Evaluating Sketchiness as a Visual Variable for the Depiction of Qualitative Uncertainty

Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, Jean-Daniel Fekete

HAL Id: hal-00717441
https://inria.hal.science/hal-00717441

Submitted on 30 Jul 2012

Evaluating Sketchiness as a Visual Variable
for the Depiction of Qualitative Uncertainty

Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, *Member, IEEE*, and Jean-Daniel Fekete, *Member, IEEE*

Abstract—We report on results of a series of user studies on the perception of four visual variables that are commonly used in the literature to depict uncertainty. To the best of our knowledge, we provide the first formal evaluation of the use of these variables to facilitate an easier reading of uncertainty in visualizations that rely on line graphical primitives. In addition to blur, dashing and grayscale, we investigate the use of 'sketchiness' as a visual variable because it conveys visual impreciseness that may be associated with data quality. Inspired by work in non-photorealistic rendering and by the features of hand-drawn lines, we generate line trajectories that resemble hand-drawn strokes of various levels of proficiency—ranging from child to adult strokes—where the amount of perturbations in the line corresponds to the level of uncertainty in the data. Our results show that sketchiness is a viable alternative for the visualization of uncertainty in lines and is as intuitive as blur; although people subjectively prefer dashing style over blur, grayscale and sketchiness. We discuss advantages and limitations of each technique and conclude with design considerations on how to deploy these visual variables to effectively depict various levels of uncertainty for line marks.

Index Terms—Uncertainty visualization, qualitative evaluation, quantitative evaluation, perception.

✦

1 INTRODUCTION

Information visualization can show not only what we know about the data but also the degree of our confidence in that data. This confidence could be considered as yet another data dimension. However, information on data quality in general—if at all available—rarely comes in numerical format. Qualitative measures of uncertainty are far more common and often come as ordinal meta-data. For instance, a utility company holds a positional confidence attribute for their assets where uncertainty is mapped to five categorical values from least certain to more certain: schematic, assumed, indicative, third party survey, and internal survey [8]. Such ordinal data can be visualized using Bertin's [4] retinal variables texture, value, or size. When visualizing uncertainty, however, a number of visual variables are considered more 'intuitive' for this domain; examples include blur, sharpness of focus, and color saturation. These variables may bear direct perceptual resemblance to what the uncertainty indicates and, thus, may provide an easier reading of uncertainty [11, 31, 45]. However, to our knowledge, no formal studies are reported to back up these observations.

In this paper, we investigate *sketchiness* as a visual variable to depict uncertainty information in line marks such as for graphs, hierarchies and route maps. The design of these sketchy lines is inspired by the field of Non-Photorealistic Rendering (NPR) and by our own observations from analyzing child and adult hand drawings. Our analogy is that the 'cleanness' of the hand-drawn lines corresponds to the quality of the data. We hypothesize that *sketchiness* is a good metaphor for the qualitative visualization of uncertainty information.

Our contribution is two-fold: (1) we provide an empirically-based method to generate line trajectories that resemble hand-drawn strokes of various levels of proficiency—ranging from child to adult strokes—where the amount of deviations in the line corresponds to the level of uncertainty in the data; (2) we present a qualitative and quantitative evaluation of four uncertainty visualization techniques using *blur*, *dashing*, *grayscale*, and *sketchiness*. In particular, we attempt

to answer the following questions: (a) can people intuitively associate sketchiness to uncertainty, (b) is sketchiness as effective for depicting uncertain information as the other visualization techniques, and (c) which method do people subjectively prefer.

The remainder of the paper is organized as follows. After discussing previous work, we first present our model for generating sketchy lines, and then present a series of studies that answer our questions on intuitiveness, accuracy of reading uncertain visualizations, and subjective viewer preference in regards to the different uncertainty visualizations.

2 RELATED WORK

The previous work that relates to our own can generally be thought of as threefold: (a) work in non-photorealistic rendering that studies the generation of primitives that appear sketchy or hand-drawn as well as their applications, (b) methods to depict uncertainty data, and (c) user studies that examine how people perceive and interpret uncertain data depictions. We discuss each of these three fields in turn.

2.1 Sketchy Lines in Non-Photorealistic Rendering

The field of non-photorealistic rendering (NPR) [21, 48] has been inspired by the many ways of traditional depiction that humans have employed over the last decades, centuries, and millennia. As such, computer-generated line drawings have been among the first goals for NPR. Early-on, for example, loose and sketchy line rendering and animation [14] have been simulated. Researchers have also developed ways to represent lines such that the line path can be separated from the (sketchy) line perturbation [16, 41]. These line models can, in turn, be applied to line renderers (e. g., [46, 53]) to generate images that resemble—to varying degrees—traditional hand-drawn ones. More recently, NPR line models have been extended to be more example-based, e. g., by taking the human arm movement into account in their generation [3]—a goal that we share for generating our sketchy lines.

In addition to reproducing a generic hand-drawn look, sketchy non-photorealistic rendering has also specifically been employed to portray uncertainty. For example, Strothotte et al. [45] used a general level of sketchiness to indicate a general notion of uncertainly in the domain of archeology visualization. They also describe examples that continuously change the line thickness/line saturation or the degree of perturbation in order to visualize continuously changing degrees of uncertainty. In the same domain, Potter et al. [38] perturb the vertices of a line-based rendering style to control the degree of sketchiness of the depiction, which in turn is used to visually express the level of confidence in a reconstruction. A similar notion of general sketchiness to indicate uncertainty was used by Nienhaus and Döllner [34] for the

- *Nadia Boukhelifa and Jean-Daniel Fekete are with INRIA, France; email: (nadia.boukhelifa,Jean-Daniel.Fekete)@inria.fr*
- *Anastasia Bezerianos is with LRI (Univ Paris-Sud & CNRS) and INRIA, France; email: anastasia.bezerianos@lri.fr.*
- *Tobias Isenberg is with the University of Groningen, the Netherlands, and with DIGITEO/CNRS/INRIA, France; e-mail: isenberg@cs.rug.nl.*

visualization of 3D shape concepts in CAD—supported by earlier findings of Schumann et al. [42]. Instead of these rather implicit notions of uncertainty visualization we are interested in a more precise analysis of how line sketchiness can be employed for qualitatively depicting a number of different levels of uncertainty, and thus in how sketchiness can be used intentionally as a dedicated visual variable.

2.2 Visual Variables to Depict Uncertainty

Many visual variables have been proposed for the depiction of uncertainty including Bertin's [4] retinal variables position, size, and value. To help visualization designers choose a technique from a myriad of possibilities, various taxonomies were proposed [32, 36, 49]. Our aim here is not to replicate these efforts but to provide an overview of widely used methods that are applicable to line primitives. We group these methods into three main categories: (1) color-based techniques that manipulate hue, saturation, or brightness dimensions; (2) focus-based techniques that modify contour crispness, transparency, or resolution; and (3) geometry-based techniques that distort line marks by applying a rendering style as in sketchiness.

Color is repeatedly used to depict uncertainty in information visualization [1, 15, 48]. Since color scales are readily available in today's graphical packages and libraries, assigning a range of a color scale to a range in the data is straightforward. Amongst all color dimensions, saturation is often preferred because the reduction of color purity conveys more intuitively the notion of degrading data quality or confidence [31, 45]. Hue is also used [39, 55, 56] and rainbow scale is often chosen to map ordinal uncertainty data up to seven levels—despite of its lack of perceptual ordering [6]. For value, darker lines suggest more certainty about an aspect of the underlying data. In this case, however, the line width may need to be adjusted to preserve perceptibility [11]. More advanced color mapping techniques are also deployed including whitening where white pixels are randomly placed or actual hue and white are blended [25, 26, 37].

Amongst focus-based techniques, blur, which is defined as the removal of high-frequency spatial detail from the information [10] has widely been used to indicate fuzziness and ambiguity in the data [7, 9, 18, 31]. For example, Bisantz et al. [5] applied blur to a set of airplane symbols to provide decision makers with a fast way to understand the level and uncertainty of a given threat. Gershon [19] used blurred versions of images in an animation loop from focused to blurry (or the inverse) to draw users' attention to uncertain objects. Kosara et al. [27] described their 'semantic depth of field' technique that uses blur to de-emphasize objects of less importance in the scene (de-emphasis of objects that are less important can also be achieved using transparency; e. g., Correa et al. [13] map uncertainty to both size and transparency). We use blur as one visual variable in our evaluation and compare it to others like sketchiness, dashing, and value.

