https://doi.org/10.2352/EI.2024.36.11.HVEI-206

Entropy Ordered Shapes as Bivariate Glyphs

Nicolas S. Holliman, Arzu Çöltekin, Sara J. Fernstad, Lucy McLaughlin, Michael D. Simpson, and Andrew J. Woods

Abstract

The natural ordering of shapes is not historically used in visualization applications. It could be helpful to show if an order exists among shapes, as this would provide an additional visual channel for presenting ordered bivariate data. Objective—we rigorously evaluate the use of visual entropy allowing us to construct an ordered scale of shape glyphs. Method—we evaluate the visual entropy glyphs in replicated trials online and at two different global locations. Results—an exact binomial analysis of a pairwise comparison of the glyphs showed a majority of participants ($n = 87$) ordered the glyphs as predicted by the visual entropy score with large effect size. In a further signal detection experiment participants ($n = 15$) were able to find glyphs representing uncertainty with high sensitivity and low error rates. Conclusion—Visual entropy predicts shape order and provides a visual channel with the potential to support ordered bivariate data.

Introduction

Glyphs are graphical representations used in visualization for presenting data that are considered clear and easy to read, as reviewed by [5]. In early work [1] identified how different visual channels (retinal variables) can be used to represent information, providing choices for encoding data in glyph designs, for example using size, color or lightness. There is however an evidence gap for the utility of commonly used visual channels, and there are only few cases where a theoretical prediction of glyph perception has been empirically validated.

One visual channel where there is currently a lack of theoretical prediction and confirmatory empirical evidence is the use of shape in glyphs. In fact, it is frequently stated that shape is not an ordered visual channel [1, 5, 32]. In this article we reconsider the perceptual order of shape and demonstrate how visual entropy can be used as a theoretical tool for predicting the outcome of empirical tests evaluating the perception of order between different shapes. Our aim is to extend the toolbox of proven methods for visualization to supplement, rather than replace, existing approaches.

Our initial motivation for this work was solving a problem in an applied science project where the representation of urban sensor data from a network of Internet-Of-Things (IoT) sensors across Newcastle-upon-Tyne was a key goal, [22]. A small subset of these data are shown in Fig. 1. The glyphs show the current hourly mean temperature for each sensor, but a representation for the uncertainty of the measure was requested. Specifically, we wanted to show the variance in the data, as this would help demonstrate which sensors could be most relied upon. This request led us to consider how to develop bivariate glyphs capable of displaying both mean temperature value and its variance together.

We hypothesise that visual entropy should relate to visual complexity, so that low visual entropy describes smooth visual signals and high visual entropy describes complex, more disordered, visual signals. This allows us to begin to define a scale of visual entropy, and to implement this, we introduce one way to quantify visual entropy using existing mathematical tools. Specifically, we address three research questions:

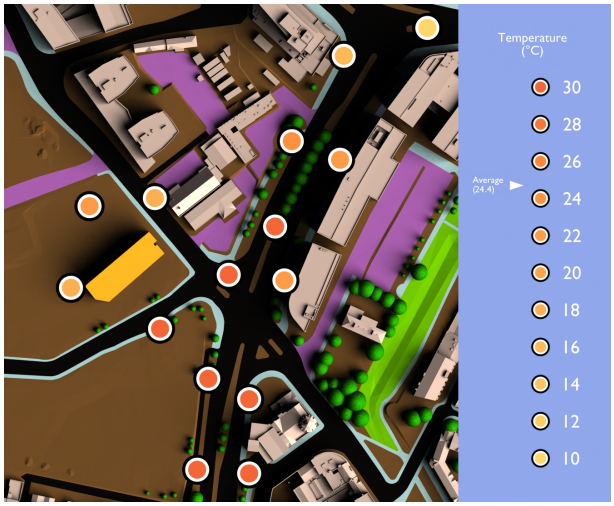

Figure 1. *Target style glyphs representing sensor values at locations in central Newcastle-upon-Tyne, the color represents the mean temperature from the last hour of readings for one sensor.*

1. Can we use theoretical visual entropy as a measure of shape complexity that predicts the human ranking order of shapes?
2. Can we demonstrate empirically through experiments that shape is an ordered visual channel?
3. Can we use visual entropy glyphs in visual search tasks?

Background

Our use case centers on the representation of uncertainty but the proposed approach has the potential be used for other bivariate data visualizations, for example including value and rate of change, or, for representing two independent variables such as temperature and NOX levels from the same sensor.

Information Theory and Visualization

A number of articles have been published on the relationship between information theory and visualization [49, 40, 10]. In Fig. 2 we illustrate how Shannon's communication pipeline ([43]) could map to the visualization pipeline. Encoding can be modelled as a process of image generation, communication as the optical path from display to retina and decoding as the process the brain uses to comprehend information encoded in the relayed image. We tend to agree with [26] that information theory is a weak match as a model for the human (neural) part of this pipeline because the human brain does not act as an ideal decoder of visual

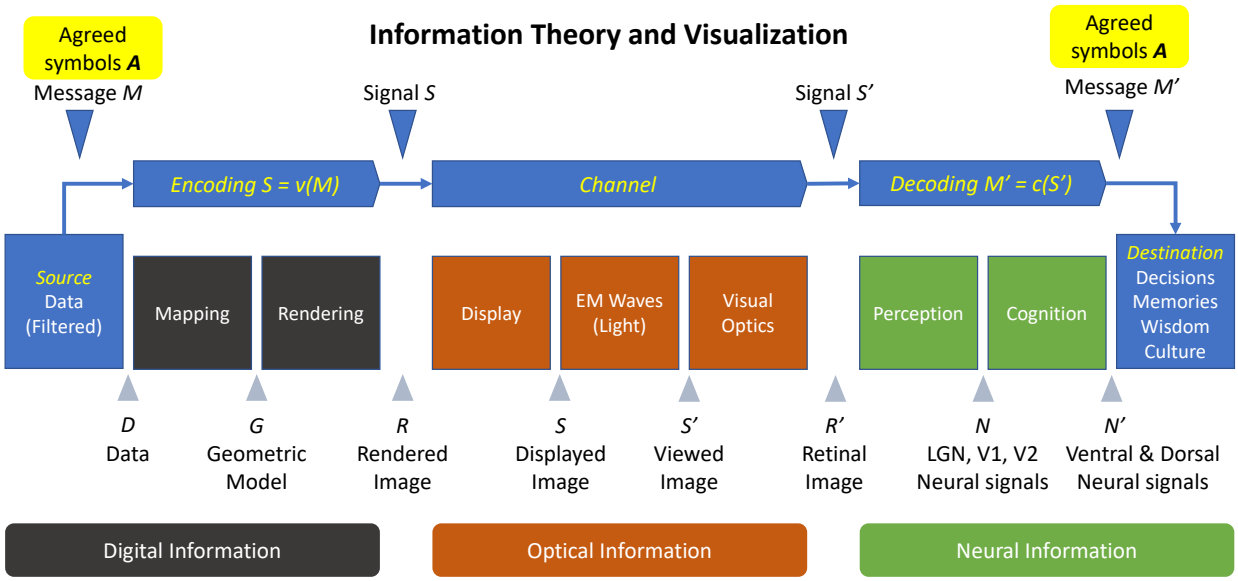

Figure 2. *The visualization pipeline mapped to an information theoretic communications pipeline as if the brain is an information theoretic receiver of visually coded messages M coded from an alphabet of symbols A, information is transformed physically twice as it flows along the pipeline from digital to optical and then from optical to neural signals.*

