graphics 23*, 1 (2016), 431–440. 2

[PRCR17] Padilla L. M., Ruginski I. T., Creem-Regehr S. H.: Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive research: principles and implications 2* (2017), 1–16. 1, 6, 7

[SA02] Singh M., Anderson B. L.: Toward a perceptual theory of transparency. *Psychological review 109*, 3 (2002), 492. 10

[Sch97] Schwartz L.: The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine 127*, 11 (1997), 966. 2

[SCRP12] Szrek H., Chao L.-W., Ramlagan S., Peltzer K.: Predicting (un) healthy behavior: A comparison of risk-taking propensity measures. *Judgment and decision making 7*, 6 (2012), 716–727. 2

[SGS*18] Schloss K. B., Gramazio C. C., Silverman A. T., Parker M. L., Wang A. S.: Mapping color to meaning in colormap data visualizations. *IEEE transactions on visualization and computer graphics 25*, 1 (2018), 810–819. 10

[SS05] Schirillo J. A., Stone E. R.: The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. *Risk Analysis: An International Journal 25*, 3 (2005), 555–566. 1, 2

[SSM23] Schloss K. B., Schoenlein M. A., Mukherjee K.: *Color Semantics for Visual Communication*. Springer International Publishing, Cham, 2023, pp. 3–37. URL: https://doi.org/10.1007/978-3-031-34738-2_1, doi:10.1007/978-3-031-34738-2_1. 6

[TLG15] THOMPSON M., LINDSAY J., GAILLARD J.: The influence of probabilistic volcanic hazard map properties on hazard communication. *J Appl*, l. 4, 6 (2015). 4

[Tve13] TVERSKY B.: Visualizing thought. In *Handbook of human centric visualization*. Springer, 2013, pp. 3–40. 2

[Vel22] VELARDE V.: Cartodb basemap styles. `https://github.com/CartoDB/basemap-styles`, 2022. 3

[WPH17] WANG C., POULIOT J., HUBERT F.: How users perceive transparency in the 3d visualization of cadastre: testing its usability in an online questionnaire. *GeoInformatica 21* (2017), 599–618. 10