DOI: 10.1111/j.1467-8659.2012.03094.x
Eurographics Conference on Visualization (EuroVis) 2012
S. Bruckner, S. Miksch, and H. Pfister
(Guest Editors)

*Volume 31 (2012), Number 3*

# Conceptualizing Visual Uncertainty in Parallel Coordinates

Aritra Dasgupta[1] Min Chen[2] Robert Kosara[1]

[1] UNC Charlotte, {adasgupt, rkosara}@uncc.edu, [2] Oxford e-Research Centre, min.chen@oerc.ox.ac.uk

**Abstract**

*Uncertainty is an intrinsic part of any visual representation in visualization, no matter how precise the input data. Existing research on uncertainty in visualization mainly focuses on depicting data-space uncertainty in a visual form. Uncertainty is thus often seen as a problem to deal with, in the data, and something to be avoided if possible. In this paper, we highlight the need for analyzing visual uncertainty in order to design more effective visual representations. We study various forms of uncertainty in the visual representation of parallel coordinates and propose a taxonomy for categorizing them. By building a taxonomy, we aim to identify different sources of uncertainty in the screen space and relate them to different effects of uncertainty upon the user. We examine the literature on parallel coordinates and apply our taxonomy to categorize various techniques for reducing uncertainty. In addition, we consider uncertainty from a different perspective by identifying cases where increasing certain forms of uncertainty may even be useful, with respect to task, data type and analysis scenario. This work suggests that uncertainty is a feature that can be both useful and problematic in visualization, and it is beneficial to augment an information visualization pipeline with a facility for visual uncertainty analysis.*

## 1. Introduction

Uncertainty is a twofold problem in visualization. On one hand, it is important for visualization to convey uncertainty in the data to the users, giving rise to the quest for effective means to measure and visually depict uncertainty [JS03]. On the other hand, the visualization process itself will introduce uncertainty. The former is primarily concerned with uncertainty in the data space, while the latter has to address sources of uncertainty in visual mapping, rendering, displaying, viewing, perception, understanding, and reasoning. While much of the existing work in the visualization literature focuses on data uncertainty [WPL96, PWL97], discussions on uncertainty stemming from the visualization process itself are still limited. In scientific visualization, there is not always a clear boundary between data uncertainty and visual uncertainty, since the visualization process often involves the manipulation of geometric primitives (e.g., errors in isosurface extraction [RLBS03, LPSW96] or in particle tracing [LB98]). Even when such geometric abstraction is considered as part of visual uncertainty, it represents only one specific type of uncertainty caused by the visualization itself. The aim of this work is to highlight the fact that there are many other types of uncertainty sources in the visualization process.

We adopt a case-based research methodology by focusing on a specific class of non-spatial data visualization, namely *parallel coordinates visualization*, which is a powerful tool for visualizing and analyzing multi-dimensional data [ID90]. Almost everyone who has used parallel coordinates has seen the dreaded black screen which is composed of over-plotted lines and conveys a high level of uncertainty but a very limited amount of useful information. In practice, many visualizations contain more subtle forms of uncertainty. For example, axes with few values create focal points where many lines meet, but it is uncertain how lines continue to the next axis; over-plotting of lines that are very close together makes it impossible to tell exactly how many points are in a particular location; the inherent resolution of the pixel grid limits the perceivable resolution of the data; etc. By focusing on a specific class of visualization, we are able to conduct a detailed analysis of a manageable set of sources and effects of uncertainty and their relationships. We believe that this methodology and the major findings of this work can also be applied to other classes of visual representations.

One of the many complexities in designing a visualization system is to properly address the trade-off: how to satisfy perceptual design principles while at the same time maintaining data fidelity during visual mapping. For large datasets, striking a balance between information loss and visual quality presents a considerable challenge. In the current visualization literature, we lack a comprehensive understanding of all the different variables and parameters in a visualization and how they might interact with each other. A theoretical foundation of *visual uncertainty* will enable a better understanding of the interplay between data and visualization properties, to build more effective means for translating data objects to visual objects [War04].

Our proposed taxonomy (Figure 1) was built from the bottom up, by considering cases of uncertainty that we observed in real parallel coordinates visualizations (Figures 2–9). Similar to the conventional way of building a taxonomy, we organized several classification schemes into a hierarchy. When an example case fell into several sub-classes, we placed it only in the most relevant category to reduce the complexity of the taxonomy.

Our contributions can be summarized as follows:

- We propose a new *taxonomy of visual uncertainty* in the context of line-based and cluster-based parallel coordinates visualization. We identify and classify various *causes* of uncertainty, relate them to the *effects* and also identify which stage of the information visualization pipeline is the source of uncertainty.
- We apply the developed taxonomy to a selection of techniques for improving parallel coordinates visualization and analyze their relative merits in relation to visual uncertainty. This demonstrates that the taxonomy can be used as a qualitative framework to evaluate visualization techniques in a structured manner.
- We identify cases where certain effects of uncertainty are useful for data analysis or even need to be introduced intentionally, like in the case of privacy-preserving visualization.

## 2. Related Work

For conceptualizing visual uncertainty, we provide context to our work by discussing the existing schemes of uncertainty, particularly those used in visualization.

### 2.1. Uncertainty in Visualization

There exists a plethora of discussions in the literature on classifying and categorizing uncertainty, for instance, in statistical forecasting, risk analysis, philosophy and psychology. For a high-level classification of uncertainty in parallel coordinates, we take into account the different perspectives provided on uncertainty by Milliken [Mil87], Norvig [RNC*95]; and Klir and Wierman [KW99]. We also refer to the typology proposed by Thomson et al. [THM*05] for relating to data-space uncertainty. While describing our taxonomy, we examine some of these classification schemes in the next section.

Most existing work in visualization relates to data-space uncertainty (e.g., [PWL97, SLSR09]) and uncertainty involving geometrical primitives, like isosurface rendering. Our conceptualization of visual uncertainty applies in case of abstract data where a spatial context is not given [TM04].

