Evaluating Visual Analytics with Eye Tracking

Kuno Kurzhals[1], Brian Fisher[2], Michael Burch[1], and Daniel Weiskopf[1]

[1]University of Stuttgart, Germany
[2]Simon Fraser University, Canada

ABSTRACT

The application of eye tracking for the evaluation of humans' viewing behavior is a common approach in psychological research. So far, the use of this technique for the evaluation of visual analytics and visualization is less prominent. We investigate recent scientific publications from the main visualization and visual analytics conferences and journals that include an evaluation by eye tracking. Furthermore, we provide an overview of evaluation goals that can be achieved by eye tracking and state-of-the-art analysis techniques for eye tracking data. Ideally, visual analytics leads to a mixed-initiative cognitive system where the mechanism of distribution is the interaction of the user with visualization environments. Therefore, we also include a discussion of cognitive approaches and models to include the user in the evaluation process. Based on our review of the current use of eye tracking evaluation in our field and the cognitive theory, we propose directions of future research on evaluation methodology, leading to the grand challenge of developing an evaluation approach to the mixed-initiative cognitive system of visual analytics.

Categories and Subject Descriptors

Human-centered computing [**Visualization**]: Empirical studies in visualization

General Terms

Evaluation

Keywords

Eye tracking, visual analytics, visualization, evaluation methods, visual cognition

BELIV'14, November 10, 2014, Paris, France

ACM 978-1-4503-3209-5/14/11…$15.00
http://dx.doi.org/10.1145/2669557.2669560

1. INTRODUCTION

Eye tracking has been widely used to measure the distribution of visual attenuation, often in connection with analyzing how well participants perform with certain tasks on visual stimuli. The task might be dependent on the environment in which eye tracking is applied. Traditionally, eye tracking has been applied in areas like psychology and marketing research [14].

The canonical early eye tracking work was that of Alfred Yarbus [55], who demonstrated that the path taken by the gaze of his observer across paintings of various naturalistic scenes was determined by the interaction of scene information and the nature of the observer's task. For example, after an initial inspection of the painting "They Did Not Expect Him", Yarbus asked the viewer to "Estimate the material circumstances of the family in the picture", "Surmise what the family had been doing before the arrival of the 'unexpected visitor"', and "Estimate how long the 'unexpected visitor' had been away from the family". This led to different patterns of eye movement (scanpaths) across the scene as the observer sought the required information. In naturalistic scenes such as these, it is thought that the "gist" of the entire scene is perceived quickly, and interacts with the observer's task and top-down knowledge to determine what areas are to be fixated and in what order [4, 33]. As eye tracking technology became more available, a large number of studies were conducted using a variety of technologies, scenes, tasks, etc. [4, 13, 49].

However, only recently eye tracking has become increasingly popular in visualization research—as a means of evaluating visualization techniques, but also as a source of data for which visualization can be used for analysis [6]. For evaluation purposes, one typically records the eye movements of study participants when they perform a given task with a visual stimulus depicting some kind of data visualization. It is thought that the measurement of spatio-temporal eye movement data may well be more diagnostic than popular summative performance variables, such as completion time and accuracy, recorded in traditional user studies. In addition, because eye movements are recorded in an ongoing basis throughout the visualization task, they can provide insight into the process of working with a visualization environment. On the challenging side, however, this spatio-temporal aspect of the eye movement data requires more sophisticated data analysis and visualization methods tailored to the tasks and stimuli of interest.

A simple mapping of the psychology research methods mentioned above onto the interpretation of eye position in visualization tasks is complicated by a number of perceptual and cognitive factors. Both natural scenes and the reduced-cue experiments used in laboratory studies by psychologists tend to be composed of discrete objects about which a decision can be made, e.g., the material circumstances question in the Yarbus example might be answered by inspection of clothing and room decor. In contrast, understanding a visualization often requires a judgment to be made based on the configuration of multiple objects or aspects of a given object (e.g., seeking clusters of points in a scatterplot), chromatic patterns as in a color map used for scalar-data visualization, the orientation of elongated objects such as in a line chart, or the relative area of sections of a pie chart. While it is clear that these kinds of tasks are not processed in the same way as object categorization [5], it is not obvious where one would expect an observer to look in order to perform tasks based on important aspects of a visualization, or how a given scanpath might be interpreted as a predictor of some particular cognitive operation. Some studies in the psychology literature have used scenes that were entirely artificial (e.g., counting and subitizing in point cloud displays [56]), and these may provide a basis for investigation. However, the problems are further complicated when interactive visualization is evaluated because dynamic stimuli require even more advanced models and evaluation methods. Even more so, the evaluation of visual analytics is challenging because it, in the ideal case, forms a mixed-initiative cognitive system—with the user interacting with the visualization environment.

These problems might be one of the reasons that the number of eye tracking studies is much smaller than the number of other user studies in visualization and visual analytics (see Section 5). In the past, high prices of eye tracking hardware and technology might have been another roadblock. Since hardware components become cheaper and cheaper and easier to handle, the potential of this technology for an extension of the evaluation in our research community is considerable. Another issue might be that it is not clear to what class of evaluation problems in visualization and visual analytics eye tracking is applied best, and which analysis methods can be employed to derive knowledge from the recorded gaze data.

In this position paper, we provide an overview of how eye tracking is currently used in the evaluation of visualization techniques and how the gaze data is analyzed. As another contribution, we describe existing cognitive models and how they can be related to eye tracking for visual analytics. Based on these ingredients, we propose a number of promising areas in which eye tracking could advance evaluation methods, sketch ways how to approach these evaluation problems, and identify open research challenges. We hope that we can stimulate other researchers to work with eye tracking in visualization and visual analytics.