codes in an information theoretic sense. Indeed, Shannon was explicit about this "semantic aspects of communication are irrelevant to the engineering problem" [43], and yet knowledge about perception and cognition is essential to the production of visualizations.

Recent work on understanding the link between the perceived and measured complexity of map visualizations suggests that psycho-physically informed models of visual complexity are able to predict subjective human rankings of complexity [42]. This emphasizes the importance of perceptual factors as well as information theoretic factors in visual cognition. While the human mind might not be an ideal decoder of information this evidence suggests it is sensitive to pure information content in a signal, in this sense it is a noisy decoder.

Uncertainty

While it is often seen as valuable to present uncertainty, it is much less often directly depicted in visualizations [23], possibly due to its complex nature. Uncertainty by definition captures our *lack of knowledge* about a value or outcome, most often expressed quantitatively, among the most widely used being Pearson's standard deviation, [28], and Fisher's variance, [17]. Uncertainty can be classified as aleatory or epistemic, depending on whether it is due to random variation or to unknown factors [33]. Semantically uncertainty is challenging, because a value that represents a lack of a property is not intuitively easy for non-experts and, at times experts to understand. One way to make the concept more accessible to humans is to rethink the language, i.e., when talking about uncertainty, it may be that a positive phrasing can be more helpful than negative, such as *degree of certainty, level of certainty, confidence level, accuracy* and *precision* [19]. There is some agreed standardization in relation to levels of uncertainty, and there are national and international standards for reporting measurement uncertainty from e.g., metrology laboratories, [46]. In addition, some weather forecasts provide a degree of (un)certainty data, for example the UK Met office provides statements on precipitation in a standard form "There is a 70% chance of rain.", in a defined time period, [31].

In this work, we aim to design and evaluate a clear visual solution to represent at least one type of uncertainty that does not rely on the viewer having statistical knowledge, but still conveys information about the uncertainty of a value which can be related back to the underlying statistical methods when needed. Perhaps the closest in concept to our aim here is the use of strings of asterisks as categorical significance codes in conventional statistical reporting and software such as R, [38].

Approaches to Uncertainty Visualization

Earlier surveys of uncertainty visualization (e.g., [35]) identify seven methods which we categorize into one of two basic approaches: Those modifying the scene directly and those adding annotations to indicate levels of uncertainty. The approach that is most effective is clearly application dependent.

Approaches to representing uncertainty visually often relate to summarizing the spread of values related to a measurement: dot plots, histograms, box plots [47], confidence intervals and probability distribution functions, [8] describe ways to do this. These often presume some basic statistical knowledge on the part of the users, and an ability to interpret meaning from a spread of values. An empirical study of glyph-based approaches is presented by MacEachren et al. (2012) where the authors claim that the often-used visual channels of lightness and fuzziness perform well on their own, [27].

There are rigorous studies, e.g., [3], highlighting that it is far from routine for visualizations to include uncertainty information even though it is fundamental to informed decision making. Contemporary workshops run by government agencies (e.g., [7]) have highlighted that even in critical operational planning situa-

tions there is a real difficulty in finding ways to convey uncertainty to high level decisions makers. It remains an open question how best to visualize uncertainty, [50, 34], particularly when a single glyph must represent both a variable's mean value and its uncertainty.

Visual Entropy

If we imagine the human brain to be an ideal Shannon decoder, then it should decode and respond to signals differently with differing levels of entropy. In practice, though, the brain is much more than a decoder of signals, it generates its own hypotheses about the world, taking decisions based on partial information and weighting information in highly non-linear ways [19]. What also seems clear is that it does not need to produce a realistic decoding of the world around it [45], it is instead very efficient in extracting and using just enough information to complete a task.

A consequence of the mind's ability to hypothesize novel ideas and impute additional information is that it does not obey the data processing inequality [13]. It can and does add information at the end of the visual pipeline, this observation leads to the conclusion that it is far from an ideal information theoretic processor in the sense of Shannon's theory.

However, here we propose that even if the brain is a noisy decoder of signals it still is a decoder of signals and will have a sensitivity to differing levels of entropy in signals. We hypothesize that we should be able to use levels of entropy as a visual cue in coding visual information, in the same way we already use varying color, brightness, size and other visual cues [1].

Following Shanon's definition of information entropy [43], we define visual entropy $V(X)$ as the average cost of coding visual symbols v_i from a visual alphabet which have a probability of appearing in the message of $p(v_i)$:

$$V(X) = \sum_{i=1}^{n} p(v_i) log_2 \left(\frac{1}{p(v_i)} \right) \quad (1)$$

The higher the visual entropy $V(X)$ the more information is contained in the visual message, the more visual complexity the message contains. This argument has close similarities to the definition of viewpoint entropy presented in [49], but we keep this theoretical definition more general, rather than incorporate specific concepts of cameras and polygons. The visual entropy of a message could also be viewed as a measure of how incompressible it is, in this sense a more complex visual message will need more coded information to be sent in the signal, it cannot be coded as a simple signal.

To transform this theoretical construct to a practically meaningful visualization cue, we will consider visual entropy as analogous to some extent to visual complexity. Visual signals with higher visual entropy have a higher visual information content, requiring more bits on average to be coded for lossless transmission. A smooth sine wave for example can be coded in fewer bits than a signal consisting of uniform white noise. Our use of the term complexity here relates to perceived visual complexity and signal incompressibility, rather than to the generation of complex phenomena from chance chaotic behavior.

To make practical use of visual entropy, we next discuss the design of visual glyphs that both represent data and its uncertainty. We consider how to practically measure visual entropy and propose an extended glyph design that uses visual entropy to represent uncertainty values. We then report an experiment testing our glyph designs to evaluate whether they can represent a scale of uncertainty. Finally, we test the glyphs in an environmentally valid application situation where we ask users to search for the most and least reliable sensors across a 3D map.