### 2.2. Visual Quality in Parallel Coordinates

*Visual quality* is a related concept to visual uncertainty. Different visual quality metrics have been proposed for analysis of multi-dimensional data, especially for parallel coordinates [DK11, TA*09]. Earlier, Bertini and Santucci [BS06] have argued for a visual optimization principle that guides the evaluation of different visualizations, facilitated through visual metrics. While there have been sporadic mentions of visual metrics in the literature, we lack a systematic approach to the problem. Bertini et al. [BTK11] point out the lack of support for meta-visualization, which would help in verifying and validating a technique and also help visualization designers better understand what works and why. Although different approaches to measuring visual quality have been suggested in the literature, the lack of objective definitions make it a difficult task to categorize the existing literature. We believe our conceptualization of visual uncertainty will provide an effective way of deconstructing complex visualizations, such as parallel coordinates, by identifying low-level causes of uncertainty and relating them to high-level concepts of perception and cognition.

## 3. Visual Uncertainty

Uncertainty is an intuitive term, with a variety of definitions and ways of measuring it. However, none of them captures the nature of visual uncertainty, mainly because it is a combination of issues in a mechanical process (the visualization pipeline), quantization (the pixel grid), and perceptual limitations (low-level vision as well as cognitive abilities and limitations). We therefore need a more nuanced view of uncertainty in visualization, which goes beyond a single definition or metric.

### 3.1. Physical vs. Perceptual Uncertainty

Two existing schemes of uncertainty are most relevant for deciding our top-level classification. One such categorization is to consider uncertainty in physical systems (physical uncertainty) and that in the human mind (perceptual uncertainty) separately. For example, in behavioral sciences and neuroscience, there is an assumption that the nervous system performs its own probabilistic estimation about events in an environment, resulting in perceived certainty or uncertainty. Such results usually differ from those obtained from measurement of the events in the environment. This is partially true in case of mixed-initiative visualization systems as the data is first processed in physical systems, on the machine side, after which the human side takes over. On the human side, we have to take perception into account as well;

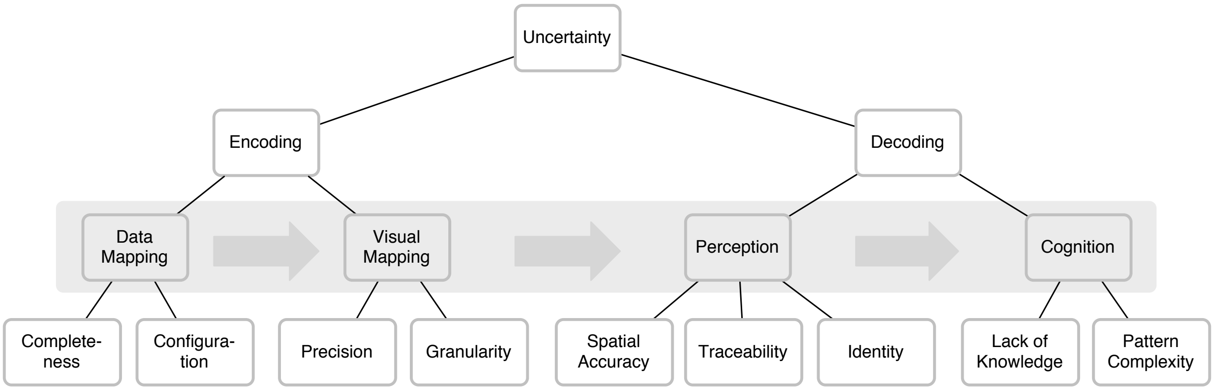


**Figure 1:** *The proposed taxonomy of visual uncertainty in parallel coordinates. The shaded area indicates the level at which it maps to the stages in typical visualization pipelines.*

uncertainty due to perception has been discussed by Russell and Norvig [RNC*95]. Holzhüter et al. [HLS*12] describe uncertainty in visualization and differentiate between input and output uncertainty. Relating the information visualization pipeline to the communication channel as discussed below, we choose encoding and decoding uncertainty to be the topmost classifying schemes.

### 3.2. The Communication Model of Uncertainty

Shannon defined information as a measure of the decrease of uncertainty for the receiver of a message [Sha48]. If visualization is viewed as a communication channel from the data space to the perceptual and cognitive mental space of the user [PAJKW08], it is important to trace the uncertainty along different stages of the pipeline, so that the information communicated to the user can be optimized. Communication of information is not the end-all in visualization, as a significant amount of transformation takes place in the mental space of the user. We believe, however, that visual communication is as important as the transformation that follows it. Like a communication channel, visualization is also associated with the encoding and decoding of information. We define *visual uncertainty* as the uncertainty that is associated with a visualization during encoding (in the screen-space) and decoding of information (in the mental space of the user).

### 3.3. Taxonomy Overview

Our taxonomy (Figure 1) separates the causes of uncertainty into two main groups: encoding (Section 4) and decoding (Section 5). The typical way of looking at uncertainty is from a decoding perspective, which includes our perceptual and cognitive processes when working with a visualization. Uncertainty is also introduced on the encoding side, however, through transformations of the data, mapping to the pixel grid, or selections of data and axes.

Uncertainty is usually the result of a number of causes, but we have attempted to narrow down the main reasons for uncertainty in specific cases. Most real-world scenarios will consist of combinations of these cases, and even within the taxonomy there is some overlap between some of the higher levels and the specific examples. As a working definition, we adopt the definition of Douglas Hubbard [Hub10], which describes uncertainty as *the lack of certainty, a state of having limited knowledge where it is impossible to exactly describe existing state or future outcome, more than one possible outcome*.

The third level of the taxonomy coincides with the stages found in visualization pipeline models like Chi's [Chi00]: data mapping and visual mapping. We add two stages on the human side of the pipeline, perception and cognition; while they are not very clearly delineated, we find them useful to structure the lowest level of the taxonomy.