Modifying the geometry of line marks can be a powerful way to convey uncertainty. For instance, Griethe and Schumann [22] argue that "wavy or dotted lines could convey less trusted relationships." Similarly, Strothotte et al. [47] show how sketch-like renditions (and the use of transparency) can express the uncertainty in archaeological reconstructions. The domain of NPR also proposed many geometry-based rendering techniques to convey uncertainty as described above. Using drawing primitives from NPR, Pang [35] used gaps in contour lines of geographic maps to encode uncertainty such that larger gaps in the contour line encode an increased uncertainty. In the related area of oceanography, Osorio et al. [2] also augmented contour visualizations using uncertainty bands that indicate different possible locations of a contour line. Luboschik et al. [30] suggested the use of dashed and wavy lines to show uncertainty in parallel coordinates. In the context of maritime situational awareness, Matthews et al. [33] depict the timeliness and quality of sensor information using icon borders—solid or broken. More relevant to our work, however, is the distorted annotation technique by Cedilnik and Rheingans [11] who distort grid lines proportionally to the amount of uncertainty in the data. Similar to our approach for sketchiness, Cedilnik and Rheingans map the amplitude of line distortion to the amount of uncertainty in the data. Our method, however, does not separate the data from the uncertainty depiction be-

cause the affected lines themselves are data carriers in our case.

2.3 Perceptual Studies and User Evaluations

Several perceptual studies and user evaluations have examined the general application of visual variables in visualizations. It is established that color is a powerful dimension to indicate data quality in general [28, 31]. Specifically for the context of multidimensional data visualization, Xie et al. [55] found that hue has stronger capacity to convey quality information for parallel coordinates than brightness or line width—even for large datasets. Color, however, may suffer from the lack of an intuitive order. MacEachren [31] showed, for example, that subjects cannot spontaneously order colors into a legend arrangement for bi-variate choropleth maps but that they can recognize order in that arrangement. The question of user preference is also pertinent to the problem of uncertainty visualization; in a user opinion survey conducted by Gerharz et al. [18] in the context of geographical information systems, people disliked whitening [26] to convey uncertainty. The authors argue that the principle of whitening is easy to understand but getting detail information from it is difficult. Li et al. [29] investigated the issue of scale for uncertainty visualization for astrophysical data and used a unified color scheme to represent log-scale distances and percentage errors. They found that participants were able to determine the amount of uncertainty using colors with 96.7% success rate. Similarly to the findings by Gerharz et al. [18], however, access to detail was difficult especially for neighboring color ranges.

Blur as an example of a focus-based uncertainty visualization technique is a well-studied visual variable in domains that go beyond information visualization (e. g., [23, 51]). Kosara et al.'s [27] evaluation of the previously mentioned 'semantic depth of field' specifically examined people's ability to read absolute blur levels. Their results show that participants can distinguish between different blur levels with good accuracy but cannot quantify this difference nor identify objects of the same blurriness. Kosara et al. thus concluded that blur can guide attention but is hard to quantify and thus may not be recommended for showing *quantitative* uncertainty. Moreover, Kosara et al. found that participants disliked looking at blurred objects. These results motivate our work in which we examine whether blur is effective in communicating *qualitative* uncertainty information.

Few quantitative evaluations exist for geometry-based methods. One exception is the work by Matthews et al. [33] who found that—in visual search tasks on maps where latency of information is indicated either by color hue (green or gold) or border style (solid or broken)—search time was faster with border style than with the color format.

More generally, the choice of which uncertainty visualization method to pick may be task-dependent [40]. Evaluating uncertainty depiction techniques for information visualization applications in general is thus challenging. For example, MacEachren [31] states that "for exploratory applications, where there is no predetermined message to communicate, we can not judge uncertainty depictions using communication effectiveness standards. We can only evaluate these depictions in terms of how they alter decision making, pattern recognition, hypothesis generation, or policy decisions." Zuk and Carpendale [57] presented a set of heuristics for uncertainty visualization evaluations and stressed the need for more research in human factors and perception. We are inspired by this call for action and conduct a comparative study between four popular visual variables traditionally used to visualize uncertainty including methods from the three described categories. For color-based methods we use the visual variable *value* (i. e., grayscale), for focus-based techniques we examine *blur*, and for geometry-based techniques we selected *dashing* and *sketchiness*. To our knowledge, there has not been a formal comparative evaluation of the use of the mentioned visual variables to facilitate an easier reading of uncertainty in visualizations that rely on line primitives. Sanyal et al.'s [40] comparative study for 1D and 2D data comes close to ours, but they evaluated different visualization methods (error bars, size of glyphs, color-mapping on glyphs, and color-mapping on data surface), and tested for different user-study tasks (search and count). To enable our evaluation of the selected variables we first describe in the next section how to mimic hand-drawn trajectories to be able to create sketchy
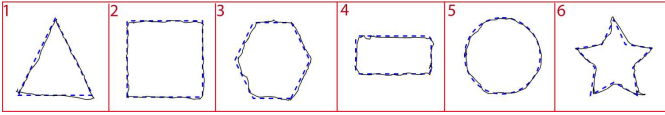
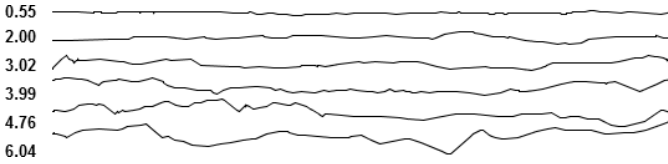Fig. 1. An eight-year-old drawing data sheet (traced with her left hand).

Fig. 2. Examples of *unfolded* hand-drawn strokes which we ranked by their average deviation (in mm) from their target shapes.

lines, before detailing our study.

3 MIMICKING HAND-DRAWN TRAJECTORIES

To to be able to synthesize lines that mimic hand-drawn ones we first studied characteristics of hand-drawn strokes collected using an Anoto pen. We asked 20 people whose ages ranged from five to 47 years (mean of 27.0 years) to trace six different shapes (see dashed blue lines in Fig. 1); each participant first drew with their dominant and later with their non-dominant hand to ensure variations in drawing proficiency. There was no time limit set for the task but participants were instructed to draw as they would *normally* but they should aim for good accuracy. To facilitate this task, we asked participants to stay within the enclosing red boxes in Fig. 1. In total, we collected 204 hand-drawn shapes. With the exception of one person, none of our volunteers had a formal drawing training experience.

Our hypothesis was that drawing proficiency can be determined by examining the average deviations of the drawn path from the template path. A quick visual inspection showed that, indeed, the subjectively more proficient-looking drawings deviated less from the template shapes. Fig. 2 shows examples of unfolded hand-drawn shapes, ranked by their average deviation from the target shapes. Within the set of hand-drawn strokes we collected, the average deviations per stroke ranged from 0.55 mm to 6.04 mm (which corresponds to 2.08–22.82 pixels at 96 ppi screen resolution). An error distribution analysis of all stroke control point deviations (signed) showed a normal distribution with a mean of 0.36 mm and a standard deviation of 2.08 mm (min: −13 mm; max: 18 mm). We decided to use the average deviation from a straight line as our way to map levels of uncertainty in data.

Mimicking hand-drawn lines, of course, requires a generative model of hand movement. Flash and Hogan [17] described such a model for straight lines, based on minimizing jerk during a stroke from $p_0 = [x_0, y_0]$ to $p_1 = [x_1, y_1]$, jerk being the derivative of acceleration. At time t, the model (improved by AlMeraj et al. [3]) generates a point:

$$x(t) = x_0 + (x_0 - x_1)(15\tau^4 - 6\tau^5 - 10\tau^3) + D$$
$$y(t) = y_0 + (y_0 - y_1)(15\tau^4 - 6\tau^5 - 10\tau^3) + D \quad (1)$$

with $\tau = \frac{t}{t_f}$, t_f being the time of the end of the stroke, and D a random value in a specified (pixel) range—D adding squiggliness to the line.

AlMeraj et al. [3] empirically defined the time sampling parameter δt according to the length of the desired line: $\delta t = 0.5s$ for lines shorter than 200 pixels, $\delta t = 0.3s$ for lines of 200–400 pixels, and $\delta t = 0.2s$ for lines longer than 400 pixels. They also fixed $t_f = 2s$ and D within $[-5, 5]$ pixels range based on empirical evidence. Using these parameters in Equation 1 yields multiple points that are connected using a smoothing spline.

3.1 Generating Levels for *Sketchiness*

Our method to produce sketchy lines takes a number of control points generated by the Flash and Hogan model and a value for D, our sketchiness parameter, to produce deviations within $[-20, 20]$ pixels. This
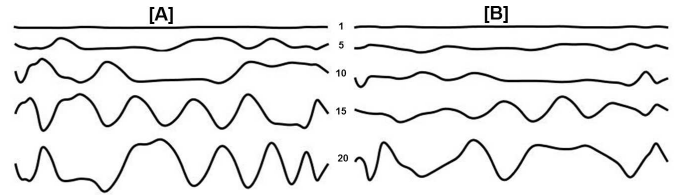

Fig. 3. Ranges (in pixel) of *sketchiness* used in the study for the mean-biased [A] and the max-biased [B] variants.

range corresponds roughly to the average stroke deviation range we observed from hand-drawn strokes (where the top range of deviation was 22.83 pixels). Since we are essentially adding noise to straight lines to produce sketchy lines that are consistently perceived by viewers as belonging to the same level, we make sure that the mean deviation stays faithful to the input D value. To achieve this correspondence we sample values from a normal distribution where the mean is set to D and the standard deviation is set empirically to $D/6$. For our studies, we produced sketchy lines where the difference between D value and actual mean is less than 0.1 pixels. This approach can be described as being *mean-biased*.