Glyphs for Urban IoT Data

In our previous research, we have implemented a number of glyph designs for representing urban environmental data in 3D city models [21, 22]. We currently use the glyph design in Fig. 3, also shown in Fig. 1, that, while located in a relevant position in 3D space, is presented to the viewer as a primarily 2D shape.

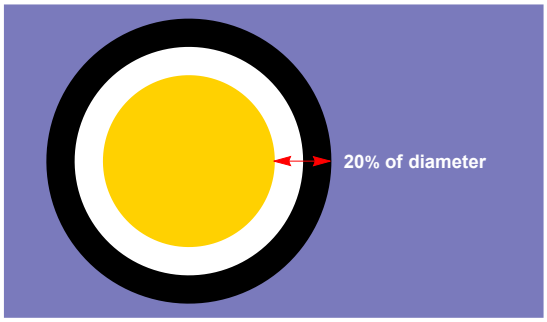

Figure 3. *The glyph design we adopted to represent a measurement from an individual IoT sensor, the two outer rings are designed to have a width of 20% of the total diameter, the colored central disc represent the sensors mean value on a predefined color scale.*

The design, shown in Fig. 3, took some aspects from that of a target of concentric rings and some from the design of the Landolt C optotypes [25]. The rings, a dark outer shape and a light inner shape, were chosen in order to highlight the glyph against both light and dark back-grounds. It also provides a level of self-contrast for the glyph. The total width of both outer rings is set to be 20% of the diameter of the whole target, matching that aspect of the Landolt optotype design.

We use the central disc to represent data value, and typically we use color to do this, following color scale standards set in the literature. In the examples here, the data value represented in the central disk is temperature and adopts the colors used by the UK Met Office [31]. In our visualizations, a color legend is usually displayed on or near the visualization.

The visualizations presented here use physically-based path tracing for the graphical rendering stage implemented with Blender Cycles [2]. Our goal is to use realistic lighting simulation to help engage viewers in the 3D image however we therefore need to make sure this realistic lighting does not alter the glyphs information carrying appearance. We do this by using flat shading of colors and rotating glyphs to face the camera.

Entropy as a Visual Cue

A question that was raised when presenting our urban data visualizations was *how much is it possible to rely on the sensor data?* Expert members of an audience are aware that different sensors can have very different accuracy and precision related in

part to their cost. To help answer this question in a visual form, we started to consider how we could represent the uncertainty of measurements at the same time as the value of measurements.

Our goal is to find a visual cue that we can use to represent variance alongside the color currently used in our glyphs for the mean value. We considered shape as a possible cue following the review in [53]. We also reviewed results from an earlier study [14] on the human perception of fractal shape where it was demonstrated that certain fractal generation parameters correlated well with perceptual ordering of perceived shape complexity. More recent work in visual search suggests the cues of size and frequency, which help form shape differences, are both reported to guide (direct) attention in visual search [52].

The subjective ordering of a set of shape glyphs was investigated in [11] alongside a range of other retinal variables detailed by [1]. Chung et al. [11] tested an overall impression of order in a given sequence of shapes and searchability of high and low-value glyphs. They did not, however, predict the perceived impression of order, nor confirm this with a pairwise study, nor did they extend their results to bivariate glyphs.

This previous work led us to hypothesise that varying the levels of visual entropy, as a measure of shape complexity, might be used as a visual cue, and ultimately allow shape to be used as a scale to represent ordinal categorical or interval numerical values. To implement and evaluate this hypothesis, we develop two things: a practical measure of visual entropy and a geometric representation for the glyphs that exhibits varying levels of visual entropy.

We start by considering how a visual signal $S = v(M)$ can be generated by a coding function v from an abstract message M. The message we encode need have no direct meaning, but for our purposes it does need to be able to represent variable levels of entropy. The signal is the geometric representation of the message that we will eventually render in our glyphs. This first step in the glyph generation is illustrated in Fig. 4.

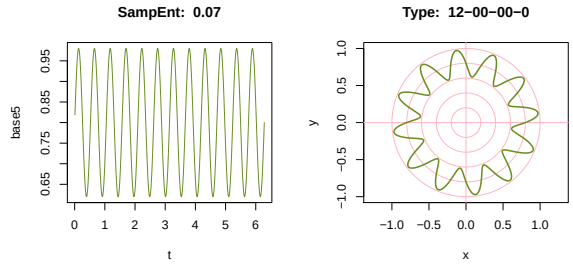

Figure 4. *The message M on the left is coded as the signal S on the right, by plotting the message in 360 degrees on a polar plot; we estimate the visual entropy of the glyph as the sample entropy of the message.*

To estimate the visual entropy of the message, we calculate the Sample Entropy [39] of the message before it is coded as a geometric shape. Sample Entropy provides an estimate for the regularity and unpredictability of a data series, it is often used for comparing time series such as electro-cardiograms. Sample entropy has low dependence on the series length and is consistent across series [54], hence we adopt it for our calculations. We use the sample entropy implementation in the R pracma package [4] with parameter settings of n=2,r=0.2 based on guidance from [54] to estimate the visual entropy in our glyphs.

We now have a route to create geometric shapes with measurably varying levels of visual entropy. To add these to our existing glyphs in Blender we export the signal shape from R and import it to Blender as an extruded polygon. These polygons are then used to replace the inner white disc in the glyph, as illustrated in Fig. 5.

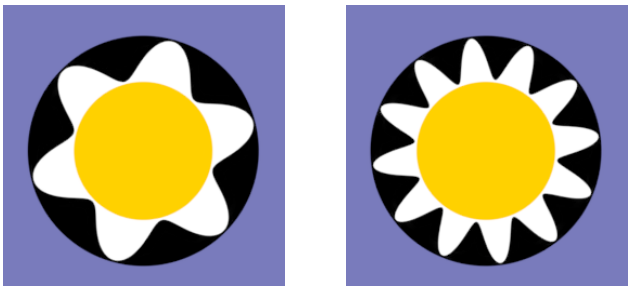

Figure 5. *Two glyphs created in Blender from signals with different sample entropy measures, the central temperature color mapping is the same in each glyph, we hypothesize the visual entropy of the surrounding shape can represent an orthogonal value such as uncertainty, or variance of the temperature value.*

The resulting designs (Fig. 5) provide an ability to create glyphs representing a value, such as mean temperature using color in the central disc and a second value, such as uncertainty in the variation of the surrounding shape. We next empirically evaluate whether viewers can naturally order glyphs of differing visual entropy.