## 4. Encoding Uncertainty

As data moves through the visualization pipeline, it gets transformed and mapped to visual coordinates and shapes. The encoding side of our taxonomy includes all the stages from data access to rendering the visualization on screen. Data acquisition and any uncertainty inherent in it is outside the scope of this work.

### 4.1. Data Mapping

In the first stage of the visualization pipeline, the user selects the data points and dimensions that are to be mapped onto the screen. In addition to the selection of the dimensions, their ordering is also determined, which is important for the patterns that will be visible once the visualization is drawn onto the screen. In contrast to data-space uncertainty, this process is entirely driven by the user, who picks which elements to show (usually in response to what is currently

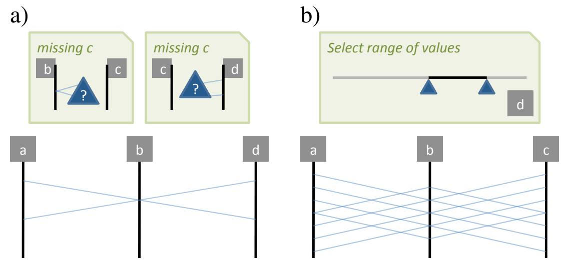


**Figure 2:** ***Completeness:*** *Choosing not to include an entire axis (a) or single values (b) prevents the user from seeing some of the data, causing uncertainty about it.*

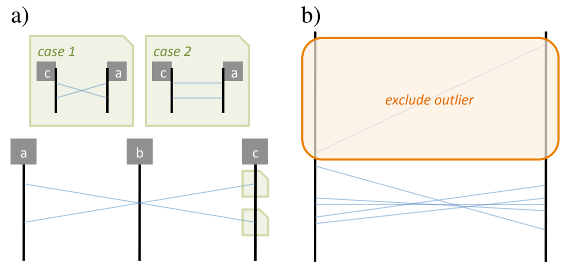


**Figure 3:** ***Configuration:*** *a) Patterns can be missed when not all possible pairs of axes are represented; b) leaving out individual data values prevents the user from seeing parts of the data;.*

shown on the screen). This process seems benign and simple, but there are many possible configurations, many of which hide potentially interesting parts of the data.

**Completeness:** While parallel coordinates can show many dimensions at once, many high-dimensional datasets are still impractical to show all at once, or the user may choose to show a smaller number to gain more space per dimension. By leaving out dimensions, potentially interesting structures are not shown on screen, causing uncertainty about the complete set of patterns in the data (Figure 2a). It is also possible to filter the data on a dimension that is not part of the visualization (Figure 2b). The most common case for doing this is when there is a time dimension in the data, in which case the visualization shows the data for only one particular time step. When not all records are shown, patterns can be hidden that would be apparent if all the data was there, resulting in further uncertainty.

**Configuration:** Even if all the axes are shown, their order is crucial to see patterns: most patterns are only visible between directly adjacent axes. Not only is it typically not feasible to try out all possible axis orderings, it is also uncommon to show the same axes several times in the same visualization [Weg90]. The wrong choice of axis ordering can thus hide important patterns without the user being able to find out what he or she is missing, leading to uncertainty (Figure 3a). A common interaction in parallel coordinates is the exclusion of outliers on an axis, typically to give the remaining data more space (Figure 3b). In contrast to the completeness case above, the missing data is chosen by visual criteria, and is typically still shown on the other dimensions (and as a line that is leaving the screen). The exact values of those outliers are lost, however.

### 4.2. Visual Mapping

When the data is drawn onto the screen in the visual mapping and rendering stage (which we treat as one step), its uncertainty increases due to the limited resolution of the pixel grid. The application of information theoretic metrics in quantifying the screen-space artifacts has been discussed by Chen and Jänicke [CJ10]. In parallel coordinates, a variety of artifacts are produced both on the axes and between them. While these also cause issues on the perception side of the taxonomy, there are really two separate phenomena at play here that need to be distinguished. The visual mapping side is also easier to assess due to its mechanical nature than the much more complex perception side.

**Precision:** The limited number of pixels on a display causes the locations of the data points to be quantized into a relatively small number of distinct values. In most real datasets, many data points end up getting mapped to the same pixel locations, and thus can no longer be differentiated. The information lost at this stage leads to uncertainty about the precise values of the data points (Figure 4a). When transparency is used, the colors of lines also mix, making it difficult to tell how many and which values are present. This is especially true when color is also used, such as for a gradient on one axis to more easily spot correlations. Even given perfect color perception, it is impossible to decode the resulting colors due to the limited resolution of the color values represented on the screen, and the resulting quantization of the colors (Figure 4b).

**Granularity:** Clustering naturally introduces uncertainty into the data, by reducing the number of values and representing them only as cluster boundaries or centroids and sizes. We are interested in the visual appearance of clusters between axes when they can be shown as polygons [NH06, DK11]. Just as in data space, the visual clusters hide the individual lines, thus removing information about the distribution of lines within the cluster, and even the number of lines in each cluster (Figure 5a). A similar issue occurs on the axes, where the locations of the points are no longer known, even when the cluster boundaries are defined by the maximum and minimum axial values of points in cluster (Figure 5b). In that case, it is not known whether the corners defining the cluster belong to the same data point or to different ones; several combinations are possible that are all equally likely (one line can run along the boundary or the

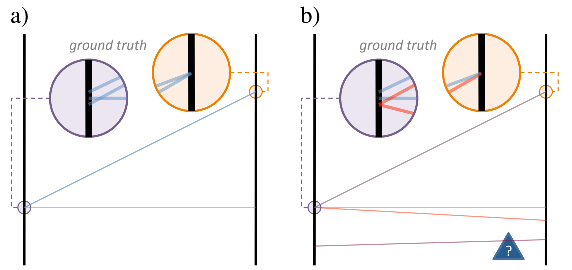


**Figure 4:** ***Precision:*** *a) Pixel binning leads to a loss in precision, which makes it impossible to read values precisely; b) the colors of lines drawn over each other make it difficult to see brushing and the precise number of lines (when transparency is used).*