An alternative strategy to generate levels consistently, described as *max-biased*, is to use the maximum deviation. We sample values for the D parameter from a normal distribution where the mean is equal to 0 and the standard deviation is equal to D. In addition, we add a constraint whereby D values outside $[-20, 20]$ pixels are not allowed.

Fig. 3 shows five major levels as generated by the two methods. As D grows, lines under the mean-biased condition get more undulated in order to preserve the overall mean value, whereas sketchy lines under the max-biased condition overall have a more flat appearance with the exception of a few peaks.

We conducted a between-subjects pilot study to compare the two variations. The pilot was identical in setup to the levels study in Section 6.2 and was carried out in two parts. First, participants were shown one line representing the maximum magnitude for *max-biased* sketchiness (100% our modulus) as well as a stimulus line that they had to express as a percentage of the modulus. Second, a similar prompt was presented for the *mean-biased* variation. Results from 20 participants (10 per variant) were analyzed and a mean error comparison of the two variations showed that the *max-biased* variation resulted in smaller errors in the perception of the sketchiness level than the *mean-biased* variation, and thus we used the former in our studies.

3.2 Assessing Generated Sketchiness Quality

We conducted an online study with 40 participants to determine if people are able to tell the generated lines (using the max-biased approach) apart from hand-drawn ones. The details of our general study setup are reported in Section 4.2.

We selected 45 samples from our library of hand-drawn strokes of three shapes (hexagons, rounded rectangles, triangles) from a deviation range between 0.9 mm and 4.9 mm (within the range we observed in real strokes), and generated the same shapes using the generative model (Fig. 4). For the hand-drawn condition we selected shapes with complete outlines and no noticeable variations in width or overshoots in corners (as much as possible). For the sketchy shapes, we varied D within $[0.4, 10]$ pixels to obtain different sketchiness levels. Our hypothesis was that people cannot tell computer- from human-generated shapes apart for D values of up to 6 pixels . This hypothesis is based on our own subjective observations that shapes start to look synthetic at $D \approx 6$: for larger D values, corners start to loose their shape and knots start to appear, something that is not associated with hand-drawn strokes. The null hypothesis was, thus, for a given sketchiness D, participants will be unable to distinguish machine-generated strokes from human-made ones (i. e., selecting the right answer 50% of the time).

The results of the online study show that, overall, 60% of the computer-generated shapes were identified as hand-drawn ones, and 71% of the real hand-drawn shapes were correctly identified as such. A
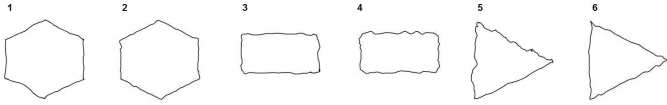
Fig. 4. Examples of hand-drawn (odd-numbered) and generated (even-numbered) shapes: each two variants have similar average deviation.
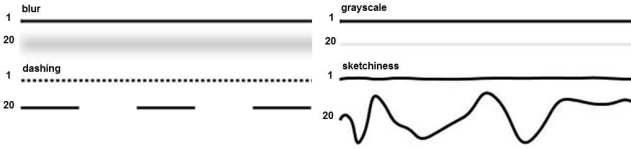

Fig. 5. Ranges of levels for the four visual variables used in our study (excluding minima).

one-tailed non-parametric Bionomial test rejected the null hypothesis ($p < .0001$), showing that participants were significantly more likely to think the computer-generated lines were human-made (thus the effect was not due to chance). Nevertheless, we did not find an effect of sketchiness level, indicating that regardless of D (i.e., even with the highest level of sketchiness tested) participants were still more likely to perceive the shapes generated by our algorithm as hand-drawn.

4 USER STUDIES: DESIGN GOALS AND REQUIREMENTS

Using our results from mimicking sketchy lines based on empirically determined parameters, we can now pursue our main objective: investigate whether sketchiness can be used as a visual variable to encode qualitative uncertainty data in information visualization. For this purpose we compared sketchiness to popular uncertainty visualization techniques with respect to three criteria: *intuitiveness*, *accuracy*, and *subjective user preference*. For intuitiveness, we wanted to investigate if and to what extent people associate sketchiness (and the other variables) to uncertainty. For accuracy, we wanted to investigate how accurately people can read values from uncertain line marks (characterized by the use of *blur*, *dashing*, *grayscale*, and *blur*) and the number of distinctly perceivable levels for each technique. Finally, we wanted to study which technique would be preferred by participants to encode uncertainty.

4.1 Generating Levels for *Blur*, *Dashing*, and *Grayscale*

We generated all blurred, dashed, and grayscale lines using the GNU Image Manipulation Program (GIMP). We generated 20 levels per visual variable; each line was 400 pixels long, three pixels wide, and used black except for the grayscale case. Fig. 5 shows the minimum and maximum levels for each visual variable (excluding minima).

For *blur*, we used a Gaussian filter for which we varied the horizontal and vertical blur radius in equal amounts from 1 to 20 pixels. For *dashing*, we selected a style where the dash/gap length grows proportionally with each consecutive level. The minimum dash/gap size was set to 3 pixels and the maximum to 60 pixels. For *grayscale*, we partitioned a linear grayscale from 4% up to value 90% into 20 discrete levels from black to white since an increase in whiteness may be seen as an increase in uncertainty in the data [26]. As described in Section 3, the 20 levels for *sketchiness* were produced using *max-biased* generation, with $D \in [-20, 20]$.

4.2 Setup of the Studies

Our studies were conducted on Amazon Mechanical Turk (AMT), inspired by previous graphical perception experiments [24]. In total, we deployed seven experiments and participants in one study could not take part in any of the other ones. We had 1176 participants in total, and subjects were paid on average $0.34 per Human Intelligence Task (HIT). To ensure that our sample participant size was representative of the general participant pool, we conducted our studies in increments (three to five blocks). For each block we calculated the mean of what

we are measuring (e.g., user subjective preference per technique) and stopped the experiment after we noted stabilization (i.e., discrepancies in mean values consistently inferior to 10%). This sampling procedure was followed throughout our experiments. Given the simple nature of the perceptual studies we were carrying out, no qualification tests were required to complete our HITs. In accordance with AMT guidelines, however, only workers with 95% or more HIT approval rate were allowed to participate. Furthermore, we dismissed work were the participant's written language was poor, indicating a possible language barrier affecting comprehension. Finally, for verification purposes, we included an unrelated answer when presenting workers with a list of options (for studies in Section 5.2 and Section 5.3). Workers were not paid for HITs where they picked the control answer and their data was not included in the analysis.

5 ON THE QUESTION OF INTUITIVENESS

The question of whether people intuitively associate sketchiness with uncertainty is pertinent to our evaluation of sketchiness as a visual variable. We define *intuitiveness* as the spontaneous association between signifier and signified: signifier being sketchiness and signified being uncertainty. If we find that there exists a spontaneous association between the two we could hypothesize that people can intuitively associate sketchiness with uncertainty. In that case they do not need to consult a legend to identify the meaning of the visual variable. Note that at this stage we are interested in the meaning associated with the representation—that it is uncertain data—and not its perceived magnitude, i.e., how uncertain it is. We, thus, conduced a study to examine the intuitiveness of sketchiness as a visual variable for uncertainty. This study was conducted in three parts. Each participant only completed one of the three parts, and each part was conducted as a between-subjects experiment.

In Part I, participants were shown visualization scenarios that included sketchiness and were asked in an open-ended question to explain the meaning of sketchiness. Their answers were then used to establish the main categories of interpretations that people spontaneously associate to sketchiness and allowed us to examine whether uncertainty is prominent among them. In Part II we ran a similar study, but participants were presented with a closed list of possible interpretations for sketchiness (multiple choice), established based on the results of Part I. The goal of Part II was to check if the extent of this association changes with the introduction of limited alternatives. In Part III we compared sketchiness to blur—a visual variable that is highly regarded in the literature as being congruent with uncertainty depiction [11, 31, 45]—using the same closed list of categories as in Part II. This list was general enough to apply to both visual variables.

For this series of studies, and based on our own experience and observations from related work, we hypothesized that:

H1 People are likely to spontaneously associate sketchiness to uncertainty in the data.

H2 People are more likely to associate sketchiness to uncertainty for abstract contexts (e.g., hierarchy) than for non-abstract contexts (e.g., map) where geometry already has an inherent meaning.

H3 Participants are more likely to associate blur with uncertainty than they are to associate sketchiness with uncertainty.