2. RELATED WORK

The evaluation of visualization techniques is challenging, but it has been acknowledged in our research community that we need good ways of assessing visualization [11, 35] and visual analytics [53]. In particular, the series of BELIV Workshops addresses the issue of how we can evaluate visualization, going beyond traditional measurements of task accuracy and completion time. For example, in the 2012 BELIV Workshop, Elmqvist and Yi [15] proposed a general approach to evaluating visualizations based on a collection of patterns. For a most recent overview of user study-based evaluation in visualization, we refer to Tory [50]. She provides her reflection on user studies and a categorization of testing methods, based on major goals such as understanding vs. evaluation as well as common methodological approaches such as quantitative experiment, qualitative observational study, inspection, and usability study. In another recent publication, Freitas et al. [17] discuss usability evaluation for information visualizations by particularly looking at it from a user-centered perspective.

Although eye tracking has a very long tradition and has been widely used in many fields, as discussed in Section 1, there is remarkably little work in the visualization literature that would address eye tracking as a means of evaluating visualization or visual analytics. For example, the paper by Goldberg and Helfman [19] is the only paper from any of the previous BELIV Workshops that would specifically address the issue of eye tracking evaluation methodology. Goldberg and Helfman present how eye tracking can be applied to evaluate simple information graphics, such as bar charts and line graphs. They focus on statistical analysis of common eye tracking metrics and visual analysis of scanpaths. Based on our research on similar user studies that included eye tracking for the evaluation of visualization techniques, we will provide a broader overview of analysis methods applied for different research questions.

Despite such little prior work on eye tracking-based testing methodology in our community, we have been witnessing a rapid increase in the number of user studies that, at least in parts, use eye tracking. One contribution of our paper is a summary and categorization of such papers; see Section 5. The majority of the papers were published in the last 4 years, showing a steep gradient of related papers. Despite the still small absolute number of such studies, we believe that this strong increase shows that eye tracking evaluation methodology is a timely topic for our research community.

Our paper is in line with previous work that reflects on how visualization is evaluated. In the context of evaluating information visualization in general, Lam et al. [31] describe seven scenarios, based on an extensive literature review of more than 800 visualization papers. They consider the "understanding of environments and work practices" and the evaluation of "visual data analysis and reasoning", "communication through visualization", "collaborative data analysis", "user performance", "user experience", and "visualization algorithms" in their survey. In a follow-up paper, Isenberg et al. [24] extend the literature review to include papers from scientific visualization. Isenberg et al. adopt the coding scheme by Lam et al., with only minor changes and extensions: they add a new category "qualitative result inspection" and change the evaluation of "visualization algorithms" to "algorithm performance". One result of both Lam et al.'s and Isenberg et al.'s reflection on the research field is that we have been witnessing a strong increase in the portion of user-related evaluation—across the different subfields of visualization. Another observation is the dominating role of the categories "user performance" and "user experience". Our paper builds a link in particular to "user performance" because eye tracking is most often used in controlled laboratory experiments that aim to measure and understand the performance of users with visualization.

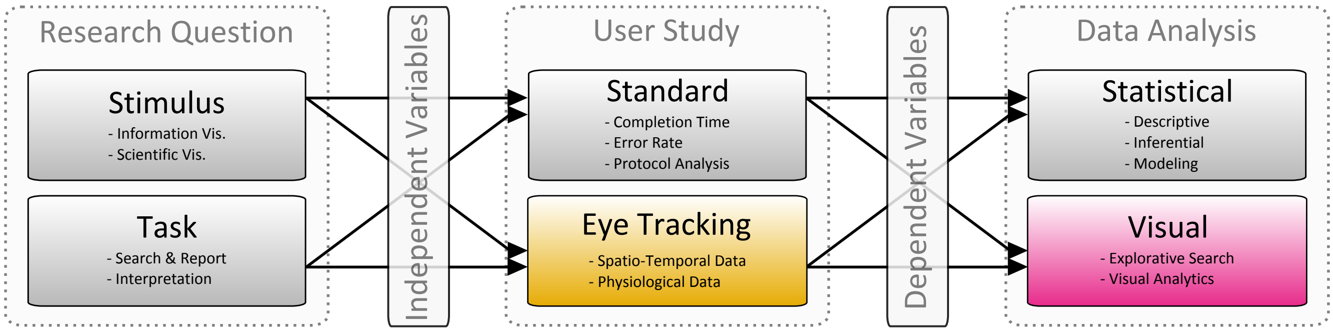

Figure 1: Pipeline of user-oriented evaluation. Stimulus and task represent the independent variables, measurements from a user study represent the dependent variables. The data analysis can then be performed with statistical methods, or in combination with a visual data analysis.

However, we also want to go beyond the traditional user performance with (isolated) visualization: in fact, we target the evaluation of visual analytics in the sense of the "science of analytical reasoning facilitated by interactive visual interfaces" [48]. As Ribarsky et al. [39] discuss, there is the general need for a science of analytical reasoning, including a human cognitive model, but they do not detail any evaluation methodology. We describe in Section 6 some cognitive models and how they relate to the evaluation of the combined cognitive system of user and visualization interface. In this sense, our paper addresses the evaluation scenario "visual data analysis and reasoning" identified by Lam et al. and Isenberg et al. However, as discussed in their papers, there is little previous work that would address the combined evaluation of such distributed cognition; much of the previous evaluation is based on case studies.