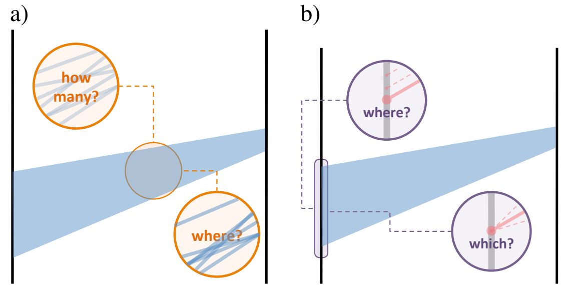


**Figure 5:** ***Granularity:*** *a) Clustering of values hides information about the internal structure of the cluster and potentially the number of items in each cluster; b) the same is true for the internal structure of the cluster and the actual locations of the original data points.*

boundary can consist of two distinct data points, for both boundaries independently).

## 5. Decoding Uncertainty

Once the information is encoded and the visualization rendered to the screen, the perceptual and cognitive processes of the user take over in interpreting that information. Decoding uncertainty occurs in the perception and cognition stages of our pipeline. We consider a source of uncertainty to be in the decoding branch only if the information concerned is fully encoded in the visualization. Without the information having been encoded first, it cannot be decoded, thus we give priority to the encoding stage. Analysis of decoding uncertainty enables us to evaluate a visualization technique by asking questions such as: is it perceptually confusing, does it incur a high level of cognitive load for reasoning, or is it only suitable for expert users who know how to interpret the visual representation?

### 5.1. Perception

In this section, we consider visual uncertainty resulting from the limits of the human vision system. Higher-level processes such as knowledge and the ability to perceive and recognize patterns are discussed in Section 5.2 on cognition.

**Spatial Accuracy:** The lack of knowledge about the exact spatial location of terminators (such as where records within a cluster are located or where the attributes are on an axis) or other geometric features (such as where two lines cross each other) causes uncertainty about the precise data represented. Perceptual accuracy concerns whether the user can differentiate visual objects from available information such as locations and colors. Sometimes, although the information is theoretically there on the screen, it can still be perceptually very difficult to perceive such information due to either the discriminative limit of the human vision system or perceptual illusion. This issue is different from (though related to) the missing information or lack of precision as discussed in Section 4.2. In the latter case uncertainty was theoretically there at the end of visual encoding, while former was caused by the human vision system.

**Traceability:** When there are many lines between adjacent axes, it becomes difficult to see individual ones in the resulting clutter. This is particularly problematic when most of the lines are almost parallel, but the ones that differ (which are often of particular interest) are hidden among or behind them (Figure 6a). Even if lines differ in the pixels at their end points, small angles between lines can cause confusion when looking at the space between axes. When lines converge onto the same pixel (or pixels that are very close together), it can become impossible to tell which line continues in which direction after that point (Figure 6b). While this can be a precision issue when the values are actually different, it becomes a pure traceability problem when the underlying data values are identical, and thus would never be mapped onto different pixels, no matter the resolution of the display. This is a common problem when categorical data is present in datasets visualized using parallel coordinates. A similar issue exists also for clusters, whose structure can be confusing due to splits and overlaps on and near axes (Figure 6c).

The common solution to the problem is interaction, which allows the user to highlight a particular record or cluster, but this is not always practical and certainly does not provide as much information as directly showing it. Users are also not always aware of traceability issues and simply fail to see subtle patterns or outliers.

**Identity:** Identity uncertainty is usually caused by many lines or clusters crossing, sometimes at low angles, making it difficult to uniquely identify a line or cluster [HHE08]. A single line can easily be hidden behind many other lines that form a solid, or almost solid, structure (Figure 7a). This case is distinct from the traceability case because it may not be apparent on the axes that the line is even there; a line that is not known to be there cannot be traced by the user. Only hints between the axes can show that this data value even exists. Overlapping clusters create similar issues, with the additional problem that they can make the user assume the

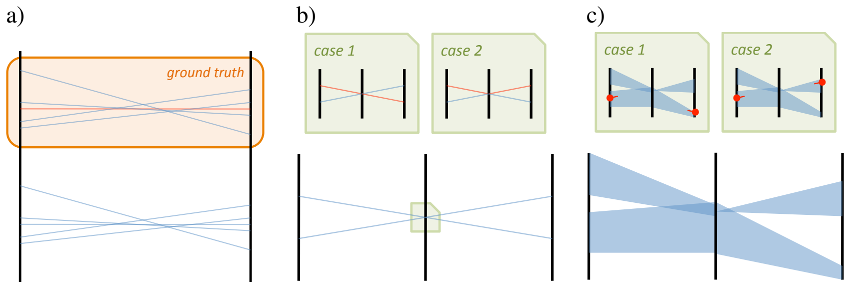


**Figure 6:** ***Traceability:*** *a) Single lines are easily hidden among others, leading to uncertainty about the exact number of lines and fine details in the data; b) lines meeting in single points, or in very small neighborhoods, on axes cause ambiguity about the multi-dimensional nature of the data; c) clusters show similar issues and are difficult to trace across multiple dimensions.*

existence of clusters that are not actually there. The lines created by overlaps can be misinterpreted as distinct clusters, and even when not it is often impossible to tell how many clusters there are (Figure 7b). A related issue making it difficult to tell how many (and which) clusters exist is when colors of cluster mix. Does the mixed color present a distinct cluster of that color or the overlap between two clusters of other colors (Figure 7c)?

### 5.2. Cognition

Cognitive uncertainty is caused by difficulties in cognitive reasoning, such as confusion and misinterpretation. Milliken [Mil87] classifies cognitive uncertainty into *state uncertainty*, *effect uncertainty* and *response uncertainty*. In data analysis, for example, state uncertainty may describe the lack of certainty about the data and information given. In the visualization context we term this category *lack of knowledge*. Effect uncertainty may describe the lack of certainty about what the information implies; response uncertainty may describe the lack of certainty about what action one should take (the latter two are outside the scope of this taxonomy).