5.1 Part I: *Sketchiness*

Scenarios. We asked people to look at different visualizations with lines as a major visual feature and to interpret what sketchy lines mean to them. We designed six different scenarios, the first four being *abstract* contexts (Fig. 6) and the last two being *non-abstract* contexts (Fig. 7): (S1) a bar chart where sketchy lines were applied to the contours of some of the bars, (S2) a family tree where sketchiness was applied to some connections between parents and children, (S3) a social network graph where sketchiness was applied to certain connections between nodes (i.e., to relationships between two connected people), (S4) a Venn diagram where sketchy outlines were applied to (parts of) the outlines of some ellipsoids that indicate set memberships, (S5) a
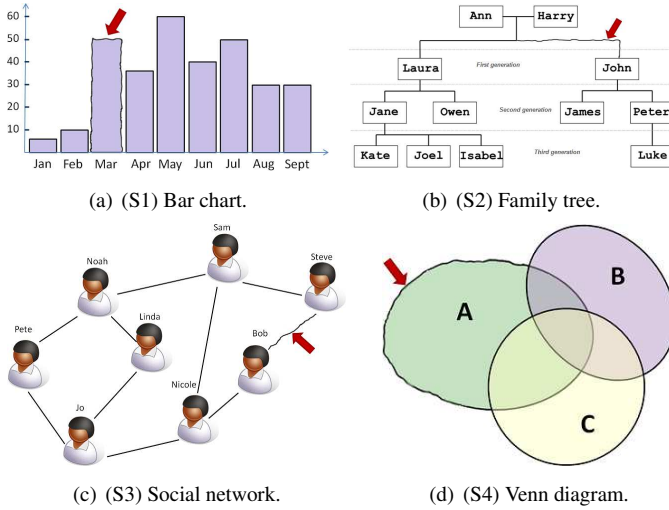
(a) (S1) Bar chart.

(b) (S2) Family tree.

(c) (S3) Social network.

(d) (S4) Venn diagram.

Fig. 6. The four abstract scenarios used in the study.

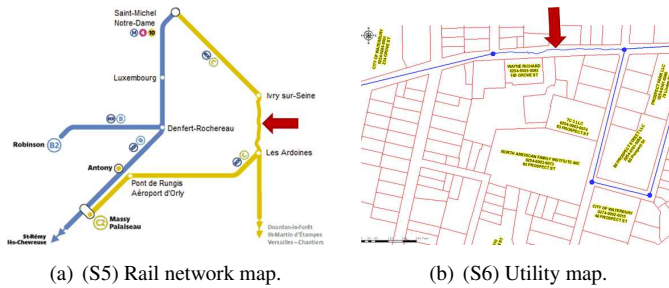(a) (S5) Rail network map.

(b) (S6) Utility map.

Fig. 7. The two non-abstract scenarios used in the study.

rail network where sketchiness was applied to some of the links between two train stations, and (S6) a utility map where sketchiness was applied to line representations of some buried assets.

For each scenario we generated five image variations, changing the percentage of lines in the visualization that were represented as sketchy (between 10% to 50% of all lines in the visualizations). The sketchy lines themselves were hand-drawn with an Anoto pen. Participants only saw one variation of each scenario.

Participants, Study Design, and Procedure. 210 participants took part in this study. Participants were split into six groups of 35 people. Each group was exposed to one of the different scenarios described above and each participant in the group saw a single variation of the scenario. Participants were first introduced to the task, provided with a short scenario description, and then presented with an image as described above. We asked participants to type two different interpretations into a text box stating what these lines convey to them. We highlighted each of the sketchy lines in question using a red arrow. Overall, our experiment consisted of:

$$
\begin{array}{rl}
& 6 \quad \text{scenarios} \\
\times & 35 \quad \text{participants per scenario} \\
\hline
= & \textbf{210} \quad \textbf{trials in total}
\end{array}
$$

After discarding entries that obviously did not represent an interpretation (e. g., jokes, unrelated answers) [52], we could analyze a total of 180 trials, 30 for each scenario.

Results. We carried a qualitative evaluation of the results; one of the authors coded all interpretations; then a second author independently encoded 40% of the total number of interpretations. The concordance rate between the two encoders was around 74%. Coding conflicts were resolved and we were able to identify six major categories of interpretations for sketchiness as given below with their association rate (derived using the mean for both interpretations combined). In

contrast to what we stated in our first hypothesis H1, we found that only **11.7%** of people associate sketchiness to uncertainty, less than we expected. However, we found no effect between the amount of sketchy lines in an image and the number of interpretations that associated it with uncertainty. The results for all categories are listed below.

- Alternative (36.7%): a different relation from what is conveyed by a straight line, which includes 'geometry' where sketchiness is attributed to the actual shape of the displayed feature.

- Qualitative (23.3%): the same relation as for a straight line but with emphasis on a particular quality other than uncertainty. Often, this was a negative quality.

- Ignore (13.9%): the exact same relation as for straight lines (no added information).

- Uncertainty (11.7%): all data quality descriptors that are related linguistically to the term uncertainty such as ambiguity, vagueness, impreciseness, doubt, or unreliability.

- Style (10.0%): intended drawing style; e. g., to draw attention to a particular part of the image.

- Glitch (4.4%): unintended style; i. e., human or computer error.

It was interesting that some participants did not spontaneously associate any semantic meaning with sketchiness (i. e., the 'ignore' category above). Because we were interested in meaningful interpretations, however, we decided to exclude this category from the list for the next two *primed* studies and thus to only use the remaining five categories.

5.2 Part II: Primed *Sketchiness*

Similar to Part I of the study in Section 5.1, we asked people to look at a line drawing and interpret what the sketchy line might mean. We provided participants with the five-category list established above, further adapted for sketchiness *abstract* (SA) and *non-abstract* contexts (SNA) as described below.

Participants, Study Design, and Procedure. 168 participants took part in this study, adopting a between-subjects design. The design of the study was identical to the study in Part I, with the exceptions that we asked participants to pick an interpretation from the provided list of options that we adapted to include the 'geometry' category for SNA contexts S5 and S6. Note that, rather than the general category names, we provided specific example interpretations that were repeatedly given by participants in Part I. For instance, for the 'geometry' category in scenario S5 we gave 'crooked or un-smooth train tracks' as the corresponding possible interpretation. Thus overall we had:

$$
\begin{array}{rl}
& 6 \quad \text{scenarios} \\
\times & 28 \quad \text{participants per scenario} \\
\hline
= & \textbf{168} \quad \textbf{trials in total (1/3 for SNA scenarios)}
\end{array}
$$

Results. Overall, people attributed sketchiness first to style (36.3%) and then to uncertainty (22.0%), showing an increase for uncertainty under the closed list condition. The next category scores, in decreasing order, were qualitative (13.69%), alternative (12.50%), geometry (10.12%), and glitch (5.36%). Separating SA and SNA scenarios, we found an important difference: in agreement with our hypothesis H2, 28.6% of participants under the *SA* condition attributed sketchiness to uncertainty in comparison to only 8.9% for the *SNA* condition. This may have implications on the type of visualization context for which sketchiness is more intuitive as a visual indicator of uncertainty.

5.3 Part III: Primed *Blur*

We replicated the previous study (Part II) for the *blur* visual variable, therefore the tasks are the same as in Section 5.2.

Participants, Study Design, and Procedure. 168 participants took part in this study, adopting a between-subjects design. Participants were split into six groups with 28 people per group. Each group was exposed to one of the different scenarios S1–S6. Participants were

first introduced to the task, were given a short description of the scenario, and were then presented with an image in which we had applied a Gaussian blur to one line (as outlined in Section 4). We asked participants to pick a single interpretation from the five-category list in Section 5.1. We highlighted the blurred line in question using a red arrow each time. Thus overall our experiment consisted of:

> 6   scenarios
> × 28   participants per scenario
> = **168   trials in total (1/3 for non-abstract scenarios)**

Results. Overall, people attributed blur mostly to style (26.2%), similar to the results for sketchiness. The second-most frequent attribution was a qualitative measure (23.8%) and only the third-most frequent one was uncertainty (22.6%). The least frequent attributions for blur were alternative (17.26 %) and glitch (10.12 %). Therefore, contrary to our hypothesis H3, sketchiness appears to be as intuitive as blur for the tested categories.

6 ON THE QUESTION OF PRACTICAL CONSIDERATIONS

After having established that *sketchiness* is just as associated with uncertainty as blur, we can now investigate the practical aspects of using *sketchiness* for this purpose, or as a visual variable in information visualizations in general. More specifically, we are interested in determining if we can express a scale using *sketchiness* as accurately as when using one of the other three visual variables traditionally used to represent uncertainty (*blur*, *grayscale*, and *dashing*). We thus compare the four techniques with respect to: (i) how close their perceived values are to the actual visual variable value; (ii) how many distinct levels can be identified by participants for each technique to use them in ordinal scales; and in (iii) whether the perception of these techniques follows previous models that express this perception mathematically.

6.1 Hypotheses

Based on our experience and observations from related work, we hypothesized that a few discrete levels of line *sketchiness* can be perceived but that people cannot accurately quantify the amount of *sketchiness* applied to a line. This led to the following hypotheses:

H4 Overall, people cannot accurately estimate the exact level for all techniques (*sketchiness*, *blur*, *grayscale*, and *dashing*), but this perception error will be different between techniques.