Except for the aforementioned paper by Goldberg and Helfman [19], none of the above papers deals with eye tracking evaluation in detail. With our paper, we want to fill this gap. We provide reflections on the current state of how eye tracking is used for evaluation in visualization and visual analytics; and we discuss the directions for future research that will allow for a broader use of eye tracking. We focus on eye tracking evaluation in the context of controlled laboratory experiments.

It should be noted that there is yet another aspect of eye tracking related to visualization and visual analytics: their use for the visual analysis of eye tracking data. This paper does not focus on this connection between eye tracking and visualization or visual analytics. Instead, we refer the reader to a recent state of the art report by Blascheck et al. [6] and a review of visual analytics techniques for eye tracking data by Andrienko et al. [2]. However, we do discuss some of the analysis problems for eye tracking data as far as they are concerned with analyzing the results of eye tracking studies with visualization and visual analytics; see Section 4.

3. EVALUATION PIPELINE

A typical user study for visualization techniques can be described by a pipeline as shown in Figure 1. Here, we assume a controlled laboratory experiment, even though many aspects carry over to other variants of user studies. A task is given to the study participant, that is to be solved by using visualization or visual analytics. The visual stimuli and choice of tasks serve as independent variables of the study. In this context, different visualization techniques and/or variations of one technique provide the basis for the visual stimuli. The task often requires the user to search and report certain aspects, or interpret the stimulus.

The performance with the task is assessed in the form of dependent variables. Standard measurements are the completion time and accuracy. However, protocol analysis is often employed as well, in particular, the "think aloud" protocol analysis [16]. Finally, the data acquired through the dependent variables is analyzed, eventually leading to conclusions regarding the study. Very well accepted is data analysis in the form of statistical inference for hypothesis testing. However, descriptive statistics and statistical modeling might be employed, too. To some degree, visually oriented data analysis might appear here as well. Nonetheless, the standard procedure for the overall user study process is oriented along the goal of hypothesis testing with statistical methods.

With eye tracking, the evaluation pipeline is extended; see the color highlighting in Figure 1. First, eye tracking provides additional dependent variables, in particular, spatio-temporal data that provides information about the participant's viewing behavior or physiological data by the pupil diameter, which can be an indicator of cognitive load [1, 27]. Due to the complexity of the spatio-temporal gaze data, we usually have to derive other, more simplified dependent variables from the raw gaze data in order to perform data analysis. Section 4 reviews typical examples of such aggregated metrics for eye tracking data. However, with data aggregation, we lose much of the information about the spatio-temporal nature of the eye tracking data. Therefore, as second major change, eye tracking studies often come with visual spatio-temporal analysis of the gaze data.

In fact, eye tracking experiments and the accompanying visual data analysis may often be used for hypothesis building, not just hypothesis testing, because they allow for a detailed "window" into how the participant works with the visualization over time. The data analysis methods suitable for eye tracking data are reviewed in the next section.

It should also be noted that the pipeline from Figure 1 targets the evaluation of visualization techniques. For visual analytics, the much more complex distributive cognitive system that includes the user and machine needs to be evaluated. To this end, we also have to include cognitive modeling of the user, as discussed in Section 6.

4. EYE TRACKING DATA ANALYSIS

With the spatio-temporal eye tracking data recorded in a user study, the data analysis can be performed by two different approaches: statistical and visual analysis.

4.1 Statistical Analysis

The raw gaze data is usually preprocessed by an appropriate filter algorithm to detect fixations and saccades; for further reading on eye tracking terminology, we refer to Holmqvist et al. [22] and Blascheck et al. [6]. The preprocessed data can then undergo statistical analysis. Typically, the data has to be further aggregated to allow for the application of statistical methods. An important class of analysis approaches is based on eye tracking metrics computed from the (preprocessed) eye tracking data. Objects or specific regions on a stimulus can be of special interest. By defining boundary shapes around these Areas of Interest (AOIs), fixation data can be mapped to the areas. The common eye tracking metrics can be separated in three categories, according to Poole and Ball [37]:

- **Fixation-derived metrics**: Fixations with or without AOI information can be processed. A common metric is defined by the number of fixations per AOI, which indicates the relevance of the AOI for the users. To compare the distribution of attention between AOIs, the sum of fixation durations may be used.
- **Saccade-derived metrics**: The characteristics of the saccades may indicate the quality of visual cues in the stimulus or the extent of visual searching. For example, large saccade amplitude can indicate meaningful cues that draw the attention from a distance, or a high frequency of saccades could come from much visual searching. Therefore, saccade-derived metrics can serve to indicate difficulties with the visual encoding.
- **Scanpath-derived metrics**: The scanpath consists of the full sequence of fixations and saccades. Therefore, scanpath-derived metrics can acquire information about visual reading strategies or pinpoint specific problems with the visualization design during the task. The transition matrix is the common approach to analyzing transition patterns between AOIs, albeit it does not represent the full sequence but only the collection of pairs of fixations from the sequence.

Once we have values from any of these metrics, we can directly apply statistical methods, including inferential or descriptive statistics as well as statistical modeling. Therefore, these metrics can serve as a basis for hypothesis testing.

A major problem is that the eye tracking metrics have to be interpreted with caution because they are no unambiguous indicator for certain characteristics of cognitive or perceptual processing. In fact, they provide a rather coarse and aggregated perspective on the participant's viewing behavior. Therefore, they are best accompanied by complementary indicators, or the eye tracking study is specifically designed to evoke and test clearly specified hypotheses. Another problem is that the metrics were typically developed for visual stimuli that are different from those from visualization; therefore, it still needs to be demonstrated that the metrics are indicators for the same characteristics.