**Lack of Knowledge:** Parallel coordinates require knowledge and experience to use for effective data analysis. Users who are unfamiliar with the way the technique depicts certain patterns may be unable to tell which pattern they are actually looking at (Figure 8a). Even when they are familiar with the technique, inconsistent axis scaling can mislead users. Parallel coordinates often scale every axis independently to make the most use of space, thus making direct comparison between them impossible, and shifting the locations of the zero on each axis. Patterns can be misinterpreted because of this (Figure 8b).

**Pattern Complexity:** Highly complex patterns in the visualization can lead to misinterpretations, even when they are correctly represented and readable on the perceptual level. While simple correlations, aggregation of values, etc., are easy to see, the superposition of different patterns can lead the user to see one pattern but ignore the other (Figure 9).

## 6. Discussion

In this section we analyze the existing research on parallel coordinates with respect to the taxonomy. Work on parallel coordinates has focused on two categories of work: qualitative tasks (clutter reduction, improving visual quality) and common analytical tasks (clustering, finding correlations, detecting outliers, privacy preservation). These tasks are based on the low-level analytical activities of a user [AES04] that are supported by parallel coordinates [AA01]. For the different uncertainty sources we analyze how these uncertainty sources reduce/enhance certain effects that are useful in some analysis scenarios. The discussions relating analytical tasks to visual uncertainty are summarized in Table 1 and described in detail in the following section.

### 6.1. Clutter

There are different definitions of clutter in the parallel coordinates literature. Peng et al. [PWR04] define clutter as the relative number of outliers to the total number of data points and aim to have a configuration which optimized with respect to outliers. The authors use reordering technique to achieve that configuration. This technique addresses the uncertainty due to configuration of the visualization: while this type of uncertainty is reduced on one hand due to preserving outliers, particular selection and ordering of axis increases the same effect of uncertainty due to possible omission of salient patterns.

*Identity*: Some clutter reduction techniques aim to reduce the number of visual elements, at the visual mapping stage. In parallel coordinates, that means reducing or manipulating number of lines that connect the data points. Sampling Lens [EBD05] is one such example where the data to be mapped on to the screen is abstracted based on density of the data points. Artero et al. [AdOL04] reduced non-important information in parallel coordinates based on the computed frequency and density plots from the original datasets. The screen space quality method [JC08] reduces clutter, while preserves the significant features in the original datasets at the same time, by filtering out data items based on distance

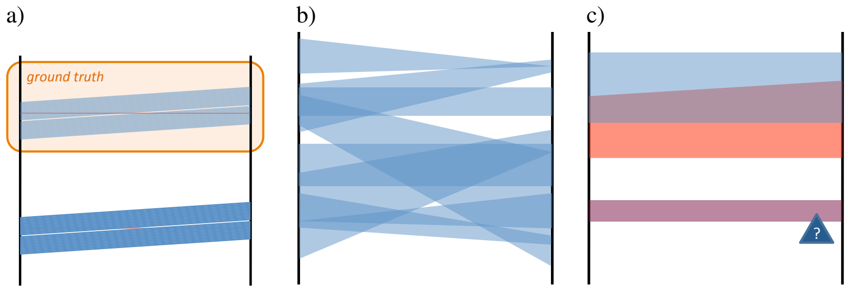


**Figure 7:** ***Identity:*** *a) Color mixing leading to confusion among identity of lines; b) overlapping clusters leading to clutter; c) color mixing among clusters lead to confusion among clusters*

transformation for data abstraction. These methods while reducing identity uncertainty, lead to lack of completeness in the visual representation.

*Traceability*: Another definition of clutter is according to Ellis and Dix [ED06], where clutter is attributed to large number of crossings and lines crossing at low angles In line-based parallel coordinates, clutter is caused by too many line crossings, several lines crossing at low angles and lots of lines converging or diverging from a small region on the axis. This relates to the uncertainty due to traceability between adjacent axes and across different axes, and also identity uncertainty on the axis. While in Pargnostics [DK10] the authors aim to retain data fidelity and reduce clutter through reordering-based optimization, in the previous case the authors reduce the number of visual elements, thereby leading to a completeness problem.

### 6.2. Clustering

Existing research on improving visual quality in parallel coordinates focuses on clustering [ZYQ*08, FWR99] in various forms.

*Configuration*: In both line-based and cluster-based parallel coordinates, binning helps in having pre-defined seeds for clustering. Binning can be either data-based or pixel-based. Pixel-based binning [NH06] helps in overcoming the problem due to high cardinality of a data-space, but leads to over-plotting. Cui et al. [CWRY06] have proposed metrics that measure the data quality. Pixel-binning therefore reduces configuration uncertainty. However, binning also leads to loss of precision and granularity uncertainty as many lines can end up on a single bin.

Low crossing angles help in the perception of proximity and similarity by inducing a Gestalt effect. For small number of data points, lines crossing at small angles generally mean lines are more or less parallel to each other, which indicates implicit clusters. In case of large number of data points, many lines crossing at low angles would tend to produce clutter. Clustering techniques in parallel coordinates aim to reduce the uncertainty related to data similarity and proximity and support analytical tasks of finding clusters within the data [AdOL04, AA04]. Information loss is intended in these cases. However uncertainty can be introduced due to lack of granularity information. The techniques do not generally convey the number of records within a cluster.