Evaluating the Visual Entropy Glyphs

Based on pilot testing in our laboratory in Newcastle (UK), we choose to evaluate the glyphs shown in Fig. 6. This set of glyphs uses a single sine wave as the generating message, at varying frequencies. These are similar to radial frequency patterns which have been shown to be detectable at low amplitudes in the psychophysics literature [51]. Recent studies support that these are discriminable shapes based on frequency differences [15], are identifiable even if only the shape convexities are visible [41] and can represent numerical order [11]. In addition, in these trials we added a numerical indication of the temperature value on the glyphs, the same value of 13.5 degrees Celsius was used in all trials.

The increase in sine wave frequency from glyph to glyph was set on a geometric scale, the frequency doubling for each new additional glyph. This is similar to the logarithmic increase between levels on a logMAR visual acuity chart [24], but here the change increases at a significantly greater rate per level (2x rather than 1.26x) in part so that the full range of visual acuity is used in a smaller number of glyphs, and in part because this evaluation is designed to determine whether an ordering of visual entropy exists rather than to determine the level of just noticeable difference (JND) between glyphs. This anticipated that, like many aspects of human perception [16, 48], there would be a logarithmic response to the visual entropy stimulus.

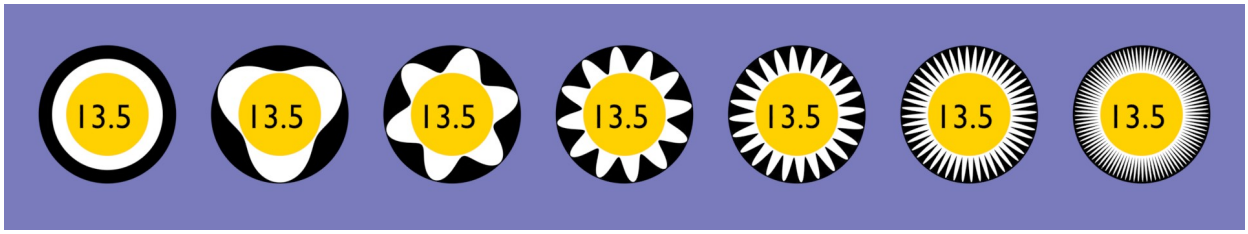

Figure 6. *The set of glyphs used in the user evaluation, ordered by the calculated visual entropy value of the generating message, as shown below each glyph, when viewed on the controlled test displays the glyph on the far right is displayed at or above the limit of human spatial acuity (10 cpd), in the results these images are referred to by the labels A to G from left to right.*

Experimental Method and Apparatus

The experiment to test the visual entropy glyphs has two objectives. First to test the hypothesis that there is a perceived rank ordering between the glyphs shown in Fig. 6 Second to demonstrate whether this order is predicted by our numerical measure of visual entropy.

To evaluate whether there is a rank ordering between the glyphs a two alternative forced choice (2AFC) method [20] of glyph image pair comparisons was implemented where all paired permutations of the set of glyphs are shown to each participant, except for those where the pairs would be the same shape. The experiment was replicated in different countries, in three different environments, two physical-in-person (Perth, Australia and Brugg-Windisch, Switzerland), and in addition, online using Mechanical Turk.

The stimulus presentation was implemented using the PsychoPy toolbox [36] and answers for each pair comparison, a left or right arrow key-press, were recorded in addition to the time taken to enter the answer. Each participant saw a different random order of pairs, determined by PsychoPy's random number generator. The stimuli for the in-person trials were presented at a viewing distance of approximately 500mm for all participants, differences between the environments are discussed below.

The Perth trials were presented on a 11.8" FHD LCD monitor built into a Lenovo laptop, the Swiss trials were presented on a 15.6" FHD LCD monitor built into a HP laptop, and the Mechanical Turk (MTurk) trials were presented remotely on unconstrained display devices. As a result of the unconstrained environment, we anticipated some variability in the outcomes of the MTurk trials and designed a set of data cleaning methods to detect outliers described below alongside the results.

The instructions for each participant read in the trial were:

```
You will see a series of image pairs.
Each image represents a value and also
represents a level of uncertainty.
More complex shapes represent more uncertainty.
Choose which image represents the most uncertain
value to you.
Left arrow for left. Right arrow for right.
```

Note, we considered the use of the word complex with some care, as it was clear that visual entropy would not be a widely understood description of shape differences in the images. Participants then began the trial where they were presented with all 42 pairwise permutations of the images in Fig. 6, this included reversed order image pairs.

Participants

Participants in the global trials gave consent for the data from the trials to be used and communicated worldwide by the investigators for the purposes of the study. We requested very limited personal data, whether participants had normal or corrected to normal visual acuity and whether they were aware of any color deficiency in their visual ability. While color deficiency was unlikely to affect the results in a shape complexity comparison experiment, we recorded this in case it had a bearing on the saliency of the central colored disc.

We used two physical locations, Perth, Western Australia ($n = 17$) and Brugg-Windisch, Switzerland ($n = 20$) and one set of online experiments with Mechanical Turk ($n = 50$). Our aim was to establish whether the three distinct locations produce broadly comparable results.

Omnibus Testing

Before evaluating the experimental hypothesis we apply a series of omnibus G-tests to check whether there are variations from our expected outcomes at the different locations. These omnibus tests allow us to understand; 1) if there is any variation from the expected proportions of correct answers within the glyph results at each location, 2) whether there is any significant variation across the replications at the three locations, 3) whether there is any unexpected variation in the data pooled across all three samples (AU, CH, MTurk) and finally 4) to use the additive property of the G-Test to check that overall the data fits our predictions. The G-test (compared to chi-squared and exact tests) is recommended for situations with larger numbers of observations [44] and when replicating studies using multiple location testing [30].

G-Test of Goodness of Fit by Location

Our first omnibus null hypothesis is that at each individual location the proportion of correct answers for each of the glyphs is the same, i.e. $1/7$ of the total number of trials. This is a reasonable null hypothesis since we designed the glyphs to be above the acuity threshold and to be clearly distinguishable on an exponential scale of increasing complexity. The alternative hypothesis is that there is a statistically significant difference in the proportions of correct answers for some of the glyphs at one or more locations, and if so this will need further analysis.

Using a G-Test for Goodness of Fit at each location we find the following; for Perth $(G = 1.5197, p = 0.9582, df = 6)$, for Swiss $(G = 0.6137, p = 0.9962, df = 6)$ and for MTurk $(G = 1.5289, p = 0.9576, df = 6)$. In each location there is no evidence (since all $p >> \alpha = 0.05$) for rejecting the null hypothesis and

therefore no evidence that there is any unexpected variation in the outcomes at each of three locations.

G-Test of Independence

The second omnibus null hypothesis tested is that the relative proportions of correct answers are the same across the three different locations. If the relative proportions are not the same we will need to analyse locations separately and will be unable to pool them.