Eye tracking data usually contains much more information than represented by the above, aggregated metrics. Therefore, statistical analysis can also be applied to data that is closer to the original gaze data. In particular, statistical modeling to predict and classify scanpaths on stimuli provides a promising approach for a more complete analysis for visualization stimuli. Here, one issue is to generate the appropriate model for the scanpath (e.g., define the appropriate AOIs) and employ the appropriate statistical methods. In this context, one can use data-mining techniques such as scanpath clustering [19], layered hidden Markov models [12], or measures for the similarity between aggregated scanpaths [21].

4.2 Visual Analysis

In general, visualization can complement statistical analysis by providing additional insight into the data by exploratory search, building hypotheses, or the presentation of confirmed analysis results [41]. The same is true of the special cases of eye tracking data analysis. In particular, visualization is a very good means of examining the spatial, temporal, or spatio-temporal aspects of the data [6].

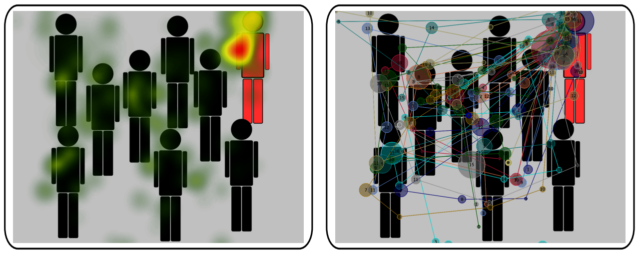

Figure 2: Attention map (left) and gaze plot (right).

The most common visualization techniques are attention maps and gaze plots (see Figure 2). Attention maps display the spatial distribution of eye tracking data on a stimulus. The data can be aggregated over time for one participant or multiple participants. Although attention maps can provide a good overview of important areas of interest on a static stimulus, the temporal component of the data is lost. In contrast, gaze plots provide a spatio-temporal perspective on fixation sequences and can be investigated to identify potential reading strategies. However, with increasing length of the scanpath, or with scanpaths from multiple participants, the visualization becomes cluttered and hard to interpret. Alternatively, transition matrices allow us to analyze transition patterns but lack the interpretation of longer transition sequences (beyond just pairs of fixations). In summary, the traditional visualization techniques are well prepared to provide a qualitative picture of the distribution of attention aggregated over time (attention maps) or of the short scanpath of a single participant—both for static stimuli. In these cases, they can be used also for eye tracking experiments with visualization or visual analytics, in particular, for exploratory data analysis and hypothesis building.

However, for more challenging research questions—including ones that work with dynamic stimuli, many participants or groups of participants, and coupled spatial and temporal structures of the gaze data—the above visualization techniques are not sufficient. Therefore, there is much, mostly recent, work in the visualization community to develop improved visualization techniques, for example, for displaying time-oriented AOI data [9], complete sequences of scanpaths [52], or spatio-temporal gaze data [30]. A particularly interesting approach is visual analytics for eye tracking data [2], combining statistical and data-mining techniques with interactive visualization; recent examples combine scanpath clustering with visualization [28] or multiple coordinated views with statistical graphics [40].

5. EYE TRACKING EVALUATION IN VISUALIZATION AND VISUAL ANALYTICS

Over the last years, we have been witnessing an increasing number of publications that included eye tracking in user studies to evaluate visualization techniques. This section summarizes and categorizes the previous examples of the eye tracking evaluation. For our systematic review, we checked the main journals (including special issues of conferences) and proceedings for visualization and visual analytics, each spanning their whole time span of publications. These include the current publication channels:

- IEEE Transactions on Visualization and Computer Graphics (TVCG)
- Computer Graphics Forum (CGF)
- Information Visualization Journal (IVS)
- IEEE Conference / Symposium on Visual Analytics and Technology (VAST)
- IEEE Pacific Visualization Symposium (PacificVis)
- International Conference on Information Visualisation (IV)

We also included older proceedings that are no longer published in this form (because they now appear in one of the above journals, or the conferences are succeeded by other conferences):

- IEEE Conference on Visualization (VIS)
- IEEE Symposium on Information Visualization (INFOVIS)
- Eurographics / IEEE TCVG Symposium on Visualization (VISSYM)
- Eurographics Conference on Visualization (EuroVis)
- Asia-Pacific Symposium on Visualization (APVIS)

From these sources, we identified 12 publications that include eye tracking in a user study for the evaluation. Table 1 summarizes the results. Several of the publication channels had no papers with eye tracking evaluation. And, obviously, there are many more user study papers with eye tracking, albeit outside our research community and, therefore, in other publication channels.

The different research questions the authors investigated with eye tracking can be summarized as follows:

1. **Distribution of visual attention [23, 25, 26, 29, 42, 46]:** Visualization techniques were compared by fixation metrics for the attention on AOIs to investigate how the techniques are perceived and to identify possible usability issues. Attention maps were applied to visualize the spatial distribution of attention on the stimuli [25, 26, 29, 42] and support the statistical results.

2. **Sequential characteristics of eye movement [7, 8, 18, 51]:** In addition to fixation-related metrics on AOIs, the transition frequencies between AOIs with transition matrices [8], transition graphs [51], and visual scanpath analysis [18] were analyzed to gain insight into how users investigate a visualization (e.g., as an explanation for a decrease in task performance). Also, gaze analysis by visual analytics was applied to identify reading strategies [7].

Table 1: Overview of the investigated visualization papers that include an eye tracking study.