H5 People can discriminate between at least three levels of *sketchiness* but more levels for the remaining techniques.

H6 We can express the relationship between reported and actual levels for all visual encoding techniques (including *sketchiness*) using a mathematical model.

6.2 Task

To assess the visual perception of the techniques we conducted a study with tasks from psychophysics, the domain that focuses on measuring relationships between perceived and actual properties of visual objects [20, 50]. Of the methods that help assess a viewer's visual perception of an object compared to its subjectively experienced magnitude, numeric estimation methods are most relevant to our research and have been frequently used in in the past (e. g., [20, 50]). Participants are shown a standard modulus object with an assigned value (e. g., 100% *sketchiness*) and are then asked to assess a second object (the stimulus) and assign it a value based on the modulus (e. g., a percentage).

In our study, participants were shown one line representing the maximum magnitude for the specific visual variable technique (100% our modulus) as well as a stimulus line that they had to express as a percentage of the modulus. We used the previously generated 20 magnitude *levels* for stimulus lines for all our visual variables. These 20 stimulus levels correspond to 5–100% of the modulus object. Participants were also shown a line representing the minimum (0%) for the technique which, for all techniques, was a three-pixel-wide straight black line.
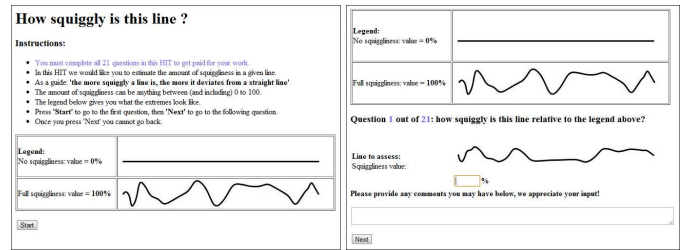

Fig. 8. The user interface for the sketchiness levels study showing the instructions and legend (left) and a magnitude estimation task (right).

6.3 Participants, Study Design and Procedure

160 participants took part in this study. A mixed factorial design was used: the *visual variable* was treated as a between-subjects factor and the *level* was treated as a within-subjects factor. Participants were split into four groups of equal sizes. Each group was exposed to a different visual variable and each participant conducted magnitude estimation tasks for all levels using that technique.

Participants were first introduced to the task. They were then presented with lines of different levels of their respective visual variable and were asked to mark on a scale from 0% to 100% how sketchy, for instance, each line was with respect to the maximum modulus line which was always visible (Fig. 8 for an example HIT). The order of presentation of the different levels was randomized across participants and techniques. We tested a total of 21 levels per technique (including level zero). Our experiment thus consisted of:

> 4   techniques (*dashing*, *blur*, *grayscale*, and *sketchiness*)
> × 21   magnitude levels (including level zero)
> × 40   participants per technique
> = **3360   trials in total**

To balance our data, we conducted our level experiments in four steps; we ensured we had 40 completed and valid HITs for one visual variable before progressing to testing the next visual variable.

6.4 Results

The metrics used in our analysis were the absolute perception error *AbsErr* and the perceived magnitude level *PerMag* of a line. As done by similar magnitude estimation studies (e. g., [12]) we define *AbsErr* as the absolute difference between the true percent of the stimulus compared to the modulus and the reported estimation percent $|reported\ level - true\ level|$. Absolute magnitude errors have a skewed distribution and, as suggested by Cleveland and McGill [12], we normalize it by using the log variation of this metric for our analysis $log_2(\frac{1}{8} + AbsErr)$. The means reported here are before normalization.

Trials were marked as outliers when metrics were beyond two standard deviations from the mean for a given technique and level. 161 trials (5% of all trials) were identified as outliers and removed from further analysis. We performed an ANOVA and post-hoc pair-wise mean comparison p-values are adjusted using the Bonferroni criterion.

6.4.1 Perception Error

We first examined how closely participants came to predicting the real level value for each technique, conducting an analysis on *AbsErr*. The overall *AbsErr* was higher for *sketchiness* (16.6%), followed by *dashing* (13.2%), *blur* (12.7%), and *grayscale* (12.3%).

AbsErr increased steadily from 5% to 20% overall as we increased the level up to 11 (half of maximum magnitude). Then we observed a steady decrease in error for all techniques from 20% to 15% for level 19, and almost 5% in level 20. The ANOVA showed that the main effect of technique on *AbsErr* was statistically significant ($F_{3,156} = 4.890$, $p < .01$). A significant interaction effect between technique and level was also present ($F_{60,3120} = 6.364$, $p < .001$). Post-hoc comparisons showed that level predictions with *sketchiness* were significantly further from the true values than for the other techniques,
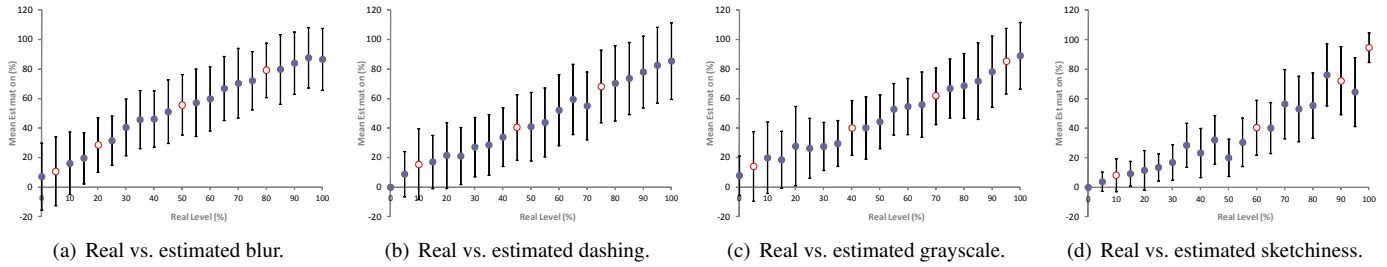
(a) Real vs. estimated blur.   (b) Real vs. estimated dashing.   (c) Real vs. estimated grayscale.   (d) Real vs. estimated sketchiness.

Fig. 9. Real vs. estimated visual variables (red cirles for distinct levels; error bars represent one standard deviation).

but this was true only after level ten, with no difference between techniques for levels up to ten (all $p < .05$). This evidence supports H4.

6.4.2   Level Groupings per Technique

We then tried to determine for each technique how many levels participants can distinguish comfortably and which level values are good candidates for an ordinal scale, assuming that our data are an accurate representation of real perception effects for the different visual variables. We thus conducted the following exploratory analysis:

We first plotted the mean and standard deviation of the *PerMag* of each level per technique (Fig. 9) and identified possible level values that are distinguishable between them. We chose as candidates levels that have similar standard deviations to their neighbors (to ensure we select representative levels for their neighborhood). Starting from the lower chosen level, we chose the next one such that the mean of the first did not overlap with the mean and the standard deviation of the next selected level, and so on. This process gave us initial estimates for the number of levels that can be distinguished per visual variable and the level value that corresponds to them (see Fig. 9(a)–(d) with the distinct levels being highlighted).

We then ran an ANOVA to see what levels were significantly different with respect to their perceived magnitude *PerMag*. We plotted the levels that are *not* significantly different on the matrices in Fig. 10 where gray squares indicate pairs of levels that were not significantly different and risk being perceived as similar (and white pairs that are significant). By inspecting these images of statistical results, we can see visual clusters of levels whose mean values are similar and thus should not all be used to represent two distinct levels. We ensured that our chosen levels (in black in Fig. 10) fall under different such visual clusters, to further supporting our level choices. In comparing the four matrices for the different techniques, *sketchiness* appears to have similar size groups but different structure especially towards the mid-scale. This may be attributed to the slightly higher perception error for *sketchiness* after level ten (as previously explained in Section 6.4.1). Our ANOVA was conducted with *visual variable* treated as a between-subjects variable and *level* as repeated measures. There was a main effect of technique on *AbsErr* ($F_{3,156} = 24.27$, $p < .0001$) and, more importantly for us, its interaction effect with level was also significant ($F_{60,3120} = 6.047$, $p < .0001$), indicating that level perception was different between variables. A post-hoc comparison showed levels that were clearly different in their perception from each visual variable technique (all $p < .05$) and are seen in white in the matrices.

We then ran a hierarchical clustering algorithm across level means for each visual variable to ensure that the levels we chose fall under different clusters found by the algorithm (Table 1). This was indeed the case, further supporting our level choices. Based on our generation process, in Table 1 we propose levels that are clearly distinguishable by technique. These levels are illustrated in Fig. 11. Indeed, at least three (four counting the maximum sketchiness level) levels were identified for *sketchiness* with four levels for blur and grayscale (supporting H5).