This hypothesis is evaluated using a G-Test of Independence with a $3x7$ contingency table (3 locations, 7 glyphs). The result is $(G = 1.1246, p = 0.99997, df = 12)$. There is no evidence (since $p >> \alpha = 0.05$) for rejecting the null hypothesis and we can conclude we can now consider the pooled results.

G-Test for Goodness of Fit for Pooled Results

The third omnibus null hypothesis is again the equal probability hypothesis for each glyph but now we apply this for the pooled results combined across all three locations (n=87).

Using a G-Test for Goodness of Fit for the pooled results we find the result $(G = 2.5378, p = 0.8642, df = 6)$ and again we cannot reject the null hypothesis (since $p >> \alpha = 0.05$) and conclude that there is no evidence for any unexplained variation between the proportions of correct answers for the glyphs in the pooled results.

Outcome by total G-value

The final omnibus test checks the overall outcome is not inconsistent with our null hypothesis of equal probability. We use the additive property of the individual location G-Tests from step *1)*, sum these to give the total G-value $(G = 3.6623, df = 18)$, looking this up in a chi-square table gives a p-value $(p = 0.99988)$. Again we cannot reject the null hypothesis (since $p >> \alpha = 0.05$) and we conclude that overall there is no evidence of any unexpected variation in the overall outcome across glyphs and locations.

In summary, over all four omnibus tests we have found no statistically significant evidence to reject the null hypothesis, that the glyphs have equal probability of being correctly chosen, and we have found no evidence to suggest there are statistically significant differences between the replication locations. We can therefore move on to consider the pooled data as a whole in testing the main experimental hypothesis.

Analysis of the Pooled Results

The design goal for the visual entropy glyphs is to create a perceived order among the glyphs that can be predicted by the visual entropy calculation for the enclosing shape.

We use the pooled response data for each glyph, shown in Table 1, to test the hypothesis that participants agree with the predicted order. The independent variable is each glyph type (A, B, C, D, E, F, G) and the dependent variable is the proportion of correct choices made about the order of each glyph in the pairwise comparisons with all six other glyphs. We deem the response "correct" when a participant's choice is as predicted by the visual entropy calculation. With $n = 87$ participants the total number of trials per glyph is $87 * 6 = 522$, and this is therefore the maximum correct score each glyph could achieve.

Table 1: Pooled results for the glyph pairwise order comparisons.

glyph	success	trials	pvalue	probability	CI low	CI high
A	469	522	<0.001	0.898	0.869	0.923
B	459	522	<0.001	0.879	0.848	0.906
C	462	522	<0.001	0.885	0.855	0.911
D	452	522	<0.001	0.866	0.834	0.894
E	456	522	<0.001	0.874	0.842	0.901
F	444	522	<0.001	0.851	0.817	0.88
F	427	522	<0.001	0.818	0.782	0.85

The null hypothesis, $H0$, for this analysis is that participants will do no better than chance at choosing between glyphs in the order predicted by the visual entropy calculations ($probability <= 0.5$) . The alternative hypothesis, Ha, is that a majority of participants do agree with the predicted order ($probability > 0.5$).

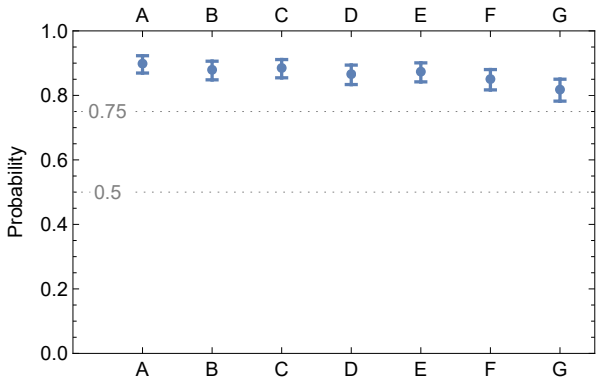

Figure 7. *Pooled results showing 95% confidence intervals, the null hypothesis, $H0$, is the $probability < 0.5$ that the majority of participants in agreement with our entropy prediction is no better than chance.*

As the results are a count of categorical choices, we apply an exact binomial test against a simple majority for each glyph. Because the test is repeated for each level (glyph) we apply a Bonferroni correction to the alpha threshold for the p-values and use $\alpha = 0.05/7 = 0.007$ as the significance level. The results of the binomial tests are given in Table 1 and illustrated in Fig. 7.

For all seven glyphs tested, the p-values are < 0.001 and as these are all smaller than α we can reject the null hypothesis of chance performance, and accept the experimental hypothesis that a majority of participants' choices agree with the predicted order.

Effect Size

One approach for estimating effect size for binomial tests, recommended in [37], is to calculate Cohen's g [12]. This method can be used to estimate the effect size for probabilities only in comparison to a chance result of 50% correct answers. Cohen's

g effect sizes are calculated for each of the glyphs as shown in Table 2 these results suggest that the effect size for all the glyphs is Large ($g > 0.25$).

As discussed in detail in [9] the practical significance of effect size depends on context. Here, given the high probability of users choosing the correct order, it seems reasonable for us to conclude the practical effect size will be strong. That is, very often users will judge the order of the glyphs as predicted, and this may be even more often if in practice we display a legend for the user to refer to.

Table 2: Effect size estimate using Cohen's g for each of the glyph binomial tests.

glyph	probability	g	effect
A	0.898	0.398	Large
B	0.879	0.379	Large
C	0.885	0.385	Large
D	0.866	0.366	Large
E	0.874	0.374	Large
F	0.851	0.351	Large
G	0.818	0.318	Large

range	effect size
0.00 <= g < 0.05	Negligable
0.05 <= g < 0.15	Small
0.15 <= g < 0.25	Medium
g >= 0.25	Large

Response Times (RT)

To evaluate the seven glyphs at the three different global locations, we had to allow the test environment to vary between the Perth, Switzerland and the online experiments. We anticipate this could introduce either random or systematic noise into the results. On the other hand, this is an ecologically valid testing context since in real-world visualization applications, the viewing environment will differ between users. We found no evidence for a statistically significant variation in accuracy (probability of choosing the predicted response) due to location or glyph type in the pairwise glyph order experiments reported above, and we now test the response times.

To select the appropriate omnibus tests, we tested for normality in the RT data using the Shapiro-Wilk method, across all locations this suggested ($W = 0.5578, p < 0.001$) we had to reject the null hypothesis and accept that the data is not normally distributed. This result also holds if we test the data separately at each location. Therefore we needed a non-parametric omnibus test and, as there is no direct equivalent for the two-way ANOVA for this experimental design, we used two separate one-way Kruskal-Wallis tests for the two independent variables location and glyphs.

