WILEY-BLACKWELL

Visual Data Mining: An Exploratory Approach to Analyzing Temporal Patterns of Eye Movements

Chen Yu, Daniel Yurovsky, and Tian(Linger) Xu
Psychological and Brain Sciences and Cognitive Science Program
Indiana University

Infant eye movements are an important behavioral resource to understand early human development and learning. But the complexity and amount of gaze data recorded from state-of-the-art eye-tracking systems also pose a challenge: how does one make sense of such dense data? Toward this goal, this article describes an interactive approach based on integrating top-down domain knowledge with bottom-up information visualization and visual data mining. The key idea behind this method is to leverage the computational power of the human visual system. Thus, we propose an approach in which scientists iteratively examine and identify underlying patterns through data visualization and link those discovered patterns with top-down knowledge/hypotheses. Combining bottom-up data visualization with top-down human theoretical knowledge through visual data mining is an effective and efficient way to make discoveries from gaze data. We first provide an overview of the underlying principles of this new approach of human-in-the-loop knowledge discovery and then show several examples illustrating how this interactive exploratory approach can lead to new findings.

INTRODUCTION

Recent advances in computing and sensing technologies have led to significant changes in the dominant methodology of science. Bell, Hey, and Szalay

(2009) predicted that while the first three paradigms of science—empirical, theoretical, and computational simulation—will continue to make incremental progress, dramatic breakthroughs will come from the new fourth paradigm: data-intensive scientific discovery. In brief, more and more scientific data captured by new instruments across multiple laboratories will become publically accessible for continued and deeper data analysis, which will result in the development of many new scientific breakthroughs.

Indeed, data-driven scientific discovery has already become popular in various research fields, such as earth sciences, medical sciences, biology, and physics. A survey of 1,700 top scientists across various fields, published in a recent special issue of *Science* titled "Dealing with Data" (Science Staff, 2011), shows that more than 60% of these experts have used or generated at least one data set larger than 1 gigabyte (GB). This result, along with the other articles in this special issue, suggests that tremendous opportunities for new research findings come from developing innovative ways to better analyze large data sets that not only provide a much deeper understanding of nature and society but also open up many new avenues of research.

In cognitive and behavioral research, technological advances (e.g., electroencephalography [EEG], functional magnetic resonance imaging, and motion tracking) are also leading to dramatic changes in the study of human behavior. With new technologies, we can ask and address new questions that we could not before. In this article, we focus on one particular technology: eye tracking. A state-of-the-art eye-tracking system can capture, with high spatial and temporal resolution, the moment-by-moment time course of visual attention. From such data, researchers can analyze and link external eye movements with underlying cognitive processes that support human behavior and decision making. Eye movement data are especially important for developmental science. Compared with well-established measurement methods used with adults (e.g., reaction time, forced-choice test, and questionnaires), access to the cognitive processes of preverbal infants is quite limited. Nonetheless, what we do know is that from the earliest stage of human development, infants actively move their eyes to gather information from their environments, and their eye movements are linked to internal cognitive processes.

The tradition of using gaze to study infant cognition dates back to the work of Robert Fantz. Fantz (1964) published a remarkable discovery: by 2 months of age, infants habituate to a repeated display, looking less and less at the screen on successive exposures to the same stimulus. This finding was incredible for two reasons. Theoretically, Fantz showed that from an early age, infants retain memories of visual experience over time, suggesting that early experience can have cascading effects on human development. But his methodological discovery—that eye movements can be used to access cognitive processes in young infants—was just as revolutionary. Over the

last 50 years, researchers have successfully used eye movements in young infants to investigate a stunning number of aspects of cognitive development, from low-level phenomena such as visual attention (Colombo, 2001; Hayhoe, 2004; Pereira, James, Jones, & Smith, 2010), short-term memory (Feigenson & Halberda, 2008; Oakes, Ross-Sheehy, & Luck, 2006), and perception of motion (Gilmore & Rettke, 2003) to questions about how infants track the statistical information in their world (Kirkham, Slemmer, & Johnson, 2002; Xu & Garcia, 2008), how infants learn language (Fernald & Hurtado, 2006; Graf Estes, Evans, Alibali, & Saffran, 2007; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Saffran, Aslin, & Newport, 1996; Smith & Yu, 2008), and even high-level questions about theory of mind (Farroni et al., 2005; Senju & Csibra, 2008) and the origins of concepts (Carey, 2009). All of these discoveries rely on gathering, analyzing, and interpreting eye movement data. Thus, all of these studies, regardless of their different research goals, are studies of infant eye movements.

The centrality of eye movement data in the analysis of infant cognition has inspired researchers to refine and extend Fantz's (1964) habituation paradigm (Ashmead & Davis, 1996; Aslin, 2007; Houston-Price & Nakai, 2004; Sirois & Mareschal, 2002; Thomas & Gilmore, 2004) and to develop a variety of other diagnostic tools: preferential looking (e.g., Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987), anticipatory looking (e.g., McMurray & Aslin, 2004), violation of expectation (e.g., Baillargeon & DeVos, 1991; Baillargeon, Spelke, & Wasserman, 1985), and even pupil dilation (Jackson & Sirois, 2009). Further, different metrics are used even within these diverse paradigms, for instance, proportion of total looking (e.g., Hollich et al., 2000; McMurray & Aslin, 2004), time to switch from distracter to target (Fernald & Hurtado, 2006), first look (Kovács & Mehler, 2009), and latency to look (D'Entremont, Hains, & Muir, 1997; Fernald, Perfors, & Marchman, 2006). Often these measures produce converging evidence (D'Entremont et al., 1997; Mani & Plunkett, 2010), but sometimes they are at odds (e.g., Arias-Trejo & Plunkett, 2010; Jovancevic-Misic & Hayhoe, 2009).

Taken together, there are several important reasons to develop novel ways to analyze and decode *temporal* patterns in gaze data. First, as briefly reviewed earlier, gaze is a rich source of data for developmental research—a unique window to access the cognitive systems of young children. Second, gaze data are highly complex because eye movements are produced moment by moment by a set of interacting processes: cognitive, perceptual, and motor (see also Hunter & Ames, 1988). Third, and more importantly, as eye movements unfold over time, they can be used to study real-time dynamics of the cognitive systems (Spivey, 2007). Gaze is used to actively gather information from the world, binding objects in the physical world to internal cognitive programs moment by moment (Ballard, Hayhoe, Pook, & Rao, 1997).

Decoding momentary eye movement data is critical to understand how external sensory data are linked with internal cognitive processes. In the adult literature, analyses of fine-grained temporal gaze patterns have already led to breakthroughs in real-time language comprehension (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), language production (Griffin & Bock, 2000), visual perception (Henderson, 2003), and perception-action coupling (Land & Hayhoe, 2001). Finally, a new trend in cognitive and developmental research is to move beyond well-controlled laboratory tasks and collect data in more naturalistic situations to understand human behavior in the real world (Aslin, 2009; Franchak, Kretch, Soska, Babcock, & Adolph, 2010; Frank, Vul, & Johnson, 2009; Hurtado, Marchman, & Fernald, 2008; Smith, Yu, & Pereira, 2011). Compared with most studies conducted in well-controlled experimental environments, this research direction with more ecological validity not only further tests and confirms whether existing findings inferred from well-designed experiments also play an essential role in more naturalistic situations but also leads to new discoveries that cannot be captured in well-controlled studies. This research venue, however, also adds particular challenges in data analysis as studies along this line usually collect data over a longer period of time—that is, more data. In addition, fewer constraints and free behaviors (e.g., free viewing) in more naturalistic situations also mean that many factors in such environments jointly influence gaze behavior—that is, more complex data. Both characteristics suggest that one may not know in advance exactly what moments should be focused on in data analysis, how to dissect multiple factors, and how to select, group, and compare data to extract meaningful, novel, and statistically reliable patterns.