*Identity and Traceability*: Zhou et al. [ZYQ*08] proposed geometry-based visual clustering to implicitly enhance the clustering in parallel coordinates by bundling the edges, and minimizing the edge curvatures and maximizing the paral-

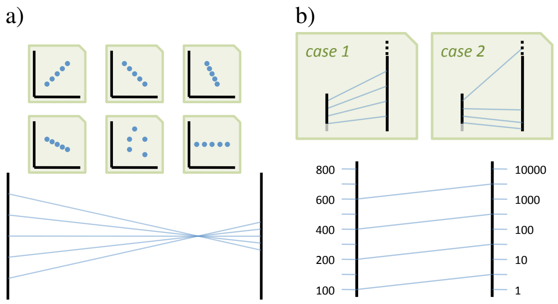


**Figure 8:** ***Lack of Knowledge:*** *a) Not knowing how to read the sometimes complex patterns in parallel coordinates leads to uncertainty about the represented pattern; b) inconsistent axis scaling, in particular because of the different locations of the zero, can lead to issues in interpretation.*

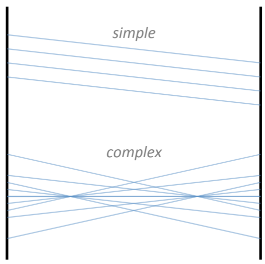


**Figure 9:** ***Pattern Complexity:*** *More complex patterns in the visualization lead to more difficulty in reading and understanding the underlying data patterns.*

| Task | Data Cardinality | Data Dimensionality | Source | Intended Effect | Unintended Effect | Utility |
|---|---|---|---|---|---|---|
| Finding Correlations | Large | | Large crossings | - Pattern Complexity | + Identity | Inverse correlation |
| | | Large | Axis selection | - Pattern Complexity | + Configuration | Correlation between dimensions |
| Detecting Outlier | Large | | Axis scaling | - Pattern Complexity | + Loss of precision | Spotting anomalies in trend |
| Clustering | Large | | Binning | - Configuration | +Precision | Seeds for clustering |
| | Small | | Low crossing angles | - Pattern Complexity | + Identity, Traceability | Clustering due to proximity, similarity |
| | | Large | Axis selection | - Pattern Complexity | + Configuration | Subspace clusters |
| Privacy-preservation | Any | | Binning | + Identity | N/A | Loss of precision and granularity |
| | | | Overlaps on the axis | + Identity | +Pattern Complexity | Uncertainty in identifying individual values |
| | | | Cluster splits | + Traceability | +Pattern Complexity | Uncertainty for traceability of sensitive clusters |

**Table 1:** *Connecting sources and effects of uncertainty to tasks and data properties (cardinality and dimensions). The positive sign indicates a particular effect of uncertainty is enhanced and negative sign implies the same is reduced. Usually, the intended effect is the reduction of a certain cause of uncertainty. In case of privacy, since increase of uncertainty is intentional, we consider the effect as being useful.*

lelism of adjacent edges at the same time. Other than reducing clutter, they also achieve reduction of uncertainty through enhancing the perception of continuity by choosing curved edges instead of lines as the basic visual elements. This reduces traceability uncertainty by violating the gestalt law of continuity among visual structures. Further, clusters can be detected by superimposing semitransparent line segments on the screen to enhance important components [ZCQ*09] and thereby reducing identity uncertainty.

Wegman and Luo [WL97] also use transparency to identify regions of high over-plotting through their dense color. Holten et al. [HVW10] have shown through user studies that improvements in visual enhancements do not always work well in practice. They have further argued for more formal evaluation measures for these techniques and we believe our definitions of visual uncertainty will help future approaches towards achieving a more quantitative basis for comparison.

### 6.3. Finding Correlations and Detecting Outliers

Line crossings, although lead to clutter in most cases, can be helpful in the case of a small number of data points, when large number of crossings at high angles is a useful representation for inverse correlations [DK10]. This enables the cognition of linear correlation, and thus reduces pattern complexity. For detecting outliers, normalization of the axes using non-linear scaling can be applied [AA01]. While this helps in reducing pattern complexity, there is significant loss in precision for the represented data.

### 6.4. Useful Uncertainty: Privacy

In case of privacy-preserving applications, contrary to the other categories mentioned above, certain effects of uncertainty are intentionally increased. For ensuring privacy data needs to be hidden, and in an interactive environment, there needs to be sufficient uncertainty to confuse the user so that he is not able to breach the intended privacy of the application. The uncertainty should, however be focused on the left part of the taxonomy tree, i.e, encoding uncertainty as increasing decoding uncertainty would degrade utility of the visualization to a much larger extent. Recently, we proposed a technique for privacy-preserving visualization [DK11], which exploits and manages the existing information loss in parallel coordinates to hide sensitive information.

*Precision and Granularity*: Loss of precision and granularity or lack of completeness are all related to information loss in visualization. While there has been sporadic mention of quantifying information loss [PAJKW08, ZK10], we still lack a framework for describing it. In privacy-preserving visualization, a clustering technique based on screen-space metrics is used to set a lower bound on the number of records per cluster. By using pixel-based binning as a staring point of the clustering process, it exploits the inherent loss of precision to mask the real values of the records. Uncertainty is also increased by the unknown location of the records within the clusters as shown in Figure 5 leading to granularity uncertainty. Thus, encoding uncertainty here is caused by both loss of precision and granularity.

*Identity and Traceability*: Cluster overlaps on the axis make it difficult for an attacker to point to individual data values due to identity and traceability uncertainty. Cluster splits as shown in Figure 10 add traceability uncertainty across the axes. These are all useful uncertainty from a privacy-preserving perspective. To optimize the utility, we should try to minimize cognitive uncertainty due to pattern complexity and also reduce identity uncertainty for axes which are less sensitive (for example, quasi-identifiers) than others (for example, sensitive attributes).

## 7. Uses of Visual Uncertainty Taxonomy

In the last section we categorized the existing literature on parallel coordinates based on our taxonomy of visual uncertainty. In this section we summarize our findings and provide some guidelines for the use of our taxonomy.

**Relating Tasks to Effects of Uncertainty.** As shown in Table 1 for the different tasks there are intended and unintended effects of uncertainty. Certain visual artifacts that are believed to increase visual uncertainty, like line crossings causing problems due to identity and traceability, can also help in identifying inverse correlations [DK10]. These cases needs to be identified when designing a visualization.