Year	Reference	Evaluation
2005	Tory et al. [51]	AOI fixation percentage and transition frequencies between different views.
2007	Huang et al. [23]	Visual investigation of gaze replay for graph layouts.
2008	Kim & Varshney [26]	Fixation percentage and fixation duration to compare attention-guiding rendering techniques.
2009	Swindells et al. [46]	AOI fixation count for comparison of parameter manipulation methods.
2009	Siirtola et al. [42]	Analysis of parallel coordinates by attention map, AOI fixation duration, and AOI fixation counts.
2011	Burch et al. [8]	Attention maps and AOI transition matrices for different tree layouts.
2011	Goldberg & Helfman [18]	AOI fixation times and visual scanpath analysis for different graph layouts.
2012	Kim et al. [25]	AOI fixations and visit durations for the comparison of two visualizations for table sorting.
2013	Bekele et al. [3]	AOI fixation percentage and duration for comparison of two groups looking at VR faces.
2013	Kurzhals et al. [29]	AOI fixation duration for the comparison of attention guiding visualizations.
2013	Burch et al. [7]	Advanced visual analysis methods (time-varying distances, time series plots, interval-based trajectory plots) for the investigation of tree layouts.
2014	Song et al. [43]	A 3D attention map for CT and MRI images was used to compare gaze of radiologists with different levels of expertise.

3. **Comparison between user groups [3, 43]:** Complementary to the previous two points, the distribution of attention between different groups was investigated. Group comparisons were performed between healthy and mentally disordered persons, or between novice and expert groups. Comparisons were based on a statistical analysis of AOI fixation metrics [3] or visual comparison of gaze point distributions [43].

These eye tracking studies mainly relied on the statistical analysis of AOI-based fixation metrics. If performed, the visual data analysis was often limited to the investigation of attention maps and scanpath visualizations. For the identification of visual reading strategies, more advanced visual analytics techniques were applied. However, none of the above studies investigated the full sequence length of scanpaths or any complex spatio-temporal characteristics of eye tracking for dynamic stimuli, let alone any cognitive aspects related to the mixed-initiative distribution of cognition in visual analytics.

6. COGNITIVE MODELS

Many of the psychology studies discussed in Section 1 were designed to build an understanding of human cognitive architecture, i.e., the aspects of human information processing that are thought to generalize across a wide range of individuals, environments, and tasks. For example, the two visual systems theory of Trevarthen (see, e.g., [20, 34]) is thought to predict changes in response to a range of visual illusions when tasks are motor vs. cognitive in nature. In order to effectively utilize eye movement information as a window onto cognitive processes in dynamic visualization environments we must move beyond naturalistic studies and laboratory investigations abstracted from those environments to focus more closely on the artificial scenes (e.g., dashboards) that we generate and the analytical cognitive processes that our visualizations are meant to support. This does not mean that we limit our evaluation to visualization systems and tasks per se, but that stimuli and tasks used in our studies should demonstrate aspects of human cognitive architecture that are important for the design and evaluation of visualization systems and the ways in which they are used to understand situations and make decisions.

One approach is to advance fundamental theories of human cognition in areas that relate to the perceptual situations and cognitive tasks that are important for the evaluation of visual information systems. Directed fundamental research studies provide knowledge of human capabilities and limitations that can be used by designers of systems for a variety of applications. The goal here is in essence to build a basic "science of analytical reasoning" specific to the kinds of operations that might be "facilitated by interactive visual interfaces" [48]. This kind of study will necessarily build upon general theories, frameworks, and methods from psychology, and many findings that result will be of interest to those conferences and journals. The specific research questions addressed, however, will be those that are most informative for the evaluation of those interfaces.

For example, Liu et al. [32] used reduced-cue experimental methods typical for psychology studies. However, its focus on transformations of the visual environment that are important for graphical visualization environments, in this case for air traffic control, suggests that it may also be of use in the design and analysis of these environments. Liu et al. began by replicating studies by Pylyshyn et al. [38] that discovered a new fundamental aspect of human cognitive architecture, the FINST ("fingers of instantiation") attentional token mechanism. This new form of attention is described as a "spatial index" that tracks multiple moving objects in the visual scene in order to support performance in a variety of tasks that depend on rapid access to information about those objects. Liu et al. hypothesized that FINSTS were important for air traffic controllers' ability to associate information from memory (such as being low on fuel) with a specific aircraft representation displayed on their screen. If the FINST link to a particular display object were to fail, the controller must then put cognitive effort into recalling the information and reestablishing its relationship with the proper display object. One possible threat to the FINST mechanism might be camera movement in the simulated 3D "fishtank VR" scenes that were to be used in proposed "NextGen" Air Traffic Control systems. Liu et al. found that participants in their experiments were able to track multiple moving targets through a surprising range of display transformations, even when those scenes were displayed in 2D. The lack of effect of these transformations on participants' psychophysical tracking functions suggested that their use in NextGen ATC systems was not contraindicated. Because the study extended previous fundamental research and was conducted using similar laboratory stimuli and tasks, the results should generalize across a range of applications.

While eye moment methods can make substantial contributions to (directed) fundamental research, they are likely to be particularly useful for translational studies that build upon what is known about the cognitive architecture to examine how it is utilized in specific situations and tasks that are more similar to a visualization approach. These studies are intended not to contribute to the psychology literature nor to evaluate a specific visualization but rather constitute an intermediate "translational" study whose results would come in the form of more structured guidance—guidelines, visual queries [54], or design actions [44] that might be of use in the design and evaluation of a class of visualizations or visual information systems. An example of this type of study is that of Po et al. [36] that mapped the two visual systems theory of Trevarthen onto the kinds of displays used in CAD of large aircraft. The results of this study were interpreted with respect to the impact of individual differences on the design of interaction with these displays.