So how can we analyze such data? Its complexity may initially suggest that one needs comparably complex statistical or mathematical analyses, which in turn require one to develop a deep understanding of sophisticated mathematical methods. Most often, however, the more sophisticated a method is, less accessible it is as fewer people would be able to understand and use it. This article presents an alternative solution—a more transparent approach to integrating top-down domain knowledge with bottom-up information visualization and visual data mining. The key idea behind this method is to leverage the computational power of the human perceptual system for pattern discovery. In this approach, scientists iteratively examine and identify underlying patterns in data and link those discovered patterns with top-down knowledge/hypotheses. This method relies on the combination of bottom-up data visualization and top-down human theoretical knowledge working together through visual data mining to lead to new findings. In the following sections, we first give an overview of the underlying principles of this new approach and then provide several examples illustrating how it can lead to new discoveries.

INTERACTIVE DATA ANALYSIS AND VISUAL DATA MINING

An interactive data analysis and visual data mining approach addresses two major challenges in analyzing gaze data. First, whenever we draw conclusions from gaze data, the final published results take the form of overall statistics derived from high density gaze streams. Thus, a huge amount of information must be reduced and summarized to be useful. But statistics and measures extracted from raw data may exclude embedded patterns or even be misleading, because multiple steps are involved in reducing raw data to final statistics and each step involves decisions about both statistical methods used to compute final results and the parameters of those methods. Different decisions may change the final outcome dramatically, leading to different results and consequently different interpretations (see also Shic, Scassellati, & Chawarska, 2008). We argue that sensible decisions about microlevel data analysis cannot be completely predetermined but must be derived at least in part from the structure of the data itself. Most often, even if we have some predictions from our experimental designs, we nonetheless lack precise predictions about the structure and patterns of gaze data at the microlevel. As we cannot specify all of the data analysis details a priori, we need insights from both top-down knowledge and raw data themselves to make sensible decisions step by step as we systematically reduce the data to extract reliable and interesting patterns.

Second, one important goal of analyzing gaze data is to find new patterns and gain new knowledge from such data. But how can we discover new and meaningful patterns if we do not know what we are looking for? Although standard statistics may find a subset of these meaningful patterns, they may miss a great deal more. We suggest here that visual data mining is a powerful approach for new pattern discovery. Discovering *new* knowledge requires the ability to detect unknown, surprising, novel, and unexpected patterns. To achieve this goal, our proposed solution is to rely on visualization techniques that allow us to easily spot interesting patterns through both our visual perception systems and our domain knowledge. This solution is feasible as it is based on two grounds. First, our eyes are quite sensitive to visual patterns. Studies in information visualization have convincingly demonstrated that humans have remarkable perceptual abilities to scan, recognize, and recall details from visually displayed information, and to detect changes in size, color, shape, or texture in visually displayed data through just a few quick glances (Shneiderman, 2002). Second, when our eyes detect patterns, our mind can naturally interpret those patterns and link them with current theories based on our domain knowledge. Further, in their book *Using Vision to Think*, Card, Mackinlay, and Shneiderman (1999) point out the crux of visual data mining—visualization techniques are a valuable way to

visually present data to reveal insights, but that discovery takes place between the ears. That is, major advances and significant new discoveries are made by people rather than directly by machine learning or visualization techniques. In practice, these techniques play a complementary role in assisting scientists to better understand data. Toward this goal, visual data mining provides an ideal framework for active exploration and sensible reduction of gaze data through human-in-the-loop knowledge discovery.

Consequently, the exploratory process should be highly iterative and interactive—visualizing not only raw data, but also the intermediate results of current statistical computations for further analysis. In this way, statistical computing and visualization can bootstrap each other—information visualization based on new results leads to the discovery of more complex patterns that can in turn be visualized, leading to more findings. Researchers play a critical role in this human-in-the-loop knowledge discovery by applying statistical techniques to the data, examining visualization results, and deciding/directing the research focus based on their theoretical knowledge. In this way, domain knowledge, computational power, and information visualization techniques can be integrated together to understand massive data sets. In the rest of this article, we present several examples of exploring both raw gaze data and derived patterns through exploratory data analysis.

We note that the goal of this article is to introduce the concept of interactive data analysis, present the general procedure of visual data mining in the context of gaze data, and demonstrate how using this interactive and exploratory approach can lead to reliable and fruitful results. In practice, the idea of visual data mining and interactive data analysis can be implemented in a variety of ways, some of which may rely on advanced computational and visualization algorithms. For the purpose of this article, two visualization techniques used here are intentionally simple and accessible through popular data analysis software: (a) visualizing continuous numerical data streams (e.g., Figures 1, 3, and 4) and continuous categorical streams (e.g., Figures 4 and 5) with different colors; (b) highlighting critical moments in time and temporal events by adding transparent blocks with different colors on top of the underlying continuous data streams (e.g., Figures 2, 3, 5, and 6). These two visualization effects are provided by most data analysis software. For example, Matlab (The MathWorks Inc., Natick, MA) and R provide drawing functions to assign color and other attributes of lines and blocks. MS Excel (Microsoft Corp., Seattle, WA) also allows users to manually set up display properties (e.g., color, texture, and transparency) through graphical user interfaces. With these two simple visualization techniques, we focus on how to visualize data and interactively explore potential patterns by following the principle/spirit of human-in-the-loop data exploration advocated in this article. In addition, more and more open-source programs (e.g., Cyber-

infrasture system, Börner, 2011; GUESS system, Adar, 2006) are becoming available for this kind of intelligent data analysis (Yu, Smith, Hidaka, Scheutz, & Smith, 2010).

VISUALIZING RAW DATA—THE FIRST STEP

Most commercial eye-tracking systems provide data at multiple levels: raw gaze data files, derived measures such as user-defined areas of interest (AOIs), fixations/saccades, and accumulated statistics such as overall looking aggregated over time. In addition, such systems may also provide additional data preprocessing functions such as low-pass filter functions for smoothing, or a gap fill-in function using interpolation to fill in missing data (e.g., Tobii Studio software). Although these functions are necessary and useful for reducing the amount of noise in eye-tracking data, it is also important for researchers to understand how these functions may unexpectedly distort raw data. More importantly, it is tempting to use just preprocessed data (derived measures and summary statistics, etc.) without accessing raw gaze data, but this may miss critical microlevel structure embedded in gaze data. Thus, it is both informative and necessary to examine raw data as the first step of eye movement data analysis. This idea is similar to graphing histograms to check for normality before conducting an analysis of variance. More generally, what is being discussed here is to access and examine raw data all the time—not directly jumping into the summarized data, but instead getting as close as we can to the raw data to "see" what empirical data can tell us.

We illustrate this point by using an example in which a group of 8-month-old infants watched a clip of a performer producing infant-directed actions. The performer demonstrated how to put a set of shapes into a shape sorter—a big plastic ball with holes matched to each of the shapes. One demonstration of putting a shape into the big ball consisted of a sequence of actions, starting from picking a shape, to showing it to infants, to rotating the ball to find the right hole, and to finally putting the shape into the hole. A similar sequence was repeated with a new shape as the target each time. As part of this naturalistic infant-directed demonstration, the performer also simultaneously and frequently generated other actions and social cues to engage infants, such as smiling, talking, waving her hands, and looking at infants' faces. Just like natural behaviors in the real world, these actions interweave with each other—sometimes two actions were sequentially generated without a clear boundary to separate the two; sometimes more than one action was conducted by different modalities (e.g., smiling while holding a shape).