Finally, to further support our findings regarding the established levels per technique, we ran a within-subjects study on AMT with 40 participants. The task was to ascendingly order lines by visual variable level by assigning a suitable rank to each line (e. g., rank one for the least blurry line and rank four for the most blurry). The order of lines and techniques was randomized. Participants took on average

	Levels	Level Values (and their clusters)
blur	4	[0, **1**, 2, 3] [**4**,5,6,7,8,9] [**10**,11,12,13,14,15] [**16**,17,18,19,20]
dashing	3	[0, 1, **2**, 3, 4, 5, 6, 7, 8] [**9**,10,11,12, 13,14] [**15**,16,17,18,19,20]
grayscale	4	[0, **1**, 2, 3] [4,5,6,7,**8**,9,10] [11,12,13,**14**,16,17] [18,**19**,20]
sketchiness	3–4	[0, 1, **2**, 3,4,5,6,8,10] [7,9,11,**12**,13] [14,15,16,17,**18**,19] [**20**]

Table 1. Distinct levels, representative values and their clusters.

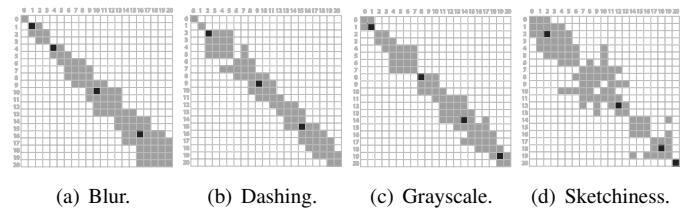(a) Blur.   (b) Dashing.   (c) Grayscale.   (d) Sketchiness.

Fig. 10. Black squares show our chosen representative levels, and gray ones indicate levels that are *not* significantly different.
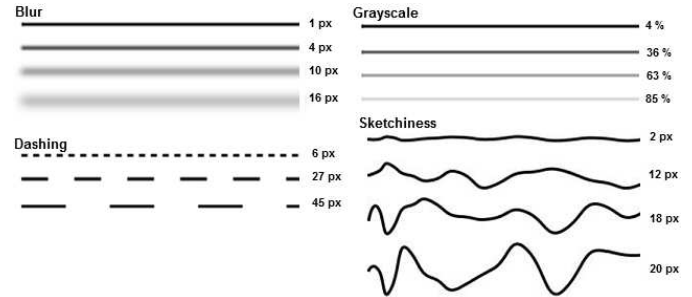

Fig. 11. Visual illustration of the established levels per technique.

126.3 seconds to complete all four parts of the study, where each part corresponds to one visual variable. Our results show that participants were able to recognize the correct order of levels (as established in Table 1) for all techniques with great accuracy (100% for both *blur* and *grayscale*, 95% for *dashing* and 90% for *sketchiness*). These results provide further validation for our established levels per technique.

6.4.3   Mathematical Description of Perceived Techniques

Previous work has attempted to mathematically describe the differences between physical and perceived magnitude of objects as collected from user studies. One popular function describing this difference is Stevens' [43] power law: $J = \lambda D^{\alpha}$, with $J = judged\ magnitude$, $D = actual\ magnitude$, $\alpha = exponent$, $\lambda = scaling\ constant$. Wagner [50] provides a meta-analysis of articles reporting values for α collected under different conditions. To the best of our knowledge, no conditions matched *dashing*, *blur*, or *sketchiness*; nevertheless there are values for *grayscale*. Stevens and Galanter [44] found an exponent of 1.2 for black-gray-white series. Given the previous discrepancies between the α varying across experiment setups [50], we decided to mathematically describe the perceived magnitude of all four different visual variables.

	$J = \lambda D^\alpha$		Regression	
	λ	α	R^2	Adjusted R^2
blur	3.149	0.729	0.993	0.993
dashing	2.188	0.774	0.974	0.973
grayscale	3.819	0.649	0.958	0.955
sketchiness	0.601	1.031	0.916	0.912

Table 2. Relationships and regression results. All shown results are statistically significant at the 99% level.

An initial curve fitting for the four techniques indicated that, indeed, a power model best fits our data (for all visual variables their respective fits had $R^2 > .9$, $p < .0001$). Using the parameter estimates from the fit we conducted a detailed regression analysis on our data to verify the mathematical relationship. Regression analysis for each of the techniques showed a very good fit (all $R^2 > .9$ and all adjusted $R^2 > .9$, $p < .0001$), as hypothesized (H6). Our results and the used coefficients are summarized in Table 2.

6.4.4 Discussion on Levels

We found that the perceived *sketchiness* is farther from the real value of a level than for the other techniques, but not by much (4%). This prediction error varied across techniques and was always $> 10\%$. This value is somewhat large and can be too prohibiting if participants attempt to accurately retrieve values from the visual representations of these techniques in real life applications (H4). We thus believe that all techniques should be used for ordinal scales but not for value retrieval.

As in previous studies, our perceived variables can be mathematically modeled using a power law (H6). The fact that our coefficients do not match others on *grayscale* as reported by Wagner [50] can be attributed to differences in study setups (as is often the case for perception experiments). An interesting observation is that the perception of sketchiness is fairly close to linear. This could be explained by our generation model that defines a sketchiness level based on the maximum distortion, as opposed to an average distortion across a line.

Using exploratory methods we were able to identify distinct levels for each of the visual variables (at least three for *sketchiness*—H5). We note that our selection is somewhat conservative. We base our selection on data that assume that each of the selected levels will be compared independently with the min/max values (e. g., when seeing a sketchy line in a visualization the user will immediately be able to determine if it indicates low, medium, or high uncertainty). Nevertheless, if viewers are presented with visual representations of other levels (e. g., comparing the *sketchiness* between two lines in a visualization, or between a line and a legend), it is possible they will be able to distinguish even more levels. This requires further investigation.

7 ON THE QUESTION OF PREFERENCE

One issue that still needs to be investigated is which techniques are preferred by people for visualizing uncertainty. Based on our own results from the intuitiveness study in Section 5 and also based on related work (notably on blur [27]) we had the following hypothesis:

H7 People prefer sketchiness to blur; but dashing and grayscale may be preferred overall due to people's familiarity with these styles.

7.1 Task

To judge people's preference about uncertainty encoding we asked participants to select one of four visual styles—*blur*, *dashing*, *grayscale*, or *sketchiness*—in a visualization scenario. To motivate participants into thinking about their choice and to avoid random answers we asked participants to justify their choice.

7.2 Participants, Study Design and Procedure

129 people participated in this study, using a between-subjects design. Participants were split into six groups of equal size. Each group was exposed to one of the six different scenarios from our first study (Section 5). Participants were first introduced to the task, and a short description of the scenario was given. They were then presented with four side-by-side versions of the same image which only differed in the rendering of a single line (depicted using *blur*, *dashing*, *grayscale*, or *sketchiness*). All values of the visual variables were selected from the middle ranges for consistency. Participants were informed that the study is about comparing different styles to show uncertainty using the aforementioned visual variables. They were asked to choose their preferred style and were requested to justify this choice. We highlighted the uncertain line in question with a red arrow in each image.

Overall our experiment consisted of:

	6	scenarios
×	32	participants per scenario
=	**192**	**trials in total**

7.3 Results

We ran our study in batches of 48 trials. For each batch we calculated the mean for each visual variable. After the third batch the means stabilized. Our results show that *dashing* is the preferred style for the participants (chosen by 68.3%), in agreement with our hypothesis (H7). However, *blur* (chosen by 15.10%) did better than both *grayscale* (chosen by 12.5%) and *sketchiness* (chosen by 3.12%).

We were able to process comments from 155 participants (80.7% of the responses). A closer look at their comments shows that the primary reason for preferring *dashing* over the other visual variables was '*noticeability*' as participants valued the ability to easily distinguish between data and uncertainty (26.5% of our participants), whilst *blur* was chosen because it was regarded as congruent to what uncertainty conveys in terms of vagueness and reduced precision (4.5% of our participants). Those who preferred *grayscale* did so because they deemed it easy to understand (2% of our participants).

We only had four comments on the use of *sketchiness*, two of which argued that it is an intuitive way to represent instability in the data. More interesting, and which may explain the low preference score for sketchiness, are comments by participants who argued against the use of *sketchiness*. The primary reason for this seems to be that *sketchiness* tends to imply a disliked notion of 'unprofessionalism' or informality. Some participants thus disregarded it as a standard visual variable (4.5%). This highlights the need for paying attention to the context in which *sketchiness* is applied as a visual variable to denote uncertainty. We note that abstract and non-abstract contexts did not show different results in terms of user preference. Application contexts however may be a more influential factor in terms of impact on user subjective influence as will be discussed in the next section.

8 DISCUSSION AND CONCLUSION

In this paper, we have reported on seven user studies to investigate the appropriateness of using *sketchiness* as a dedicated visual variable for depicting uncertainty. We started by studying characteristics of collected hand-drawn strokes. Findings from this evaluation and existing work on a generative model of hand movement fed into our method for synthesizing sketchy lines that mimic stroke characteristics of hand-drawn lines. Our mapping for uncertainty depiction consists of varying the amount of pixel deviations from a straight line in accordance to the amount of uncertainty in the data. An online study using Amazon Mechanical Turk provided evidence that lines generated by our model were significantly more often thought to be handwritten than not.