It is worthwhile to note that these methods are in addition to the summative evaluation of a specific application or visualization in the context of use, i.e., usability evaluation. Such laboratory study of the full application must of course be done for each mapping of cognitive architecture onto a real-world application design.

7. FUTURE DIRECTIONS

With the availability of cheap eye tracking hardware and its ease of use, there are no longer any technological obstacles for using eye tracking in user-based evaluation; in particular, in controlled laboratory studies, we can essentially record gaze data for free along with any traditional study procedure that aims to test task performance. Therefore, the big overall challenge is to make sense out of the eye tracking data and relate this data to something we want to learn about the visualization tested and the cognitive processes involved. As discussed before, there are already several examples of eye tracking studies in visualization: they mostly work with statistical analysis of quite aggregated data, for well defined hypotheses, and with traditional visual analysis by attention maps and gaze plots. In fact, many other laboratory studies could adopt these approaches to testing and data evaluation, adding a better understanding of reasons for task performance. Therefore, our general recommendation is that eye tracking should be considered as a testing method whenever you plan and design a laboratory study.

However, we see the real value of eye tracking going beyond what is possible now. Based on our reflections on the state of the art in the previous sections, we discuss relevant directions for future research on evaluation methodology. We begin with more technologically oriented research questions asking for short term action, and end with long term grand challenges.

7.1 Exploratory Data Analysis and Hypothesis Building

Well known statistical methods can be applied once we have a clearly defined hypothesis and an eye tracking experiment set up accordingly. The interesting question is how we can design such an eye tracking experiment, in particular, for the complex visual representations and tasks in applications of visualization and visual analytics. Here, we see a great potential for improved data analysis methods that could work on eye tracking data acquired in less constrained preliminary studies. In fact, **visual analytics** will certainly play a major role here [2], in particular, for the complex spatio-temporal nature of the eye tracking data *and* the (dynamic) stimulus data, and by combining data-mining, statistical, and interactive visualization methods.

One analysis aspect is most relevant, albeit difficult: improved **scanpath analysis**. So far, the studies focused on the spatial aspect of the recorded gaze data. Temporal aspects of the data, such as AOI sequences, provide important information about reading strategies but were often neglected completely or only partially covered through transition matrices. Therefore, better visual analysis for long sequence information needs to be developed.

Another relevant analysis aspect is concerned with **group comparisons.** The comparison of different user groups, especially the comparison of experts and novices, provides valuable information about different viewing behavior. To this point, group comparisons with eye tracking for visualization techniques are rare but could help improve the learning curve for visualization techniques by guiding novices with the knowledge from experts' gaze patterns. Therefore, there should be support for comparative visual analysis between groups.

A third aspect is the **combination of eye tracking data with additional time-oriented data**. For example, the temporal evolution of the dynamic stimuli needs to be understood to build the context for the gaze data. Or, the eye tracking data can be combined with information about logged interaction such as mouse or key-stroke data, as to obtain deeper insights in the usability of interactive visualization applications and visual analytics systems.

A practical aspect is concerned with making the newly developed analysis methods available to other researchers. Reflecting a general discussion in our community, we see the need for **disseminating codes, tools, and systems** so that improved analysis can be adopted quickly. One way is to have advanced analysis methods included in professional software by the vendors of eye tracking hardware; however, this approach might not always work due to the latency in this software development process and because not all of our analysis problems will be sufficiently relevant for the broader eye tracking audience. Therefore, there should also be dissemination of software (prototypes) developed, including complete analysis systems but also partial codes. For example, we have already shared our system "ISeeCube" [28] with other eye tracking researchers and plan a public release for the future.

7.2 Evaluation Procedures

We not only see the need for improved data analysis but also for extended evaluation procedures and protocols.

The "think aloud" protocol analysis [16] is a method commonly used by HCI researchers to elicit user reports of subtasks that take place in the course of accomplishing a given task with a specific user interface. This approach has much to recommend it when the goal is to produce a GOMS (Goals, Operators, Methods, and Selection rules [10]) model of the task as it takes place in a particular interface, to detect operational errors, and to document the process of repair of those errors. Since many perceptual processes are unconscious, user reports of how they detect known patterns and discover new patterns in data are not well elicited by this method. Taken in the context of cognitive models of task performance, the analysis of eye movement patterns may provide greater insight into unconscious aspects of task performance. The **combination of think aloud protocol analysis and eye tracking** may well lead to insight into both task performance using visual information systems and the perceptual processes that allow users to understand information presented by those systems.

One approach to this could be taken from the work of Tanenhaus et al. [47]. By interpreting eye movements as they occur in response to verbal task instructions, they were able to show a much tighter integration of linguistic understanding and overt attention to objects in the environment. Tanenhaus et al. showed that spoken instructions can guide eye movements in real time, with close coupling of hearing and eye position. The application of this method to visualization tasks might support a deeper theoretical understanding of visualization use as well as guidance for the design and evaluation of visualization environments. A translational approach to the use of eye movement research that might be productive for visualization researchers would be to adapt the methods used by Tanenhaus et al., leading to user studies with **verbal task instructions in combination with eye tracking**.

7.3 Translational Evaluation of Human Cognition

To move beyond the evaluation of usability and the techniques of visualization we must build an understanding of human cognition as it is shaped by visual information systems. This builds upon work done in psychology and cognitive sciences. Since these studies are not well-suited for application to visualization stimuli and tasks, we must move beyond off-the-shelf psychology and build translational studies in partnership with interested cognitive scientists. Eye movement records are a strong candidate for a **boundary**

object [45]: a method and data source that can be interpreted from both psychological and visualization perspectives, acting as a bridge between cognitive and computing science. Defining boundary objects such as eye tracking protocols and methodologies constitutes a grand challenge for visualization and visual analytics.