Compared with participants in many eye-tracking studies designed to measure their response to a particular stimulus at a single moment, infants in this example needed to continuously decide to what information they attended moment by moment. Different kinds of research questions can be addressed by analyzing gaze data in this example, such as how infants perceive goal-directed actions, in what ways they are sensitive to social cues generated by the performer, how they track moving objects, and how they respond to the performer's momentary bodily actions. Importantly, these questions are studied in a naturalistic context in which each action generated by the performer is embedded in streams of sequential actions from the same modality and concurrent actions from other modalities. Sequential multimodal behaviors produced in more naturalistic studies pose a particular challenge in data analysis compared with discrete actions in separate trials. Basically, where should we start in data analysis? Which types of actions, based on our prior knowledge, may be the best candidates to obtain reliable patterns from the observers' gaze data? Would the infants still generate predictive looking toward objects with the co-presence of social cues from the performer? Would the infants still be very sensitive to social cues in the context of ongoing manual demonstration? How would they switch their attention between the performer's face and objects and would these switches share similar patterns across individual infants? Would parameters used in previous data analyses still be applied to this study? Would patterns extracted in studying a single discrete action (e.g., timing of predictive looking) still hold or would cognitive loads in this more naturalistic (and more complex) context require additional resources and additional time to process information and respond? One might come up with so many questions that it would not be practically feasible to try out all of these ideas one by one in an ad hoc way given the amount of efforts required in implementing and examining each idea. Instead, we need a more efficient and systematic way to quickly identify a few questions among those that are more like to lead to interesting and reliable patterns. In the following sections, we will use this example, as representative of infant eye-tracking studies, to illustrate how we can interactively and visually examine infant eye movement data from more naturalistic studies.

Figure 1 visualizes several streams of gaze location data recorded from participants. Each colored stream shows the same location variable (y-coordinate, etc.) from one of the participants as s/he watched the action sequence. Most often, visual attention was concentrated on either the performer's face (high y values) or objects on the work space (low y values). Given the output of many eye-trackers and eye-tracking software, it is relatively easy to quickly calculate predefined statistics, for instance, average looking time to the ball or the performer's face or hand.
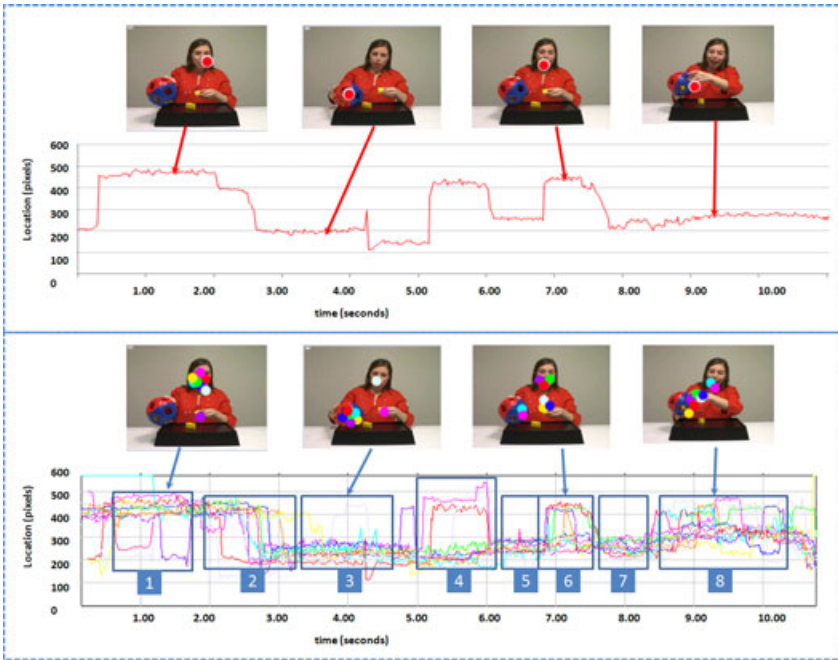
Figure 1 Visualizing raw data streams collected from one (top) and multiple (bottom) participants. By visually exploring multiple streams in parallel (both vertically and horizontally), researchers can easily detect the shared gaze patterns across participants—those temporal moments that most participants generated the same behaviors or perceived the same information. Meanwhile, the gaze data also demonstrate individual differences. Researchers can also identify those individuals who diverge from the whole group and examine when such deviations occur. Moments labeled (1), (5), and (7) reveal consistent gaze data while moments labeled (2), (3), (4), (6), and (8) show individual differences.

Although such results are certainly informative, visualizing the data and developing new hypotheses about how participants processed the information moment by moment will allow for richer and more fine-grained analyses of the infants' behavior. In particular, as all of the data streams in this example are temporally aligned (watching the same video, etc.), we can easily compare the time series to finding both shared, frequently occurring patterns and as well as surprising patterns *across* participants to inform and guide further quantitative data analysis. In the following section, we will discuss several different aspects of gaze data observed through raw data visualization.

Group Behavior

Participants as a whole group shared an overall trend: first, looking at the performer's face at the moment labeled (1), then, switching to objects at the moments labeled (2), and then, consistently looking at those objects with occasional glances back to the face at the moment labeled (3). Shared patterns can be easily spotted by examining multiple time series vertically and identifying when multiple streams merge into a big "river." This observation through visualization is informative in two ways. First, it tells us at what moments the mean measure is both meaningful and representative of the whole group. Second, the elaborate relationship between the performer's actions and consistent eye movement patterns across participants at those moments suggest further analyses. For example, one may ask questions about the latency to look at the face (rather than just the mean duration of fixation to the face). Further, one may compare the latencies between looking at the face with looking at a target object, or more generally, latencies between different events (e.g., the performer's smile as a social event, a speech act or a manual action) may reveal how different kinds of information (e.g., social versus nonsocial; visual versus auditory) were processed in these events. It is worth noting that the example illustrated in Figure 1 includes only eight participants for the illustration purpose, though in practice, one may need to work with gaze data from a much larger number of participants. The overall pattern from a large group can be similar to that of a small group—a main "river" with minor rivulets; or alternatively, there may be several "rivers," each formed by a subset of participants. In the later case, one can select those individuals sharing similar patterns to be a subgroup—first studying each subgroup one by one, and then comparing patterns across subgroups.

Individual Differences

This visualization also provides an easy way to see which individuals differed from the group's average behavior. Further, we can find easily *at what moments* and *in what ways* those outliers allocated their attention differently. For example, at the moments labeled (1), when the performer looked toward observers, all but two of the participants looked to the performer's face. The other two participants (in purple and orange red) maintained fixation on an object at different times. Similarly, only two participants (in red and pink) looked at the performer's face at the moments labeled (4) while others attended to the performer's hand actions and objects. In addition, we can also identify some local patterns. For example, at the moments labeled (6), when the performer was speaking while simultaneously holding an object,

approximate half of the whole group attended to her face as a response to her speech and facial expressions, while the rest continued to fixate on the objects. Without seeing and knowing this essential bimodality, we would be misled by using the mean of gaze locations to represent the whole group. Thus, inspection of raw data makes us aware of which assumptions hold, and which are violated. This data visualization serves as an example to illustrate the importance of accessing raw data so that we can better calculate and interpret derived statistics. One may argue that an alternative way to achieve this goal would be to extract various statistics (e.g., in addition to mean, also median, 25 percentile, 75 percentile, histogram, or more complicated statistical measures). Taken together, those statistics would ultimately allow us to obtain a better picture of the overall distribution of the data. However, the power of information visualization is that we can easily and quickly see and notice regular and irregular patterns with just several glances.