**Design Choices.** The design choices for encoding informa-

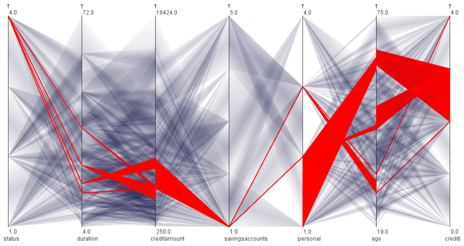

**Figure 10:** *Privacy-preserving parallel coordinates are an example for the usefulness of controlled uncertainty [DK11].*

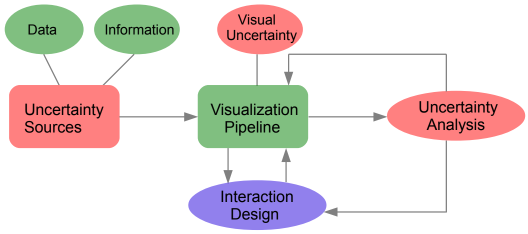


**Figure 11:** *While the traditional visualization pipeline has a linear structure, analysis of visual uncertainty provides a feedback loop for iterative refinement of the visual representation and the associated interaction techniques.*

tion have to be informed by the trade-offs between the different effects of uncertainty. For example, in parallel coordinates, clusters can be either represented by closely bundled polylines or by solid polygons. While the granularity information is available in the first case, it is absent in the second. For the specific task that a user wants to perform, we have to apply the appropriate visual representation. In the privacy case, since granularity information is something we want to hide, polygonal clusters would be the design choice.

**Framework for Systematically Defining Metrics**: Bertini et al. [BTK11] have pointed out the subjectivity in the choice of quality metrics in general. With the taxonomy shown in Figure 1, one can systematically design metrics to measure different types of uncertainty, for example, along the line of leaf nodes in Figure 1. We can qualitatively evaluate a metric based on what it measures and what it does not measure, and also identify aspects of uncertainty where no metric has been proposed. We can also combine metrics by following the tree from leaf nodes to the root. The direction of research involving use of metrics needs to be pursued further, so that analysts' trust in visualizations [BTK11] can be established more effectively.

**View Optimization.** In previous work [DK10] we showed the benefits of optimizing parallel coordinates by using screen space metrics. Similarly analysis of the causes and effects of visual uncertainty can inform the preceding stages of visualization and help refine the output. While the data-space uncertainty factors can be studied using the framework proposed by Correa et. al. [CCM09], our model will serve as a feedback loop (Figure 11), that is absent in the current linear structure of the visualization pipeline model.

**Interaction Design.** Certain types of visual uncertainty, like the one involving configuration can be reduced by interaction techniques. For example, reordering reduces configuration uncertainty, brushing reduces identity uncertainty; certain axis selections during the clustering task can enable one to visualize subspace clusters. We believe our taxonomy can be used to as a bridge between human factors and visual information in the context of interaction design (Figure 11).

## 8. Conclusions and Future Work

In this paper we have introduced the concept of *visual uncertainty* and proposed a taxonomy according to existing concepts in the uncertainty literature. Our work can be extended for further theoretical research on building a taxonomy for uncertainty in different types of visualization techniques, besides parallel coordinates. We believe such a taxonomy for visual uncertainty can serve as a foundation for future techniques to be developed that takes into account the issues related to uncertainty and develops means to address them in terms of visual design. We have also illustrated the application of uncertainty in privacy-preserving data analysis scenarios, where we intentionally hide information from the user and uncertainty thus becomes a desirable artifact. Based on our existing taxonomy we want to build a complete model of screen-space uncertainty in visualization, and develop metrics that quantify it. This will be useful to either reduce uncertainty if desired or precisely control it for purposes such as privacy-preserving information visualization.

## References

[AA01] ANDRIENKO G., ANDRIENKO N.: Constructing parallel coordinates plot for problem solving. In *Proceedings Smart Graphics* (2001), pp. 9–14. 6, 8

[AA04] ANDRIENKO G., ANDRIENKO N.: Parallel coordinates for exploring properties of subsets. *Proceedings Coordinated and Multiple Views in Exploratory Visualization* (2004), 93–104. 7

[AdOL04] ARTERO A. O., DE OLIVEIRA M. C. F., LEVKOWITZ H.: Uncovering clusters in crowded parallel coordinates visualizations. *Proceedings Information Visualization* (2004), 81–88. 6, 7

[AES04] AMAR R., EAGAN J., STASKO J.: Low-level components of analytic activity in information visualization. *Proceedings Information Visualization* (2004), 111–117. 6

[BS06] BERTINI E., SANTUCCI G.: Visual quality metrics. In *In Proceedings,BELIV Workshop* (2006), pp. 1–5. 2

[BTK11] BERTINI E., TATU A., KEIM D.: Quality metrics in high-dimensional data visualization: an overview and systematization. *IEEE Tansactions on Visualization and Computer Graphics 17*, 12 (2011), 2203–2212. 2, 9

[CCM09] CORREA C., CHAN Y., MA K.: A framework for uncertainty-aware visual analytics. In *Visual Analytics Science and Technology, 2009.* (2009), IEEE, pp. 51–58. 9

[Chi00] CHI E.: A taxonomy of visualization techniques using the data state reference model. In *Proceedings Information Visualization* (2000), pp. 69–75. 3

[CJ10] CHEN M., JÄNICKE H.: An information-theoretic framework for visualization. *Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1206–1215. 4

[CWRY06] CUI Q., WARD M., RUNDENSTEINER E., YANG J.: Measuring data abstraction quality in multiresolution visualizations. *Transactions on Visualization and Computer Graphics 12*, 5 (2006), 709–716. 7

[DK10] DASGUPTA A., KOSARA R.: Pargnostics: Screen-space metrics for parallel coordinates. *Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1017–26. 7, 8, 9