RT omnibus testing

We considered the RT variation between the glyphs (seven levels) first using a one-way Kruskal-Wallis test. The result is ($\chi^2 = 0.24288, df = 6, p = 0.9997$) and as a result we cannot reject the null hypothesis $H0$ that the difference in the means of the RT between the glyphs is simply due to chance, with $p >> \alpha = 0.005$

We use the same type of non-parametric test to consider the overall effect of location (three levels) on the response times with the result ($\chi^2 = 79.428, df = 2, p < 0.001$). In this case we do reject the null hypothesis and accept the alternative hypothesis that there seems to be statistically significant variation between the response times in each location beyond that expected by chance.

Table 3: Post-hoc Dunn test results comparing the response time (RT) between all location pairs, in all cases the null hypothesis is rejected.

location	statistic	df	pvalue	alpha
MTK–PTH	−4.935	2	<0.001	0.0167
MTK–SWS	−8.476	2	<0.001	0.0167
PTH–SWS	−2.598	2	0.0094	0.0167

To understand more about the variation between locations we use the Dunn post-hoc test against the test criterion $\alpha = 0.05/3 = 0.01667$, results are shown in Table 3. In all three possible location pair comparisons the null hypothesis can be rejected suggesting the RT differ statistically significantly between each of the locations.

As noted above we anticipated some variation between locations in at least some of the results due to test environment differences, and this analysis does suggest that some uncontrolled factors, such as varying screen size (phone or laptop/desktop), response method (mouse click or touch), or the experimental context (in person vs remote) are effecting the response times by location.

Application Domain Testing

Our results above suggest that the participants perceive the visual entropy glyphs in an order that is in agreement with our visual entropy predictions. We should therefore be able to use them to represent ordered categorical information or quantized numerical data on interval or ratio scales. In this section we examine the use of visual entropy glyphs in a specific context. Our urban digital twin application, see Fig. 1, has a requirement that we display both a sensor's mean value and its variance, so that end users, for example policy makers, can see at a glance which sensors they can rely on most.

Previously, we displayed sensor data as the mean value over an hour using the MetOffice color scale for temperature, here we now also visualize uncertainty as the variance calculated over the same hour using visual entropy glyphs, see Fig. 8. To set the range of the uncertainty scale, we need to calculate the range of variance for sensors in view, potentially over the whole city, so that we can calculate the minimum and maximum values on this scale.

Experimental Method and Apparatus

To evaluate whether this representation can be effective, we designed a signal detection experiment that requires the participants to search for a glyph based on its level of uncertainty, as recommended as a follow on experiment by [6] and similar to the

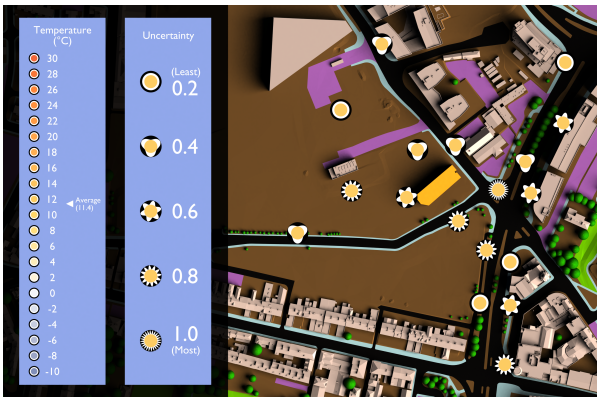

Figure 8. *The urban temperature data visualization showing both hourly mean temperature values using the MetOffice color scale and the variance of those values using our new visual entropy scale, this image is an example of the high uncertainty target-present stimulus used in the experiment described below.*

search task implemented by [11]. This was a target-present/target-absent visual search for either a low uncertainty target or a high uncertainty target. A total of fifteen participants (n=15), students and staff at Newcastle University (UK), each viewed forty images (therefore amounting to a total of 600 trials), searching for the least uncertainty glyph in ten target-present and ten target-absent images and the same again for the highest uncertainty glyph.

The display used for this experiment was a Microsoft Surface Pro 4, a display with 2736x1824 0.094mm square pixels at a nominal viewing distance of 500mm. Given this geometry, we calculated that the 24-cycle glyph was at the 10 cpd limit of human vision for this display, therefore we selected the set of five glyphs shown in Fig. 8. Once again, we used PsychoPy to present the stimulus to the participants. Ethics approval was granted, details of which were given earlier.

Our hypothesis in this study is that the target-present glyphs should be easy to find because of the choice of log-scale increments in the generating frequency and as a result the discriminability should be high. If this is case, we also hypothesize that there should be a response time difference between target-present and target-absent trials.

Results of the Domain Testing

We analyzed the outcome data in R using the psycho package [29] from a total of 300 responses per glyph type (highest and lowest uncertainty), of which 150 were target-present and 150 target-absent in each group. The confusion matrix for the low uncertainty glyph searches is shown in Fig. 9. As hypothesized, the low uncertainty glyph is easy to find in visual search tasks, with high discriminability, d', while the response bias, β is towards answering (n) target-absent.

The confusion matrix for the high uncertainty glyph in Fig. 10 gives similar results, with a slightly higher response bias towards target-absent.

Response times were analyzed using two-way within-subjects t-tests, the results are shown in Table 4. As was hypothesized, target-absent trials took significantly longer to complete (approximately twice as long) than target-present trials.

There was no statistically significant difference between

		Correct choice: Y	Correct choice: N
Participant choice	Y	148 (TP)	0 (FP)
	N	2 (FN)	150 (TN)

Low uncertainty glyph search.

Parametric indices
d' = 4.774
beta = 4.816

Figure 9. *Confusion matrix for the low uncertainty glyph visual search from 150 target-present and 150 target-absent trials. There is a slight response bias towards false negatives (target-absent when it is present).*

		Correct choice: Y	Correct choice: N
Participant choice	Y	142 (TP)	0 (FP)
	N	8 (FN)	150 (TN)

High uncertainty glyph search.

Parametric indices
d' = 4.276
beta = 11.856

Figure 10. *Confusion matrix for the high uncertainty glyph visual search from 150 target-present and 150 target-absent trials.*

mean glyph search times for low and high uncertainty glyphs in the target-absent condition, and nor did we expect one as the search task is essentially the same.

There was a statistically significant difference ($p < 0.05$) between target-present search times for the low- and high- uncertainty glyphs, with it taking on average 0.3 seconds longer to find the more complex, high uncertainty glyph. This was also supported by comments from some participants who reported the more complex glyph was harder to search for.

In summary, the results from the application domain testing seem in agreement with our general hypotheses from the earlier sections: Participants could, with low error rates, search for and successfully find glyphs with different levels of uncertainty. We also identified the possibility that glyphs with higher visual entropy (more complex shapes) are slower to search for than those with low visual entropy.