Temporal Acuity

Humans actively move their eyes to gather just-in-time visual information for real-time internal processes. Thus, examining the dynamics of eye movements has yielded many insights into how human cognitive and perceptual systems work (Land & Hayhoe, 2001; Tanenhaus et al., 1995). Visualization of raw data shown in Figure 1 allows us to see the exact timing of attention switches between the face and objects. For example, moments labeled (6) show close synchrony across participants in switching their attention to the performer's face. Thus, based on visualization, we learn that extracting and interpreting patterns based on precise timing between the performer's actions and participants' consistent responses at this particular moment may lead to interesting results. In another case, at the moments labeled (2), all of the participants switched their attention from the performer's face to the objects, but the exact timing differed across participants, spanning a window of approximately 1.5 sec. This suggests that timing measures extracted at this moment can serve as a good candidate to capture individual differences that might predict other stable behavior patterns. In another moment labeled (8), when the performer demonstrated how to put the piece into a hole while generating an exciting sound to engage observers, in a window of 1.5 sec (from 8.5 to 10 sec), participants mostly focused on the actions and objects but occasionally glanced at the performer's face. The exact timing of these glances varied across participants. This visualized pattern suggests that: (a) within this particular period of time, they seemed to consistently go back to check the performer's face at some time; and (b) with different timings, they seemed to spend a similar amount of time on the

performer's face. Therefore, a reliable pattern in (8) is the average duration of face look (but not the exacting timing of face look). More generally, these observations can inform follow-up decisions about what moments and what aspects of gaze patterns (e.g., attention switches, timings of those switches, or gaze durations) should receive more rigorous analysis. What we want to highlight here is the value of visualizing and examining raw data to reveal both regular and irregular patterns in raw data, and consequently, to make more informed decisions. If we skip this step and simply derive some statistical measures without taking the raw data into account, statistical analyses alone would be noisier (at best) or misleading (at worst).

Spatial Acuity

Human visual acuity varies across the visual field. The foveal region occurs at the center of gaze having the highest acuity that is concerned with high spatial frequency information with fine details, and the periphery is used to evaluate coarse information with low spatial frequency. In light of this, in addition to microlevel temporal patterns, raw gaze data provide high-resolution spatial information—exact gaze locations but not just AOIs. For example, at the moments labeled (1), even though most participants looked at the face, they fixated on different parts of the face (with different y values). This provides useful information on the following AOI definition—should the whole face be defined as an AOI or should different parts of the face (e.g., eye and mouth areas) be individual AOIs? Thus, as defining AOIs and reducing raw gaze data into AOI streams is a necessary step in most data analyses, examining raw gaze locations can lead us to make more careful decisions on finding appropriate spatial resolutions to define AOIs without losing useful information.

To summarize, visualizing raw data allows researchers to quickly gather information on various aspects of eye movements, such as group behavior, individual differences, and temporal and spatial patterns. This information is critical to subsequent statistical analysis in two important ways. First, the information encoded in raw data may be filtered in high-level derived statistics. For instance, AOI-based representations do not have high spatial resolution. Similarly, accumulated looking times do not contain temporal information that can be critical in many perception, vision, and action experiments. By visually examining raw data, we can get a better sense of inherent properties of gaze data which can provide critical insights to guide us to determine which aspects of the data to focus on next and which sensible parameters to be chosen in the following data analysis. Second, data visualization and examination is an efficient way to

conduct exploratory data analysis as we can simultaneously detect and examine multiple potentially interesting patterns and then actively connect what we see with top-town theories and principles to identify gaze patterns that either confirm our hypothesis and support existing theories or discover new patterns and new findings that may ultimately lead to new advances. Subsequently, these patterns need to be quantified using mathematical and statistical techniques. But the real intellectual challenge in this venture comes not from quantification, but from initial exploratory pattern discovery.

EVENT-BASED TEMPORAL PATTERN DISCOVERY

Given a first round of visually examining raw data, the next step is to investigate further and extract reliable patterns. The idea of exploratory gaze data analysis we propose here follows the general principles of building scientific data visualization systems: ''overview, zoom & filter, details-on-command'' (Shneiderman, 1996, p. 337, 2002). More specifically, we start by examining raw data and then zoom into particular moments for closer investigation. Further, this process is iterative; we may go back to the overview mode and select another subset of data to be closely examined.

This exploratory data analysis is implemented through event-based pattern discovery. The central idea is to define *moments-of-interest* using a temporal event representation and then study how eye movement data dynamically change within and outside of individual events. Formally, an event variable $y = \{<y_1, t_{s1}, t_{e1}>, <Y_2, t_{s2}, t_{e2}>, \ldots <y_m, t_{sm}, t_{em}>\}$ is a set of time intervals with onset t_{sm}, and offset t_{em} timestamps of each instance of a specific event y_m. For example, the time intervals when a certain visual stimulus is displayed can be represented as an event variable; the time when a sound or a spoken word is played can be captured as another event; and the moment that an performer is smiling or waving her hands can be defined as another event. In practice, an event variable can be defined both from top-down knowledge and experimental design, and from bottom-up patterns discovered from raw data. In general, there are four ways that an event can be defined and extracted:

- Events based on experimental design: In a typical eye-tracking experiment, visual and auditory stimuli are presented at certain moments in time. Therefore ,we can define events based on stimulus presentations to capture appearance and disappearance of certain visual stimuli, onsets and offsets of spoken words, and durations of executed actions (smiling or laughing).

- Events based on top-town knowledge: Based on research questions asked in the study, we most often have specific hypotheses and predictions about what moments in time might be interesting. Predictions based on top-down knowledge can be captured by event representations to allow us to further examine exactly what happen at those moments. For example, if we predict that participants may look at the target hole even before the performer put the corresponding shape into the hole, we can create an inserting-shape event and examine gaze patterns right before this event.
- Events based on bottom-up patterns: Events can also be extracted based on bottom-up patterns observed in examining raw data. For example, the highlighted areas in Figure 1 (labeled 3, 5, and 7) can form a new event depicting those moments during which the majority of participants consistently gazed at the same locations. Then we can use that event to determine what leads to this consistency and why it ends. Thus, we define an event based on a certain aspect of gaze data and then use it to further study other aspects of gaze data.
- Combined events: In addition, we can generate high-order events (e.g., AˆB) based on two basic events A and B, each of which can be either a top-down or bottom-up event. For example, a combined event can be defined as the moments that the performer is smiling while talking (i.e., combining the smiling event with the talking event), or the moments during which both a target visual object is displayed and a spoken word is heard (i.e., combining a spoken word event with a visual object event).