As an evaluation methodology, we see such boundary objects as linking functions for **multi-scale evaluation**. In this context, multi-scale is interpreted in an abstract sense: it reaches from low-level perception, over mid-level descriptions of tasks, all the way to human cognitive processes. The challenge is to define the details of such linking so that the existing models on the different levels can be connected quantitatively.

Bridging cognitive and computing science with eye tracking as a boundary object is only a temporary solution. As systems become more richly interactive and user develop perceptual and cognitive capabilities based on their experience with increasingly sophisticated visualization environments, we will find ourselves in the position of studying cognitive processes that are distributed between one or more human users and complex computational processes that provide non-trivial contribution to the overall cognitive system. At this point, we face the grandest challenge: to **understand and design mixed-initiative cognitive systems** where the mechanism of distribution is interaction with visualization environments.

8. ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG) under grants WE 2836/5-1 and WE 2836/1-2.

9. REFERENCES

[1] E. W. Anderson. Evaluating visualization using cognitive measures. In *Proc. Wksp. Beyond Time and Errors: Novel Eval. Meth. for Vis.*, pages 5:1–5:4, 2012.

[2] G. L. Andrienko, N. V. Andrienko, M. Burch, and D. Weiskopf. Visual analytics methodology for eye movement studies. *IEEE Trans. Vis. and Comput. Gr.*, 18(12):2889–2898, 2012.

[3] E. Bekele, Z. Zheng, A. Swanson, J. Crittendon, Z. Warren, and N. Sarkar. Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *IEEE Trans. Vis. and Comput. Gr.*, 19(4):711–720, 2013.

[4] I. Biederman. Perceiving real-world scenes. *Sci.*, 177(4043):77–80, 1972.

[5] I. Biederman and P. C. Gerhardstein. Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff. *J. Exp. Psychol.: Human Perception and Performance*, 21(6):1506–1514, 1995.

[6] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye tracking data. In *Proc. EuroVis State of the Art Reports*, 2014.

[7] M. Burch, G. Andrienko, N. Andrienko, M. Höferlin, M. Raschke, and D. Weiskopf. Visual task solution strategies in tree diagrams. In *Proc. IEEE Pac. Vis. Symp.*, pages 169–176, Feb 2013.

[8] M. Burch, N. Konevtsova, J. Heinrich, M. Höferlin, and D. Weiskopf. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *IEEE Trans. Vis. and Comput. Gr.*, 17(12):2440–2448, 2011.

[9] M. Burch, A. Kull, and D. Weiskopf. AOI Rivers for visualizing dynamic eye gaze frequencies. *Comput. Gr. Forum*, 32(3):281–290, 2013.

[10] S. K. Card, A. Newell, and T. P. Moran. *The Psychology of Human-Computer Interaction.* L. Erlbaum Associates, 1983.

[11] S. Carpendale. Evaluating information visualizations. In *Information Visualization: Human-Centered Issues and Perspectives*, pages 19–45. Springer, 2008.

[12] F. Courtemanche, E. Aïmeur, A. Dufresne, M. Najjar, and F. Mpondo. Activity recognition using eye-gaze movements and traditional interactions. *Interacting with Computers*, 23(3):202–213, 2011.

[13] M. DeAngelus and J. B. Pelz. Top-down control of eye movements: Yarbus revisited. *Vis. Cogn.*, 17(6):790–811, 2009.

[14] A. Duchowski. A breadth-first survey of eye-tracking applications. *Behav. Res. Meth. Instr., & Comput.*, 34(4):455–470, 2002.

[15] N. Elmqvist and J. S. Yi. Patterns for visualization evaluation. In *Proc. Wksp. Beyond Time and Errors: Novel Eval. Meth. for Vis.*, pages 12:1–12:8, 2012.

[16] K. A. Ericsson and H. A. Simon. *Protocol Analysis: Verbal Reports as Data.* MIT Press, 1993.

[17] C. M. D. S. Freitas, M. S. Pimenta, and D. L. Scapin. User-centered evaluation of information visualization techniques: Making the HCI-InfoVis connection explicit. In W. Huang, editor, *Handbook of Human Centric Visualization*, pages 315–336. Springer, 2014.

[18] J. Goldberg and J. Helfman. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Info. Vis.*, 10(3):182–195, 2011.

[19] J. H. Goldberg and J. I. Helfman. Comparing information graphics: A critical look at eye tracking. In *Proc. Wksp. Beyond Time and Errors: Novel Eval. Meth. for Vis.*, pages 71–78, 2010.

[20] M. A. Goodale. How (and why) the visual control of action differs from visual perception. *Proc. Roy. Soc. B: Bio. Sci.*, 281(1785):20140337, 2014.

[21] T. Grindinger, A. T. Duchowski, and M. W. Sawyer. Group-wise similarity and classification of aggregate scanpaths. In *Proc. Symp. ETRA*, pages 101–104, 2010.

[22] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* Oxford University Press, 2011.

[23] W. Huang. Using eye tracking to investigate graph layout effects. In *Proc. Asia-Pac. Symp. Vis.*, pages 97–100, 2007.

[24] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller. A systematic review on the practice of evaluating visualization. *IEEE Trans. Vis. and Comput. Gr.*, 19(12):2818–2827, 2013.

[25] S.-H. Kim, Z. Dong, H. Xian, B. Upatising, and J. S. Yi. Does an eye tracker tell the truth about

visualizations?: Findings while investigating visualizations for decision making. *IEEE Trans. Vis. and Comput. Gr.*, 18(12):2421–2430, 2012.