Conclusion

There are many visual representations of uncertainty in technical publications that work for statisticians and for technical audiences. It is harder to find good visual representations of uncertainty in the everyday media and in documents intended for

Table 4: Signal detection experiment mean response times, all differences are statistically significant ($p < 0.0125$ with Bonferroni correction), except between glyph absent conditions.

Response times(s)	Low glyph	High glyph	pvalue
Glyph present (y)	1.49	1.78	0.0019
Glyph absent (n)	3.26	3.1	0.28
pvalue	<0.001	<0.001	

non-technical high-level decision makers.

We set out an argument for the use of visual entropy as a visual coding scale for visually transmitted information. We hypothesised that even though the human brain is not an ideal information theoretic signal receiver, it should still be sensitive to varying levels of entropy in signals. Intuitively, we can consider entropy in this case to be analogous to visual complexity. We then set out to rigorously evaluate this approach by creating a set of glyphs and using those glyphs to represent uncertainty. We addressed three research questions:

- Can we use visual entropy as a measure of shape complexity that predicts the human ranking of simple and complex shapes? We demonstrated we were able to predict human ranking of glyphs, using sample entropy as a proxy for visual entropy, with high confidence.
- Can we demonstrate empirically through experiments that shape is an ordered visual channel? We demonstrated a natural ranking order among our proposed visual entropy glyphs allowing us to represent ordinal categorial, or numerical interval, data on a discrete scale.
- Can we use visual entropy glyphs in visual search tasks? We demonstrated users could successfully search for glyphs with predefined levels of uncertainty in an urban digital twin visualization of temperature sensors.

There is a similarity between our work and the ideas described in [18], where an informal argument for the use of wavelength and amplitude for representing uncertainty is made. Here we have provided a theoretical basis for this approach, a rigorous evaluation of glyph ranking and an initial application domain validation of the effectiveness of visual entropy glyphs.

We believe that visual entropy provides a useful concept with which to reason about glyph shape ordering. We have shown that we can measure it, predict human ranking of glyphs using this measure and apply these glyphs in a 3D visualization environment. We believe we have presented a rigorous novel approach for visualizing ordinal data that has potential application to bivariate data including uncertainty visualization. A Python library *Vizent* is available on Github that implements these ideas.

Ethical approval

Ethical approval was granted at Newcastle University, UK, ref:8775, to run a global experiment (inside and outside the EU) and was also independently requested at Curtin University, ref:181113, Perth, Australia. All participants agreed to a Newcastle University approved consent form.

Acknowledgments

The authors wish to thank: the Alan Turing Institute for funding under EPSRC grant EP/N510129/1 and for Professor Holliman's Turing Fellowship; Professor Jenny Read and Dr Kevin Wilson (Newcastle University) for their helpful insights; Northumbria VRV Studio for the VNG 3D model of Newcastle; the EPSRC UKRIC funded Urban Observatory at Newcastle for sensor data; the Curtin HIVE and members of the CIC at Curtin University, Perth, WA; Dr Ronni Bowman (DSTL) and Matt Butchers (KTN) for inspirational workshops on uncertainty visualization for high-level decision makers; all participants who took part in our experiments, and finally, Professor David Firth (Univ. of Warwick) for invaluable advice on statistical methods. A more detailed preprint of this article is available on *Arxiv*.

References

[1] J. Bertin and W. J. Berg, *Semiology of Graphics: diagrams, networks, maps*. University of Wisconsin press Madison, 1983.

[2] Blender Online Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2024. [Online]. Available: http://www.blender.org

[3] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz, "Overview and state-of-the-art of uncertainty visualization," in *Scientific Visualization*. Springer, 2014, pp. 3–27.

[4] H. W. Borchers, "Practical Numerical Math Functions [R package pracma version 2.2.2]." [Online]. Available: https://cran.r-project.org/web/packages/pracma/index.html

[5] R. Borgo, J. Kehrer, D. H. S. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications," in *Eurographics 2013 - State of the Art Reports*.

[6] N. Boukhelifa, A. Bezerianos, T. Isenberg, and J.-D. Fekete, "Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2769–2778, 2012, publisher: IEEE.

[7] V. Bowman, "Uncertainty Visualisation Study Group Briefing Notes," London, 2018.

[8] K. Brodlie, R. A. Osorio, and A. Lopes, "A review of uncertainty in data visualization," in *Expanding the frontiers of visual analytics and visualization*. Springer, 2012, pp. 81–109.

[9] P. Cairns, *Doing Better Statistics in Human-Computer Interaction*. Cambridge University Press, 2019.

[10] M. Chen and H. Jäenicke, "An information-theoretic framework for visualization," *IEEE TVCG*, vol. 16, no. 6, pp. 1206–1215, 2010.

[11] D. H. S. Chung, D. Archambault, R. Borgo, D. J. Edwards, R. S. Laramee, and M. Chen, "How ordered is it? On the perceptual orderability of visual channels," in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 131–140.

[12] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Assoc., Hillsdale, NJ., 1988.

[13] T. M. Cover, *Elements of Information Theory*. Wiley, 1999.

[14] J. E. Cutting and J. J. Garvin, "Fractal curves and complexity," *Perception & Psychophysics*, vol. 42, no. 4, pp. 365–370, 1987.

[15] J. E. Dickinson, K. Haley, V. K. Bowden, and D. R. Badcock, "Visual search reveals a critical component to shape," *Journal of vision*, vol. 18, no. 2, p. 2, 2018.

[16] G. T. Fechner, D. H. Howes, and E. G. Boring, *Elements of Psychophysics*. Holt, Rinehart and Winston New York, 1966, vol. 1.

[17] R. A. Fisher, "XV.—The correlation between relatives on the supposition of Mendelian inheritance." *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, vol. 52, no. 2, pp. 399–433, 1919.

[18] J. Görtler, C. Schulz, D. Weiskopf, and O. Deussen, "Bubble treemaps for uncertainty visualization," *IEEE TVCG*, vol. 24, no. 1, pp. 719–728, 2017.

[19] D. Hardman, *Judgment and Decision Making: Psychological perspectives*. John Wiley & Sons, 2009, vol. 11.

[20] M. J. Hautus, N. A. Macmillan, and C. D. Creelman, *Detection Theory: A User's Guide*. Routledge, 2021.

[21] N. Holliman, M. Turner, S. Dowsland, R. Cloete, and T. Picton, "Designing a Cloud-based 3D Visualization Engine for Smart Cities," *Electronic Imaging*, vol. 2017, no. 5, pp. 173–178, 2017.