As shown in Figure 2, events can be presented as colored bars, with lengths corresponding to their durations. Visualizing temporal events themselves allows us to examine how frequently each temporal event happens
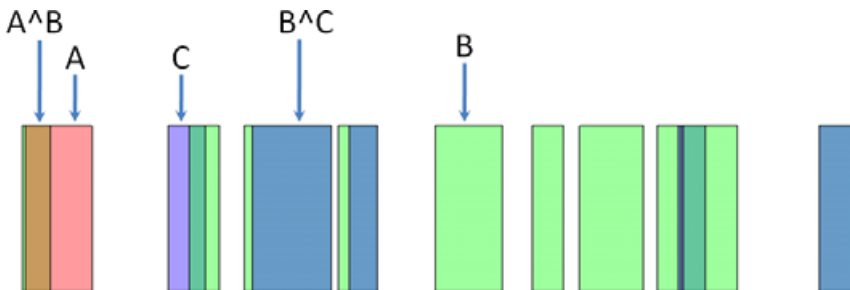

Figure 2 Visualization of temporal events. Each event type is represented by one unique color. The overlapping between colors creates a new color indicating a logical conjunction of more than one event.

over time, how long each instance of an event takes, how one event relates to other events, and whether an event appears more or less periodically or whether there are other trends over time. All of those potential (repeated or unique) patterns can be quickly examined by scanning through the visualized event data streams. We can then connect these patterns back to top-down research questions and hypotheses. Moreover, we can see not only an event itself but also the conjunction of two events because the overlapping of two colors/areas will generate a new color. For example, if two binary events A and B are labeled as ''red'' and ''green'' then we can see four color categories: red—''A^¬B,''   green—''¬A^B,''   yellow—''A^B,''   and white—''¬A^¬B.'' In a more general case with more events, a color-based event visualization scheme is still able to represent all possible logical conjunctions of those events.

Further, we can exploit this technique to find complex patterns hidden in continuous gaze data *conditioned* on temporal events—what trends and patterns can be extracted from continuous gaze data when certain events happen. Our approach is to overlap color bars representing events on the top of continuous gaze streams. Figure 3 shows multiple raw gaze streams overlapped by several events, allowing us to visually spot potential patterns across those data streams and to examine how those patterns are affected by the events.

In particular, as temporal events define important moments over time (both onsets and offsets), we can pay special attention to those highlighted moments to discover potentially interesting temporal patterns. For instance, at around the 7 sec mark in Figure 3, right before the performer started speaking, a subset of participants already started switching their attention to her face. If participants looked at a target location even before a stimulus was presented there or a motion event happened there, we would be able to see the allocation of their attention before the onset of an event to detect potential predictive looking behaviors. As predictive looking behaviors may happen in different ways across participants (some
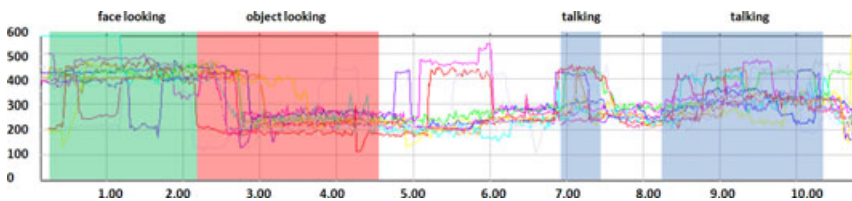

Figure 3 Visualizing raw gaze data (generated by participants) with multiple temporal events (generated by the performer) allows us to compare underlying patterns of continuous gaze data within and across different events.

earlier and some later), visualizing all of the data allows us to capture both the overall pattern and potential individual differences. Similarly, we can visually examine responsive looks and gaze reaction times by zooming into the moments right after the onset of an event. We can measure latency to shift gaze to the target and duration of fixation. For instance, while the performer switched her attention to an object after face looking (around 2 sec in Figure 3, from the ''green'' event to the ''red'' event), participants also all followed her gaze toward the target object in joint reference. Through visualized data, we can see quite different latencies across participants in this attention switch, indicating that further data analysis should take individual differences into account. In addition, we can also examine whether participants maintained sustained attention at a certain location while an event was on and whether their attention drifted away when the event ended. For example, during the face-looking moments (in the ''green'' event), participants shown in Figure 3 consistently looked back at the performer's face and by doing so created more eye contact moments. For another example, during the first talking event in Figure 3 (around 7 sec), participants switched their attention away from the face approximately at the same time that the talking event ended.

While predictive looking, responsive looking, and sustained attention are well studied topics, summary statistics may miss important patterns because of the complexity of spontaneous eye movements and individual variation. Visual data mining provides a unique way to quickly examine data to gain insights about which moments (before, after, and during) should be focused given a certain event, what data should be included (or excluded) to capture certain patterns, what appropriate statistical methods should be used to compute statistics, and what appropriate parameters (e.g., what timing threshold to capture predictive looking) should be selected. Without such insights from examining microlevel data, just applying predetermined approaches with standard and default parameters (e.g., those developed by eye-tracking companies for general purposes) may not be able to accurately capture elaborate patterns in real-time gaze data. To further illustrate this point, we present two data processing techniques to capture two different aspects of temporal patterns that leverage event-based exploratory data analysis to move beyond the raw data.

Temporal Profiles

This representation is based on one used in psycholinguistic studies to capture overall temporal dynamics across a related class of events (Allopenna, Magnuson, & Tanenhaus, 1998). The basic implementation consists of two simple steps: (a) taking multiple gaze streams synchronized in time and then

computing the aggregated value at each time point; and (b) plotting a temporal profile by putting together aggregated values from individual moments and overlapping event bars on the profile. Continuous data streams can take two forms. In the case of numerical data stream (e.g., raw x and y coordinates as shown in Figure 1), the aggregated value is the mean of data points at a given time. In many other cases, we first convert (x,y) into a categorical value indicating an AOI. To follow the examples in Figures 1 and 2, Figure 4 (left) shows AOI streams from a group of participants wherein two colors are used to represent looking at two different AOIs (the performer's face or an object). Figure 4 (middle) shows the results of temporal profiles by aggregating data across participants. Each data point in a temporal profile represents an estimated probability that participants as a group gazed at a certain AOI (a face or an object) at that moment. Temporal profiles provide easy access to information about the dynamics of attention in a task, and how these dynamics play out in concert with the input stimuli. For instance, consensus can be seen in the moments when a majority of participants looked at the face AOI (e.g., 90% in Figure 4 means 9 of 10 participants gazing at that location). Real-time attention dynamics can be seen in the group transition from focusing on one AOI (e.g., the face) to another (e.g., the object). In this example, two AOIs can be viewed as competing for participants' attention, and we can see how this competition emerges over time. Finally, the interaction of group attention dynamics with events can be seen by overlapping
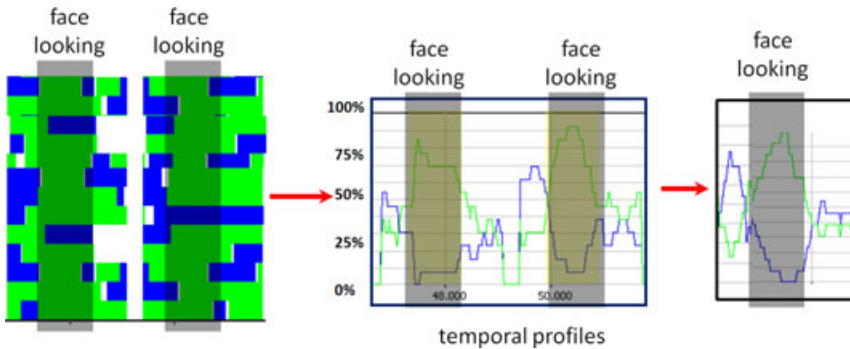

Figure 4   From multiple areas of interest (AOI) streams (left), two probabilistic temporal profiles (middle) are computed, which indicate the probabilities of looking at each AOI at a particular moment in time. In this way, the overall dynamic trend from multiple profiles can be examined together. In addition, overlapping events (shaded areas) with the temporal profiles highlights interesting moments in time, which allows us to see the patterns within those instances as well as the overall trend before and after events. Next, two instances from the same "face-looking" event are integrated to generate aggregated temporal profiles (right) around event "face looking."

event bars with temporal profiles. For example, before the onset of the "face-looking" event wherein the performer looked up toward observers, participants looked more toward the object AOI; and right after the onset of the "face-looking" event, there was an increase in looking at the face AOI, suggesting that participants looked back to the performer in response to her action. All this makes temporal profiles from a whole group an informative intermediate step from raw data to final statistics.