[26] Y. Kim and A. Varshney. Persuading visual attention through geometry. *IEEE Trans. Vis. and Comput. Gr.*, 14(4):772–782, 2008.

[27] J. Klingner, R. Kumar, and P. Hanrahan. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proc. Symp. ETRA*, pages 69–72, 2008.

[28] K. Kurzhals, F. Heimerl, and D. Weiskopf. ISeeCube: Visual analysis of gaze data for video. In *Proc. Symp. ETRA*, pages 43–50, 2014.

[29] K. Kurzhals, M. Höferlin, and D. Weiskopf. Evaluation of attention-guiding video visualization. *Comput. Gr. Forum*, 32(3):51–60, 2013.

[30] K. Kurzhals and D. Weiskopf. Space-time visual analytics of eye-tracking data for dynamic stimuli. *IEEE Trans. Vis. and Comput. Gr.*, 19(12):2129–2138, 2013.

[31] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Trans. Vis. and Comput. Gr.*, 18(9):1520–1536, 2012.

[32] G. Liu, E. L. Austen, K. S. Booth, B. D. Fisher, R. Argue, M. Rempel, and J. T. Enns. Multiple-object tracking is based on scene, not retinal, coordinates. *J. Exp. Psychol.: Human Perception and Performance*, 31(2):235–247, 2005.

[33] G. R. Loftus and N. H. Mackworth. Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol.: Human Perception and Performance*, 4(4):565–572, 1978.

[34] A. D. Milner and M. A. Goodale. Two visual systems re-viewed. *Neuropsychologia*, 46(3):774–785, 2008.

[35] C. Plaisant. The challenge of information visualization evaluation. In *Proc. Adv. Vis. Interf.*, pages 109–116, 2004.

[36] B. A. Po, B. D. Fisher, and K. S. Booth. Pointing and visual feedback for spatial interaction in large-screen display environments. In *Proc. Third Int'l Symp. Smart Gr.*, volume 2733 of *Lect. Notes Comp. Sci.*, pages 22–38. Springer, 2003.

[37] A. Poole and L. Ball. Eye tracking in human-computer interaction and usability research: Current status and future prospects. In C. Ghaoui, editor, *Encyclopedia of Human-Computer Interaction*, pages 211–219. Idea Group Inc., 2006.

[38] Z. W. Pylyshyn. Situating vision in the world. *Trends Cogn. Sci.*, 4(5):197–207, 2000.

[39] W. Ribarsky, B. D. Fisher, and W. M. Pottenger. Science of analytical reasoning. *Info. Vis.*, 8(4):254–262, 2009.

[40] G. Ristovski, M. Hunter, B. Olk, and L. Linsen. EyeC: coordinated views for interactive visual exploration of eye-tracking data. In *Proc. Conf. Info. Vis. (IV)*, pages 239–248, 2013.

[41] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Trans. Vis. and Comput. Gr.*, 19(12):2366–2375, 2013.

[42] H. Siirtola, T. Laivo, T. Heimonen, and K.-J. Raiha. Visual perception of parallel coordinate visualizations. In *Proc. Conf. Info. Vis. (IV)*, pages 3–9, 2009.

[43] H. Song, J. Yun, B. Kim, and J. Seo. GazeVis: interactive 3D gaze visualization for contiguous cross-sectional medical images. *IEEE Trans. Vis. and Comput. Gr.*, 20(5):726–739, 2014.

[44] B. Spence. The broker. In A. Ebert, A. Dix, N. D. Gershon, and M. Pohl, editors, *Human Aspects of Visualization*, volume 6431 of *Lect. Notes Comp. Sci.*, pages 10–22. Springer, 2011.

[45] S. L. Star. The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. In M. Huhns and L. Gasser, editors, *Readings in Distributed AI*. Morgan Kaufmann, 1988.

[46] C. Swindells, M. Tory, and R. Dreezer. Comparing parameter manipulation with mouse, pen, and slider user interfaces. *Comput. Gr. Forum*, 28(3):919–926, 2009.

[47] M. K. Tanenhaus, M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy. Integration of visual and linguistic information in spoken language comprehension. *Sci.*, 268(5217):1632–1634, 1995.

[48] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.

[49] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychol. Rev.*, 113(4):766–786, 2006.

[50] M. Tory. User studies in visualization: A reflection on methods. In W. Huang, editor, *Handbook Human Centric Visualization*, pages 411–426. Springer, 2014.

[51] M. Tory, M. S. Atkins, A. E. Kirkpatrick, M. Nicolaou, and G.-Z. Yang. Eyegaze analysis of displays with combined 2D and 3D views. In *Proc. IEEE Vis. Conf.*, pages 519–526, 2005.

[52] H. Y. Tsang, M. Tory, and C. Swindells. eSeeTrack – visualizing sequential fixation patterns. *IEEE Trans. Vis. and Comput. Gr.*, 16(6):953–962, 2010.

[53] J. J. van Wijk. Evaluation: A challenge for visual analytics. *IEEE Comput.*, 46(7):56–60, 2013.

[54] C. Ware. Visual queries: The foundation of visual thinking. In S.-O. Tergan and T. Keller, editors, *Knowledge and Information Visualization*, volume 3426 of *Lect. Notes Comp. Sci.*, pages 27–35. Springer, 2005.

[55] A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967.

[56] G. J. Zelinsky, R. P. N. Rao, M. M. Hayhoe, and D. H. Ballard. Eye movements reveal the spatiotemporal dynamics of visual search. *Psychol. Sci.*, 8(6):448–453, 1997.