[22] N. S. Holliman, M. Antony, J. Charlton, S. Dowsland, P. James, and M. Turner, "Petascale cloud supercomputing for terapixel visualization of a digital twin," *IEEE TCC*, vol. 10, no. 1, pp. 583–594, 2022.

[23] J. Hullman, "Why authors don't visualize uncertainty," *IEEE TVCG*, vol. 26, no. 1, pp. 130–139, 2019.

[24] A. J. Jackson and I. L. Bailey, "Visual acuity," *Optometry in Practice*, vol. 5, pp. 53–68, 2004.

[25] E. Landolt, "Nouveaux opto-types pour la détermination de l'acuité visuelle," *Arch. d'Opht*, vol. 19, pp. 465–471, 1899.

[26] A. M. MacEachren, *How maps work: representation, visualization, and design*. Guilford Press, 2004.

[27] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan, "Visual semiotics & uncertainty visualization: An empirical study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2496–2505, 2012.

[28] M. E. Magnello, "Karl Pearson and the establishment of mathematical statistics," *ISR*, vol. 77, no. 1, pp. 3–29, 2009.

[29] D. Makowski, "The psycho package: an efficient and publishing-oriented workflow for psychological science," *Journal of Open Source Software*, vol. 3, no. 22, p. 470, 2018.

[30] J. H. McDonald, "Handbook of Biological Statistics," 2022. [Online]. Available: http://www.biostathandbook.com/repgtestgof.html

[31] Met Office, "Key to symbols and terms - Met Office," 2022. [Online]. Available: https://www.metoffice.gov.uk/guide/weather/\symbols#temperaturesymbols

[32] T. Munzner, *Visualization Analysis and Design*. CRC press, 2014.

[33] T. O'Hagan, "Dicing with the unknown," *Significance*, vol. 1, no. 3, pp. 132–133, 2004.

[34] L. Padilla, "How to understand your climate uncertainty," *Interactions*, vol. 29, no. 1, p. 30–32, Jan 2022.

[35] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha, "Approaches to uncertainty visualization," *The Visual Computer*, vol. 13, no. 8, pp. 370–390, 1997.

[36] J. W. Peirce, "PsychoPy—psychophysics software in Python," *Journal of neuroscience methods*, vol. 162, no. 1-2, pp. 8–13, 2007.

[37] Peter Statistics, "Analysing a binary variable, Effect size: Cohen's g," https://peterstatistics.com/CrashCourse/2-SingleVar/\Binary/Binary-2b-EffectSize.html.

[38] R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria, 2023. [Online]. Available: https://www.r-project.org/

[39] J. S. Richman, D. E. Lake, and J. R. Moorman, "Sample entropy," in *Methods in enzymology*. Elsevier, 2004, vol. 384, pp. 172–184.

[40] J. Rigau, M. Feixas, and M. Sbert, "Conceptualizing birkhoff's aesthetic measure using shannon entropy and kolmogorov complexity." in *CAE*, 2007, pp. 105–112.

[41] G. Schmidtmann, B. J. Jennings, and F. A. A. Kingdom, "Shape recognition: convexities, concavities and things in between," *Scientific reports*, vol. 5, p. 17142, 2015.

[42] S. Schnur, K. Bektaş, and A. Çöltekin, "Measured and perceived visual complexity: A comparative study among three online map providers," *Cartography and Geographic Information Science*, vol. 45, no. 3, pp. 238–254, 2018.

[43] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[44] Stats Test, "G-Test of Goodness of Fit." [Online]. Available: https://www.statstest.com/g-test/

[45] E. Svarverud, S. Gilson, and A. Glennerster, "A demonstration of 'broken' visual space," *PLoS One*, vol. 7, no. 3, p. e33782, 2012.

[46] B. N. Taylor and C. E. Kuyatt, "Guidelines for evaluating and expressing the uncertainty of NIST measurement results," 1994.

[47] J. W. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.

[48] L. R. Varshney and J. Z. Sun, "Why do we perceive logarithmically?" *Significance*, vol. 10, no. 1, pp. 28–31, 2013.

[49] P.-P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich, "Viewpoint selection using viewpoint entropy," in *Proceedings of the Vision Modeling and Visualization Conference 2001*, ser. VMV '01. Aka GmbH, 2001, p. 273–280.

[50] T. von Landesberger, D. W. Fellner, and R. A. Ruddle, "Visualization system requirements for data processing pipeline design and optimization," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 8, pp. 2028–2041, 2017.

[51] F. Wilkinson, H. R. Wilson, and C. Habak, "Detection and recognition of radial frequency patterns," *Vision research*, vol. 38, no. 22, pp. 3555–3568, 1998.

[52] J. M. Wolfe and T. S. Horowitz, "Five factors that guide attention in visual search," *Nat. Hum. Behav.*, vol. 1, no. 3, pp. 1–8, 2017.

[53] J. Wolfe, "Visual Search," in *Attention*, H. Pashler, Ed. Psychology Press, 1998, ch. Visual Sea.

[54] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets," *Annals of biomedical engineering*, vol. 41, no. 2, pp. 349–365, 2013.

Author Biography

Nicolas S. Holliman Ph.D. (University of Leeds 1990) Computer Science, B.Sc. (University of Durham, 1986) joint hons Computing with Electronics. He is currently Professor of Computer Science at King's College London, Director of CUSP London and is a Fellow of the Alan Turing Institute in London.

Arzu Coltekin Ph.D. (Helsinki University of Technology, now Aalto, 2006) Photogrammetry and Geographic Information Science. She is currently a Professor at, and Director of, the University of Applied Sciences and Arts Northwestern Switzerland's Institute of Interactive Technologies.

Sara J. Fernstad PhD (Linkoping University, Sweden, 2011) Visualization and Inter-action, MSc (Linkoping University, Sweden, 2007) Engineering and Media Technology. She is currently a Senior Lecturer in Data Science in the School of Computing at Newcastle University (UK) and Director of Postgraduate Research.

Lucy McLaughlin M.Sc. (Newcastle University) , B.Sc. (Newcastle University) was an Alan Turing project funded Research Software Engineer and was a Ph.D. candidate on the EPSRC funded CDT in Cloud Computing and Big Data at Newcastle University.

Michael D. Simpson Ph.D. (Newcastle University, 2016) Computing Science, M.Sc. (Newcastle University, 2009) Computer Games Engineering, B.Sc. (Newcastle University, 2008) Computing Science. He is currently a Research Software Engineer at Newcastle University, specializing in data visualization and 3D.

Andrew J. Woods PhD (Curtin), MEng (Curtin), BEng (Curtin). He is currently an Associate Professor at Curtin University. As head of the Curtin University HIVE (Hub for Immersive Visualization and eResearch) he has made contributions to through his research, development and utilization of 3D technologies.