An additional step is to aggregate temporal profiles across multiple instances of the same event to generate accumulated profiles for that event type. As shown in Figure 4 (right), gaze data from multiple instances of the same event are aligned (to maintain timing relative to the event) and integrated to obtain aggregated probability profiles. The alignment can be based on either the onsets or the offsets of individual instances of the event depending on their relative importance to the research question. By integrating multiple instances while keeping relative temporal information with regard to a certain event, the accumulated temporal profiles can reveal statistically more reliable (and stronger) temporal patterns. In the example shown in Figure 4 (right), we can see a clear trend of attention from the object to the face and then back to the object around the face-looking event when the performer initialized a face look and then looked away.

Event-Driven Response Timing

As temporal profiles provide an overview of the dynamic trend of gaze patterns over time from the whole group, they cannot provide detailed information about when participants switched their gaze from one area to the other. The precise timing information is critical as it reveals just-in-time information requested by internal cognitive processes (Hayhoe & Ballard, 2005). In many cases, it is informative to know how participants respond to a certain event and whether they adjust their attention during the event (e.g., the onset and offset of a spoken word, the appearance and disappearance of visual stimuli, or the starting and ending of an action). For instance, participants might not attend to the target at the onset of the event, but only after they perceive the event as a response to such an event. Therefore, we can measure whether they switch to the target after the event onset by computing the timing latencies between the event onset and the first fixation on the target. Figure 5 shows timing latency histograms from two events, showing participants responded to event B more promptly than to event A. Again visualizing these results allows us to discover not only whether two timing latencies are different, but more importantly to closely examine *in what ways* they differ. For example, with this visualization, we can easily see that a
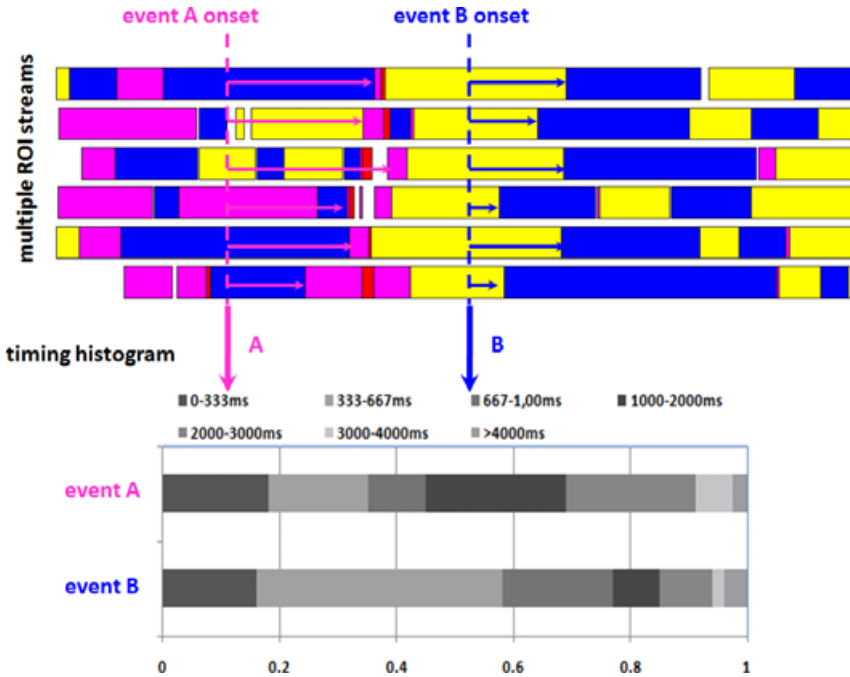
Figure 5 Computing and visualizing timing histograms. Top: timing latencies (arrows, etc.) from the onset of an event (indicated by a vertical dot line) to the moment of gazing at a certain areas of interest. There are two event onsets in this example. Bottom: two timing latency histograms are computed by gathering all the latencies from individual participants, and a comparison between the two histograms to two events show the differences of how promptly participants reacted to these two events.

large proportion of participants generated their first fixation on the "blue" AOI as a response to event B within a timing window between 333 ms and 667 ms. In contrast, their responses to event A vary across individual participants—more or less evenly distributed up to 3,000 ms.

In summary, temporal profiles and event-driven response timing are two complementary measures—one focusing on the overall dynamics over time from a whole group and the other focusing on precise timing between events and gaze data from individual participants. These bottom-up patterns, when extracted, can be rendered with meaningful interpretations based on top-down knowledge. In this way, this interactive event-based exploration with informed decisions from the combination of top-down knowledge and bottom-up patterns in data will ultimately lead to clean, meaningful and reliable results.

Grouping and Comparing Gaze Data

In most studies, results take the form of a comparison between gaze data from different conditions, or between an experimental condition and a base-line condition. This requires the whole data set to be divided into groups/subsets, and then, the behaviors from different groups to be compared. Shared patterns between groups can be viewed as consistent and reliable gaze behaviors as different patterns show the impact of factors that are used to define grouping. Given various ways to extract different patterns from raw data as intermediate results (e.g., temporal profiles or timing histo-gram), a necessary next step toward generating final statistics in this frame-work of exploratory data analysis is to form different groups and compare patterns between groups.

While grouping data based on different experimental conditions is straightforward, data can also be grouped in other ways: either using addi-tional information independent of gaze data or gaze data themselves. In the first case, most studies collect additional measures along with gaze data. For instance, the MacArthur-Bates Communicative Developmental Inventory (CDI) may be used to measure the vocabulary sizes of participants. In some studies, EEG or psychophysiological measures are also taken together with gaze data. These measures can be first analyzed separately and then subse-quently used to group gaze data for further data analysis. Alternatively, those measures may define some critical moments in time which can be used to divide continuous gaze data into several segments and then group and compare those segments. Moreover, one can group data based on existing patterns extracted from data. For instance, participants can be grouped based on their gaze behaviors at a certain moment to predict gaze behavior at other moments (see good examples from Amso & Johnson, 2006; John-son, Davidow, Hall-Haro, & Frank, 2008; Johnson, Slemmer, & Amso, 2004). In particular, some experiments are designed with two stages: habitu-ation/training and testing. Therefore, testing results can be used to group gaze data during training to draw inferences about learning processes.

Here, we use an example from Yu and Smith's (2011) cross-situational word learning study in which young infants were exposed a sequence of training trials followed by a testing phase. In studies such as this, one can group individual participants in different ways. For example, gaze data at test can be first analyzed based on preferential looking (Golinkoff et al., 1987) by counting the proportion of accumulated time looking at the correct referent when hearing a corresponding word. This learning result (based on gaze data at test) can then be used to group participants into two group-s—strong and weak learners. Alternatively, one can group participants based on gaze behaviors in early training trials to see whether these

behaviors are predictive for where they looked in later training or at test. Further, one can group participants based on other measures such as their CDI scores to examine whether gaze behaviors generated in this laboratory learning task relate to vocabulary development in the real world. For each grouping, different gaze patterns (e.g., number of fixations, duration of fixations, latencies) can be extracted within a group and then compared across groups. Now that we face with another challenge: given various ways to group data and given various patterns that can be extracted within individual groups, what kinds of grouping on what patterns are both meaningful and interesting? Similar to other data analysis challenges discussed in the article, the goal here is to use top-down knowledge to quickly examine the results from a variety of possible combinations of groupings and extracted patterns within groups, and identify those that may lead to reliable results.