8.1 *Sketchiness* as a Visual Variable

We then investigated whether *sketchiness* is an appropriate visualization technique for representing uncertainty. The three questions we attempted to answer were:

(a) Can people intuitively understand sketchiness *as an indicator of uncertainty?* In a series of studies we found that if users are asked what a sketchy line represents, their most common reaction is that it is associated with a different semantic relation from what is conveyed by a straight line (other than uncertainty). When given a multiple choice of possible explanations for sketchiness, the number of people who associated *sketchiness* with uncertainty almost doubled. This was lower

than expected; nevertheless, a similar study showed a very similar effect for *blur*, a visual variable which is traditionally associated with uncertainty in visualizations. Thus, although for the majority of people *sketchiness* was not associated with uncertainty, we feel that given a legend, *sketchiness* is a viable alternative to other visualization techniques such as *blur*, *dashing*, or *grayscale*, which can now be used to represent other types of information in the data.

(b) Is sketchiness *as effective for depicting uncertain information as the other visualization techniques?* Our results were encouraging for practically applying *sketchiness* as a visual variable to information visualizations. We compared *sketchiness* with other visual variables in a magnitude estimation task (it is very common in information visualization, e. g., to compare a given object to a legend). Our studies showed that *sketchiness* was only slightly more error-prone than other techniques. In fact, none of the techniques was very accurately perceived. This indicates that none of them is particularly well suited for true value retrieval, and the common practice of using them in small ordinal scales of uncertainty is indeed a sound one. Our mathematical model for predicting all visual variables indicates that this error is predictable for all techniques (including *sketchiness*). To help uncertainty visualization designers we proposed, based on our findings, a number of levels for ordinal scales for each visual variable (three or four excluding maxima). The number of levels we identified and proposed are close in number between *sketchiness* (three) and other techniques. Our process for selecting levels is based on the assumption that estimates are made when comparing to a maximal value object.

(c) Which method do people subjectively prefer? Sketchiness was not the preferred visual variable to encode uncertainty (*dashing* was preferred). Moreover, and in the context of our provided scenarios, some participants commented negatively on the 'informal' and 'unprofessional' look of sketchy lines. There are, however, some contexts where informality is advantageous and thus where *sketchiness* may be a more preferable choice such as for conveying initial design ideas [42]. We thus conclude that *sketchiness* as a visual variable is a viable additional choice for depicting uncertainty in ordinal data, but further work is needed to investigate the appropriate visualization contexts where it can be deployed.

8.2 Contexts and Design Guidelines

Translating our findings into guidelines, we recommend *sketchiness* for conveying *qualitative* uncertainty on lines with certain caveats:

- for uncertainty depiction, *sketchiness*—as well as the other variables we studied—require a legend;
- *sketchiness* is as intuitive as *blur*;
- *sketchiness* is not suitable for spatial contexts such as maps to indicate uncertainty, because changes in the geometry tend to be perceived as related to the actual underlying spatial features;
- conversely, *sketchiness* is more appropriate in abstract contexts such as hierarchies and diagrams where accurate readings of intermediate line points is not required, but not for line graphs for instance where geometry is also perceived as related to values change;
- sketchy lines with high perturbations occupy more space and may increase clutter; therefore, *sketchiness* with many levels of encoding may not be suitable for dense displays (the same applies for *dashing*);
- *sketchiness*—as well as the other variables we studied—can only convey 3–4 levels reliably;
- *sketchiness* is perceived as informal; this may be desirable in some contexts (e. g., when showcasing preliminary designs) and less in others; and
- *sketchiness* is not the preferred encoding for uncertainty but is acceptable unless the context requires formality.

8.3 Limitations and Future Work

We do not make the claim that our results are generalizable to all kinds of sketchy lines; our notion of sketchiness is specifically defined by a range of pixel deviations from a straight-line and implemented using a sampling model that is max-biased. Other methods to generate sketchy lines—for instance the mean-biased method—may yield different results notably for the number of levels that can be distinguished by the viewer. *Sketchiness* is a promising technique and we hypothesize that it can be extended beyond the tested deviation range to potentially support more levels.

To determine the number of levels that can be visually distinguishable we asked participants to carry out a low-level perceptual task. It would be interesting to test for higher-level tasks such as search tasks in various information visualization contexts; and to study the effect of shape and area on the perception of the sketchiness level. Another extension to this study would be to control for line width, especially for *dashing*, where it is likely that both the length and width of the line play a role in the perceptibility of its connectedness; and to add repeated trials in order to show the stability of the perceived levels.

Finally, we focused in this paper on *sketchiness* as a visual variable to encode uncertainty information; it would be interesting to investigate the link between sketchiness and engagement, in particular in cases where informality is advantageous, such as for conveying initial design ideas. Such a more general look at sketchy rendering in the context of information visualization has recently been carried out [54] which looks beyond our focus on uncertainty and also investigates higher-level tasks, indicating that indeed engagement is affected.

REFERENCES

[1] J. C. J. H. Aerts, K. C. Clarke, and A. D. Keuper. Testing Popular Visualization Techniques for Representing Model Uncertainty. *Cartography and Geographic Information Science*, 30(3):249–261, July 2003. doi> 10.1559/152304003100011180

[2] R. S. Allendes Osorio and K. W. Brodlie. Contouring with Uncertainty. In *Theory and Practice of Computer Graphics*, pp. 59–65. Eurographics Association, Goslar, Germany, 2008. doi> 10.2312/LocalChapterEvents/TPCG/TPCG08/059-065

[3] Z. AlMeraj, B. Wyvill, T. Isenberg, A. A. Gooch, and R. Guy. Automatically Mimicking Unique Hand-Drawn Pencil Lines. *Computers & Graphics*, 33(4):496–508, Aug. 2009. doi> 10.1016/j.cag.2009.04.004

[4] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. ESRI Press, Redlands, California, 2010.

[5] A. M. Bisantz, T. Kesevadas, P. Scott, D. Lee, S. Basapur, P. Bhide, P. Bhide, and P. Bhide. Holistic Battlespace Visualization: Advanced Concepts in Information Visualization and Cognitive Studies. Technical report, UB Engineering, University of Buffalo, USA, June 2002.

[6] D. Borland and Russell. Rainbow Color Map (Still) Considered Harmful. *IEEE Computer Graphics and Applications*, 27(2):14–17, Mar./Apr. 2007. doi> 10.1109/MCG.2007.323435

[7] R. P. Botchen, D. Weiskopf, and T. Ertl. Texture-Based Visualization of Uncertainty in Flow Fields. In *Proc. IEEE Visualization*, pp. 647–654. IEEE, Los Alamitos, 2005. doi> 10.1109/VISUAL.2005.1532853

[8] N. Boukhelifa and D. J. Duke. Uncertainty Visualization – Why Might it Fail? In *CHI Extended Abstracts*, pp. 4051–4056. ACM, New York, 2009. doi> 10.1145/1520340.1520616

[9] R. Brecheisen, B. Platel, B. ter Haar Romeny, and A. Vilanova. Illustrative Uncertainty Visualization of DTI Fiber Pathways. *The Visual Computer*, 2012. To appear. doi> 10.1007/s00371-012-0733-9

[10] R. Brown. Animated Visual Vibrations as an Uncertainty Visualisation Technique. In *Proc. GRAPHITE*, pp. 84–89. ACM, New York, 2004. doi> 10.1145/988834.988849

[11] A. Cedilnik and P. Rheingans. Procedural Annotation of Uncertainty Information. In *Proc. IEEE Visualization*, pp. 77–84. IEEE Computer Society, Los Alamitos, 2000. doi> 10.1109/VISUAL.2000.885679

[12] W. S. Cleveland and R. McGill. Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science*, 229(4716):828–833, Aug. 1985. doi> 10.1126/science.229.4716.828

[13] C. D. Correa, Y.-H. Chan, and K.-L. Ma. A Framework for Uncertainty-Aware Visual Analytics. In *Proc. VAST*, pp. 51–58. IEEE Computer Society, Los Alamitos, 2009. doi> 10.1109/VAST.2009.5332611

[14] C. Curtis. Loose and Sketchy Animation. In *SIGGRAPH Technical Sketches*, p. 317. ACM, New York, 1998. doi> 10.1145/281388.281913

[15] S. Diepenbrock, J.-S. Prani, F. Lindemann, H.-W. Bothe, and T. Ropinski. Interactive Visualization Techniques for Neurosurgery Planning. In *Eurographics Short Papers / Dirk Bartz Prize for Visual Computing in Medicine*, pp. 13–16. Eurographics Association, Goslar, Germany, 2011.

[16] A. Finkelstein and D. H. Salesin. Multiresolution Curves. In *Proc. SIG-GRAPH*, pp. 261–268. ACM, New York, 1994. doi> 10.1145/192161.192223

[17] T. Flash and N. Hogans. The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model. *The Journal of Neuroscience*, 5(7):1688–1703, July 1985.