[DK11] DASGUPTA A., KOSARA R.: Adaptive privacy-preservation using parallel coordinates. *Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2241–2248. 2, 4, 8, 9

[EBD05] ELLIS G., BERTINI E., DIX A.: The sampling lens: making sense of saturated visualisations. In *CHI'05 Extended Abstracts on Human factors in Computing Systems* (2005), pp. 1351–1354. 6

[ED06] ELLIS G., DIX A.: Enabling automatic clutter reduction in parallel coordinate plots. *Transactions on Visualization and Computer Graphics 12*, 5 (2006), 717–724. 7

[FWR99] FUA Y.-H., WARD M. O., RUNDENSTEINER E. A.: Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings, Visualization* (1999), pp. 43–50. 7

[HHE08] HUANG W., HONG S.-H., EADES P.: Effects of Crossing Angles. In *Proceedings Pacific Visualization Symposium* (2008), pp. 41–46. 5

[HLS*12] HOLZHÜTER C., LEX A., SCHMALSTIEG D., SCHULZ H.-J., SCHUMANN H., STREIT M.: Visualizing uncertainty in biological expression data. In *Proceedings Visualization and Data Analysis* (2012). 3

[Hub10] HUBBARD D.: *How to measure anything: Finding the value of intangibles in business*. Wiley, 2010. 3

[HVW10] HOLTEN D., VAN WIJK J.: Evaluation of cluster identification performance for different pcp variants. *Computer Graphics Forum 29*, 3 (2010), 793–802. 8

[ID90] INSELBERG A., DIMSDALE B.: Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proceedings Visualization* (1990), pp. 361–378. 1

[JC08] JOHANSSON J., COOPER M.: A screen space quality method for data abstraction. *Computer Graphics Forum 27*, 3 (2008), 1039–1046. 6

[JS03] JOHNSON C., SANDERSON A.: A next step: Visualizing errors and uncertainty. *Computer Graphics and Applications 23*, 5 (2003), 6–10. 1

[KW99] KLIR G., WIERMAN M.: *Uncertainty-based information: Elements of generalized information theory*. Springer Verlag, 1999. 2

[LB98] LOPES A., BRODLIE K.: Accuracy in 3d particle tracing. *Mathematical Visualization: Algorithms, Applications and Numerics* (1998), 329–341. 1

[LPSW96] LODHA S., PANG A., SHEEHAN R., WITTENBRINK C.: Uflow: Visualizing uncertainty in fluid flow. In *Proceedings Visualization* (1996), pp. 249–254. 1

[Mil87] MILLIKEN F.: Three types of perceived uncertainty about the environment: State, effect, and response uncertainty. *Academy of Management review* (1987), 133–143. 2, 6

[NH06] NOVOTNY M., HAUSER H.: Outlier-preserving focus+context visualization in parallel coordinates. *Transactions on Visualization and Computer Graphics 12*, 5 (2006), 893–900. 4, 7

[PAJKW08] PURCHASE H., ANDRIENKO N., JANKUN-KELLY T., WARD M.: Theoretical foundations of information visualization. In *Information Visualization: Human-Centered Issues and Perspectives*. Springer, 2008, pp. 46–64. 3, 8

[PWL97] PANG A., WITTENBRINK C., LODHA S.: Approaches to uncertainty visualization. *The Visual Computer 13*, 8 (1997), 370–390. 1, 2

[PWR04] PENG W., WARD M., RUNDENSTEINER E.: Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Proceedings Information Visualization* (2004), pp. 89–96. 6

[RLBS03] RHODES P., LARAMEE R., BERGERON R., SPARR T.: Uncertainty visualization methods in isosurface rendering. In *Eurographics* (2003), pp. 83–88. 1

[RNC*95] RUSSELL S., NORVIG P., CANNY J., MALIK J., EDWARDS D.: *Artificial intelligence: a modern approach*. Prentice hall, 1995. 2, 3

[Sha48] SHANNON C. E.: A mathematical theory of communication. *The Bell System Technical Journal 27* (1948), 379–423. 3

[SLSR09] SKEELS M., LEE B., SMITH G., ROBERTSON G.: Revealing uncertainty for information visualization. *Information Visualization 9*, 1 (2009), 70–81. 2

[TA*09] TATU A., ALBUQUERQUE G., ET AL.: Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Proceedings Visual Analytics Science and Technology* (2009), pp. 59–66. 2

[THM*05] THOMSON J., HETZLER E., MACEACHREN A., GAHEGAN M., PAVEL M.: A typology for visualizing uncertainty. In *Proceedings SPIE* (2005), vol. 5669, pp. 146–157. 2

[TM04] TORY M., MÖLLER T.: Rethinking visualization: A high-level taxonomy. *Symposium on Information Visualization* (2004), 151–158. 2

[War04] WARE C.: *Information visualization: perception for design*. Morgan Kaufmann, 2004. 2

[Weg90] WEGMAN E.: Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association 85* (1990), 664–675. 4

[WL97] WEGMAN E., LUO Q.: High dimensional clustering using parallel coordinates and the grand tour. *Computing Science and Statistics 28* (1997), 361–368. 8

[WPL96] WITTENBRINK C., PANG A., LODHA S.: Glyphs for visualizing uncertainty in vector fields. *Transactions on Visualization and Computer Graphics 2*, 3 (1996), 266–279. 1

[ZCQ*09] ZHOU H., CUI W., QU H., WU Y., YUAN X., ZHUO W.: Splatting the Lines in Parallel Coordinates. *Computer Graphics Forum 28*, 3 (2009), 759–766. 8

[ZK10] ZIEMKIEWICZ C., KOSARA R.: Embedding Information Visualization Within Visual Representation. *Advances in Information and Intelligent Systems* (2010), 307–326. 8

[ZYQ*08] ZHOU H., YUAN X., QU H., CUI W., CHEN B.: Visual clustering in parallel coordinates. *Computer Graphics Forum 27*, 3 (2008), 1047–1054. 7