Figure 6 shows two extracted patterns from one grouping—categorizing participants into strong and weak learners based on testing results. More specifically, Figure 6 (top) shows the average number of eye fixations over individual training trials. Both infants who show strong learning at test (the strong learner group) and those who show weak learning at test (the weak learner group) have similar eye movement patterns at the beginning, but roughly around 3–5 trials into the experiment, their looking patterns begin to diverge. Weak learners generated more and shorter fixations while strong learners maintained more stable attention switches through the whole training. At the end of training, both strong and weak learners had similar overall attention switches again. Figure 6 (bottom) shows the average lengths of the longest fixations over training trials. A longer fixation may indicate more stable attention and therefore better learning. Infants who showed stronger learning at test consistently generated longer fixations trial by trial while the infants who showed weaker learning at test always produced shorter fixations especially at the end of training. The overall results from Figure 6 suggest that strong and weak learners show quite different gaze patterns from the beginning to the end of learning. On the early learning trials, weak learners tend to generate more and shorter fixations. However, the most significant differences are in the middle of learning; weak learners generate more attention switches than strong learners. Moreover, given the approximate same number of attention switches (given that there are no major differences in the longest looking duration), the other fixations generated by strong learners are more stable than those in weak learners. Finally, the end of training is clearly characterized by longer fixations by the stronger learners. As more details on both extracted patterns and interpretations of those patterns relating to how underlying attentional processes may lead to better or worse word learning can be found in Yu and Smith (2011), the main purpose here is to illustrate visualizing these patterns from the two learner
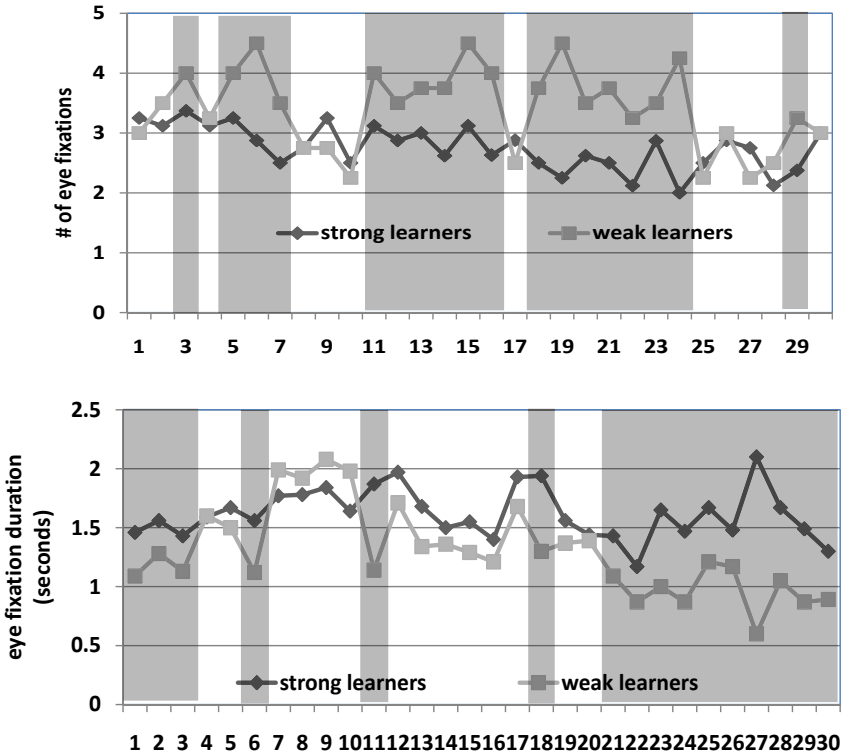
Figure 6 Participants are grouped as strong and weak learners. Three metrics are extracted from gaze data in individual learning trials: (top) the number of fixations in each learning trial; and (bottom) the average of the longest fixation durations within learning trials. Shaded areas highlight those moments that are statistically significant between the two learner groups.

groups facilitate exploratory data analysis to allow us easily examine, identify, compare, and interpret potentially interesting patterns.

GENERAL DISCUSSION

Data-driven knowledge discovery is a promising new trend in scientific research, already producing new findings and theories across a number of fields. The central problem in this endeavor is the discovery of novel patterns from large data sets. Even when we have theoretically motivated predictions at the macrolevel, we may not be able to perfectly predict fine-grained gaze patterns at the microlevel embedded in high density eye gaze data. The

solution lies in developing a systematic way to explore complex data and discover interesting and unexpected patterns that we do not know in advance. Visualizing raw data and intermediate results is a powerful way to facilitate this data exploration process which allows researchers to interactively explore potential interesting patterns visually displayed.

Visual data mining approaches can be viewed as an integration of information visualization techniques with human domain expertise. The goal is not to develop a computer program or model that can automatically extract such patterns, but rather to leverage researchers' visual pattern recognition and domain knowledge. Human-in-the-loop data analysis involves active and iterative examination and exploration of visually displayed patterns to select useful information and guide further knowledge discovery. This practice of visual examination and visual pattern matching has been used in various domains. For example, when presented with two fingerprint images and asked to perform a matching task, a novice without any fingerprint examination experience is unable to perceive much information from two seemingly similar fingerprint images. A fingerprint expert viewing the same two fingerprints, on the other hand, can confidently identify key visual features from those two images and compare and match those features to reach a conclusion (Busey & Parada, 2010). Similar to fingerprint experts, researchers studying human development and cognition possess domain knowledge from their training. When presented with visualized gaze data, our low-level system can easily identify (expected or unexpected) patterns, and our domain expertise can render those patterns with scientific interpretations. Moreover, through interactive visualization, we can further look for specific patterns to iteratively and systematically decipher the information revealed by complex gaze data.

Toward this goal, this article does not intend to offer a comprehensive review of all of the visualization methods available, but rather serves as an introduction of visual data mining of gaze data. There are two important take-home messages. First, analysis of complex data, such as eye movements, involves multiple processing steps between raw data and final statistics. Because different decisions at intermediate steps can result in different conclusions, and the information required to make sensible decisions is rarely fully available a priori, one effective way to ensure correctness of data processing is to look back to raw data and compare intermediate results with the raw data from which those results are derived. Second, to produce new findings from complex data, we need to conduct interactive data analyses by integrating top-down domain knowledge and bottom-up pattern discovery. While many machine learning and computational modeling techniques are available for analyzing behavioral data, they cannot simply be used ''off the shelf'' to automatically find new patterns that we do not yet
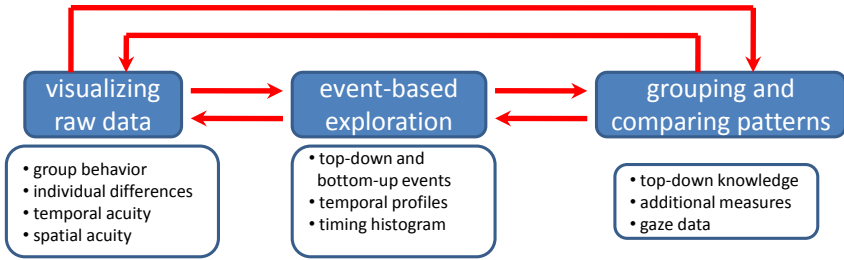
Figure 7 Interactive data mining of eye movement consists of three components/steps: visualizing and overviewing raw data; closely examining moments of interest on some portion of data selected based on events; and finally grouping and comparing patterns to compute statistics.

know. Compared with "blind" data mining, what we suggest in this article is that researchers with top-down theoretical knowledge need to be in the loop of data mining and data analysis. Exploratory data analysis requires theoretical guidance to be practically feasible in a sea of complex data. In fact, we argue that the most intellectually challenging part is for researchers to gain insights from data by linking domain knowledge with extracted patterns. For this purpose of intelligent data analysis, data visualization and examination provide an ideal platform to facilitate this process.