[18] L. E. Gerharz and E. J. Pebesma. Usability of Interactive and Non-Interactive Visualisation of Uncertain Geospatial Information. In *Proc. Geoinformatik*, pp. 223–230. ifgiPrints, Münster, Germany, 2009.

[19] N. D. Gershon. Visualization of Fuzzy Data Using Generalized Animation. In *Proc. IEEE Visualization*, pp. 268–273. IEEE, Los Alamitos, 1992. doi> 10.1109/VISUAL.1992.235199

[20] E. B. Goldstein. *Sensation and Perception*. Brooks/Cole Publishing, Pacific Grove, USA, 5ᵗʰ edition, 1999.

[21] B. Gooch and A. A. Gooch. *Non-Photorealistic Rendering*. A K Peters, Ltd., Natick, 2001.

[22] H. Griethe and H. Schumann. The Visualization of Uncertain Data: Methods and Problems. In *Proc. SimVis*, pp. 143–156. SCS Publishing House e.V., Erlangen, Germany, 2006.

[23] J. R. Hamerly and C. A. Dvorak. Detection and Discrimination of Blur in Edges and Lines. *Journal of the Optical Society of America*, 71(4):448–452, Apr. 1981. doi> 10.1364/JOSA.71.000448

[24] J. Heer and M. Bostock. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proc. CHI*, pp. 203–212. ACM, New York, 2010. doi> 10.1145/1753326.1753357

[25] T. Hengl and D. J. J. W. A. Brown. Pixel and Colour Mixture: GIS Techniques for Visualisation of Fuzziness and Uncertainty of Natural Resource Inventories. In *Proc. Accuracy*, pp. 300–308. Delft University Press, Delft, the Netherlands, 2002.

[26] T. Hengl and N. Toomanian. Maps Are Not What They Seem: Representing Uncertainty in Soil-Property Maps. In *Proc. Accuracy*, pp. 805–813. Instituto Geográfico Português, Lisboa, Portugal, 2006.

[27] R. Kosara, S. Miksch, H. Hauser, J. Schrammel, V. Giller, and M. Tscheligi. Useful Properties of Semantic Depth of Field for Better F+C Visualization. In *Proc. VisSym*, pp. 205–210. Eurographics Association, Goslar, Germany, 2002.

[28] M. Leitner and B. P. Buttenfield. Guidelines for the Display of Attribute Certainty. *Cartography and Geographic Information Science*, 27(1):3–14, Jan. 2000. doi> 10.1559/152304000783548037

[29] H. Li, C.-W. Fu, Y. Li, and A. J. Hanson. Visualizing Large-Scale Uncertainty in Astrophysical Data. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1640–1647, Nov./Dec. 2007. doi> 10.1109/TVCG.2007.70620

[30] M. Luboschik, A. Radloff, and H. Schumann. Using NPR-Rendering Techniques for the Visualization of Uncertainty. In *Posters of IEEE InfoVis*. IEEE Computer Society, Los Alamitos, 2010.

[31] A. M. MacEachren. Visualizing Uncertain Information. *Cartographic Perspectives*, 13(Fall):12–19, 1992.

[32] A. M. MacEachren, A. Robinson, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32(3):139–160, July 2005. doi> 10.1559/1523040054738936

[33] M. Matthews, L. Rehak, A.-L. Lapinski, and S. McFadden. Improving the Maritime Surface Picture with a Visualization Aid to Provide Rapid Situation Awareness of Information Uncertainty. In *Proc. IEEE TIC-STH*, pp. 533–538. IEEE, Los Alamitos, 2009. doi> 10.1109/TIC-STH.2009.5444441

[34] M. Nienhaus, F. Kirsch, and J. Döllner. Sketchy Illustrations for Presenting the Design of Interactive CSG. In *Proc. IV*, pp. 772–777. IEEE Computer Society, Los Alamitos, 2006. doi> 10.1109/IV.2006.97

[35] A. Pang. Visualizing Uncertainty in Geo-spatial Data. In *Proc. Workshop on the Intersections between Geospatial Information and Information Technology*, 2001.

[36] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to Uncertainty Visualization. *The Visual Computer*, 13(8):370–390, Nov. 1996. doi> 10.1007/s003710050111

[37] J. J. Pfeiffer, Jr. Using Brightness and Saturation to Visualize Belief and Uncertainty. In *Proc. Diagrams*, pp. 279–289. Springer-Verlag, Berlin, 2002. doi> 10.1007/3-540-46037-3_27

[38] K. Potter, A. Gooch, B. Gooch, P. Willemsen, J. Kniss, R. Riesenfeld, and P. Shirley. Resolution Independent NPR-Style 3D Line Textures. *Computer Graphics Forum*, 28(1):52–62, Mar. 2009. doi> 10.1111/j.1467-8659.2008.01297.x

[39] P. J. Rhodes, R. S. Laramee, R. D. Bergeron, and T. M. Sparr. Uncertainty Visualization Methods in Isosurface Rendering. In *EG Short Papers*, pp. 83–88. Eurographics Association, Goslar, Germany, 2003.

[40] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. J. Moorhead. A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1209–1218, Nov./Dec. 2009. doi> 10.1109/TVCG.2009.114

[41] S. Schlechtweg, B. Schönwälder, L. Schumann, and T. Strothotte. Surfaces to Lines: Rendering Rich Line Drawings. In *Proc. WSCG*, volume 2, pp. 354–361, 1998.

[42] J. Schumann, T. Strothotte, A. Raab, and S. Laser. Assessing the Effect of Non-photorealistic Rendered Images in CAD. In *Proc. CHI*, pp. 35–42. ACM, New York, 1996. doi> 10.1145/238386.238398

[43] S. S. Stevens. *Psychophysics*. Transaction Publishers, New Brunswick, USA, 2ⁿᵈ edition, 1975.

[44] S. S. Stevens and E. H. Galanter. Ratio Scales and Category Scales for a Dozen Perceptual Continua. *Journal of Experimental Psychology*, 54(6):377–411, Dec. 1957. doi> 10.1037/h0043680

[45] T. Strothotte, M. Masuch, and T. Isenberg. Visualizing Knowledge about Virtual Reconstructions of Ancient Architecture. In *Proc. CGI*, pp. 36–43. IEEE, Los Alamitos, 1999. doi> 10.1109/CGI.1999.777901

[46] T. Strothotte, B. Preim, A. Raab, J. Schumann, and D. R. Forsey. How to Render Frames and Influence People. *Computer Graphics Forum*, 13(3):455–466, Aug. 1994. doi> 10.1111/1467-8659.1330455

[47] T. Strothotte, M. Puhle, M. Masuch, B. Freudenberg, S. Kreiker, and B. Ludowici. Visualizing Uncertainty in Virtual Reconstructions. In *Proc. EVA Europe*, p. 16. EVA Conferences International/GFaI, Berlin, 1999.

[48] T. Strothotte and S. Schlechtweg. *Non-Photorealistic Computer Graphics. Modeling, Animation, and Rendering*. Morgan Kaufmann Publishers, San Francisco, 2002.

[49] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. A Typology for Visualizing Uncertainty. In *Proc. Visualization and Data Analysis*, pp. 146–157. SPIE, 2005. doi> 10.1117/12.587254

[50] M. Wagner. *The Geometries of Visual Space*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2006.

[51] R. Watt and M. Morgan. The Recognition and Representation of Edge Blur: Evidence for Spatial Primitives in Human Vision. *Vision Research*, 23(12):1465–1477, 1983. doi> 10.1016/0042-6989(83)90158-X

[52] W. Willett, J. Heer, and M. Agrawala. Scented Widgets: Improving Navigation Cues with Embedded Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, Nov. 2007. doi> 10.1109/TVCG.2007.70589

[53] G. A. Winkenbach and D. H. Salesin. Computer-Generated Pen-and-Ink Illustration. In *Proc. SIGGRAPH*, pp. 91–100. ACM, New York, 1994. doi> 10.1145/192161.192184

[54] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. Sketchy Rendering for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), Nov./Dec. 2012. In this issue.

[55] Z. Xie, S. Huang, M. O. Ward, and E. A. Rundensteiner. Exploratory Visualization of Multivariate Data with Variable Quality. In *Proc. VAST*, pp. 183–190. IEEE, Los Alamitos, 2006. doi> 10.1109/VAST.2006.261424

[56] B. Zehner, N. Watanabe, and O. Kolditz. Visualization of Gridded Scalar Data with Uncertainty in Geosciences. *Computers & Geosciences*, 36(10):1268–1275, Oct. 2010. doi> 10.1016/j.cageo.2010.02.010

[57] T. Zuk and S. Carpendale. Theoretical Analysis of Uncertainty Visualizations. In *Visualization and Data Analysis, Proc. SPIE-IS&T Electronic Imaging*, pp. 606007/1–14. SPIE, 2006. doi> 10.1117/12.643631