Based on the two principles summarized earlier and inspired by the visual information mantra "overview first, zoom and filter, then details-on-demand" (Shneiderman, 1996, p. 337), this article describes an interactive data mining method to analyze temporal behavioral data consisting of three components as shown in Figure 7:

- Visualizing raw data allows us to gain an overview of the entire data set to compare raw data across participants and identify both consistent patterns across participants and unique patterns at certain moments. This step serves two important roles: it gives us insights on what aspects of gaze data should be further explored, and what computational methods would be appropriate to capture the patterns observed.
- Event-based temporal pattern discovery allows us to zoom in and systematically and closely examine temporal patterns at certain moments.
- Grouping and comparing patterns is the final step to generate statistically reliable patterns and compare these patterns between different subgroups and experimental conditions.

In each component, we introduced a set of tools and guidelines with examples on how to effectively explore new patterns that may lead to new

discoveries. Critically, this process must have two critical characteristics. First, it needs to be highly interactive. For example, after we visually examine raw data, we may go to the next stage of event-based exploration to study the moments of interest captured by a certain event variable. At that stage, we may need to go back to an overview stage to verify the patterns we spot in event-based exploration and interpret these patterns in the context of global data. Similarly, when we group and compare patterns at the final step, statistical results derived may not best capture certain properties of the gaze data. Hence, we need to go back to raw data or event-based visualization to come up with better ways in calculating final statistics.

Second, each component (and a set of guidelines and techniques described in each component) can be used individually as a part of data analysis. For example, even if one has predetermined which statistics are appropriate beforehand, it is still beneficial to visualize and examine raw data and connect that with calculated results to better understand those results. Similarly, researchers may decide to start with derived data, such as AOI data, but not raw gaze data (x and y coordinates on the tracking plane). In this case, event-based exploration can be applied directly to derived data.

Finally, we note that general principles and ideas advocated in this article can be carried out across many forms of implementation, from MS Excel, to MATLAB and R, to specialized visualization systems. Most visualization examples in this article can be generated relatively easily using existing software. To follow the principles and spirit of human-in-the-loop data exploration, one can simply use MS Excel to generate both visualized results shown in Figures 1, 2, and 3, and temporal profiles and timing histograms shown in Figures 4 and 5. In addition, there are more and more (advanced) resources available for researchers to use. The examples in this article were derived using a visual data mining system that we have developed and used for several years (Yu, Zhong, Smith, Park, & Huang, 2009). This system is designed to analyze temporal sequences such as gaze data and will be made available to the research community. In addition, there are many other open-source projects that are targeted at building research tools for scientists (e.g., Cyberinfrastures developed by Börner, 2011). These techniques provide rich resources for exploratory data analysis and visual data mining for researchers. In summary, with more and more gaze data collected in various laboratories, and with more and more information visualization and data mining techniques available for researchers, we suggest that exploratory data analysis and visual data mining of gaze data are a promising and feasible direction to lead to fruitful results and advance our knowledge of cognition and development.

REFERENCES

Adar, E. (2006). GUESS: A language and interface for graph exploration. *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems*, 791–800.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Amso, D., & Johnson, S. P. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology*, *42*(6), 1236–1245.

Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, *105*, 63–80.

Ashmead, D. H., & Davis, D. L. (1996). Measuring habituation in infants: An approach using regression analysis. *Child Development*, *67*, 2677–2690.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*, 48–53.

Aslin, R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science Official Publication of the American Academy of Optometry*, *86*(6), 561–565.

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development*, *62*, 1227–1246.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*, 191–208.

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–767.

Bell, G., Hey, T., & Szalay, A. (2009). Beyond the data deluge. *Science*, *323*, 1297–1298.

Börner, K. (2011). Plug-and-play macroscopes. *Communications of the ACM*, *54*(3), 60–69.

Busey, T. A., & Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin & Review*, *17*(2), 155–160.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*, San Francisco, CA: Morgan Kaufman.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*, 337–367.

D'Entremont, B. D., Hains, S. M. J., & Muir, D. M. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, *20*, 569–572.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*, 668–670.

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborn' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, *102*, 17245–17250.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences*, *105*, 9926–9930.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than in isolation. *Developmental Science*, *9*, F33–F40.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98–116.

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010). Head-mounted eye-tracking of infants natural interactions: A new method. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, 21–27.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants attention to faces during the first year. *Cognition*, *110*(2), 160–170.

Gilmore, R. O., & Rettke, H. R. (2003). Four-month-olds' discrimination of optic flow patterns depicting different directions of observer motion. *Infancy*, *4*, 177–200.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*, 23–45.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*, 254–260.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279.

Hayhoe, M. (2004). Advances in relating eye movements to cognition. *Infancy*, *6*(2), 267–274.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3), i–vi, 1–123.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*, 341–348.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In L. P. Lipsitt (Ed.), *Advances in child development and behavior* (pp. 69–95). New York: Academic Press.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31–F39.

Jackson, I., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science*, *12*, 670–679.

Johnson, S. P., Davidow, J., Hall-Haro, C., & Frank, M. C. (2008). Development of perceptual completion originates in information acquisition. *Developmental Psychology*, *44*(5), 1214–1224.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*(2), 185–201.

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *The Journal of Neuroscience*, *29*, 6234–6238.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Kovács, A. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, *106*, 6556–6560.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research*, *41*, 3559–3565.

Mani, N., & Plunkett, K. (2010). In the mind's ear: Evidence for implicit naming in 18-month-olds. *Psychological Science*, *21*, 908–913.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*, 203–229.

Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science*, *17*, 781–787.

Pereira, A. F., James, K. H., Jones, S. S., & Smith, L. B. (2010). Early biases and developmental changes in self-generated views. *Journal of Vision*, *10*, 1–13.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Science Staff. (2011). Dealing with data. Challenges and opportunities. Introduction. *Science*, *331*(6018), 692–693.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *9*, 668–671.

Shic, F., Scassellati, B., & Chawarska, K. (2008). The incomplete fixation measure. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 111–114.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *IEEE Proceedings of Visual Languages*, 336–343.

Shneiderman, B. (2002). Inventing discovery tools: Combining information visualization with data mining. *Information Visualization*, *1*, 5–12.

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. *Trends in Cognitive Sciences*, *6*, 293–298.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568.

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17.

Spivey, M. J. (2007). *The continuity of mind.* New York: Oxford University Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Thomas, H., & Gilmore, R. O. (2004). Habituation assessment in infancy. *Psychological Methods*, *9*, 70–92.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*, 5012–5015.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180.

Yu, C., Smith, T. G., Hidaka, S., Scheutz, M., & Smith, L. B. (2010). A data-driven paradigm to understand multimodal communication in human–human and human-robot interaction. In P. R. Cohen, N. M. Adams & M. R. Berthold (Eds.), *Advances in intelligent data analysis IX, LNCS 6065* (pp. 232–244). Berlin/Heidelberg: Springer Verlag.

Yu, C., Zhong, Y., Smith, T., Park, I., & Huang, W. (2009). Visual data mining of multimedia data for social and behavioral studies. *Information Visualization*, *8*, 56–70.