Does an Eye Tracker Tell the Truth about Visualizations?: Findings while Investigating Visualizations for Decision Making

Sung-Hee Kim, Zhihua Dong, Hanjun Xian, Benjavan Upatising, and Ji Soo Yi, *IEEE Member*

(a) SimulSort (b) Typical Sorting

Fig. 1. Comparing two screenshots of the total aggregated fixation duration of 10 participants for 10 trials. The red area indicates longer duration of fixations. The two interfaces compared are (a) SimulSort, a tabular visualization with simultaneously sorted columns, and (b) Typical Sorting, a table with a one-column sorting feature.

Abstract—For information visualization researchers, eye tracking has been a useful tool to investigate research participants' underlying cognitive processes by tracking their eye movements while they interact with visual techniques. We used an eye tracker to better understand why participants with a variant of a tabular visualization called 'SimulSort' outperformed ones with a conventional table and typical one-column sorting feature (i.e., Typical Sorting). The collected eye-tracking data certainly shed light on the detailed cognitive processes of the participants; SimulSort helped with decision-making tasks by promoting efficient browsing behavior and compensatory decision-making strategies. However, more interestingly, we also found unexpected eye-tracking patterns with SimulSort. We investigated the cause of the unexpected patterns through a crowdsourcing-based study (i.e., Experiment 2), which elicited an important limitation of the eye tracking method: incapability of capturing peripheral vision. This particular result would be a caveat for other visualization researchers who plan to use an eye tracker in their studies. In addition, the method to use a testing stimulus (i.e., influential column) in Experiment 2 to verify the existence of such limitations would be useful for researchers who would like to verify their eye tracking results.

Index Terms—Visualized decision making, eye tracking, crowdsourcing, quantitative empirical study, limitations, peripheral vision.

1 Introduction

An eye tracker is a potentially useful tool for information visualization (InfoVis) researchers because its basic premise is that it can tell where a person looks. In addition, as long as the "eye-mind hypothesis" [34] holds, eye-tracking results can reveal the underlying cognitive processes of a human user. In this case, the eye is literally the window of the mind. For this particular reason, some InfoVis researchers who are interested in the cognitive aspects of a visualization user often rely on eye-tracking methods (e.g., [6, 11, 46, 32]). In addition, visualization tools have been proposed to analyze eye-tracking data (e.g., [48]).

We are also researchers who would like to see the person's mind while investigating visualization tools supporting multi-attribute decision making, where one has to choose the best option among many candidates after reviewing the multiple attributes of each candidate (e.g., choosing a college or a nursing home). Since such multi-attribute decision making often involves overwhelming information and laborious cognitive processes, various visualization techniques have been proposed (refer to [25] for reviews). Some recent empirical evidence also has demonstrated that such techniques lead to better decision quality and satisfaction [1, 35, 38, 40, 15]; however, the gap in the previous literature is that there is no empirical explanation of how these visualization techniques have helped with decision making beyond a simple confirmation of their effects. For example, studies using a visualization tool called SimulSort (or SS) [16, 15] empirically showed that the participants who used SS made higher-quality decisions in a shorter amount of time than made the participants who used a regular table with a typical single-column sorting technique: Typical Sorting (or TS); however, these empirical studies cannot clearly explain *how it happened*.

To fill this gap, in this paper, we conducted an eye-tracking study to investigate how visual aids influenced the participants' browsing behaviors and decision-making strategies that eventually influence decision quality [10, 29]. The eye-tracking study partially showed that the decision quality difference actually came from the changes in the decision strategies that the participants employed. Though this finding is only meaningful to a relatively small number of researchers who would like to combine InfoVis and decision science, such a finding is one of the first pieces of empirical evidence showing the *how* part and also one of major contributions of this paper.

Interestingly, we had another unexpected finding of potential value to a larger audience. While conducting the study (Experiment 1), we came across unexpected results: We believed that a certain part of the visualization interface was seen by participants, but the eye tracker did not capture it. To verify our suspicion, we conducted an additional crowdsourcing-based study (Experiment 2). It revealed that our suspicion was correct, and it turned out to be clear evidence of a limitation of the eye-tracking method: the incapability of capturing peripheral

- *Sung-Hee Kim, Zhihua Dong, Benjavan Upatising, and Ji Soo Yi are in the School of Industrial Engineering at Purdue University. E-mail: {kim731, dong17, benjavan, yij}@purdue.edu.*
- *Hanjun Xian is in the School of Engineering Education at Purdue University. E-mail: hxian@purdue.edu.*

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

vision. In addition, the method to use a testing stimulus for Experiment 2 (influential column for our experiment) could be instrumental to other InfoVis researchers who would like to confirm the validity of their eye-tracker results.

In summary, the contributions of this paper are as follows:

- We believe this paper is one of the first eye-tracker studies empirically showing how visual aids can promote decision quality and efficiency;
- We identified a limitation of eye-tracking methods in understanding the effects of visual aids: the incapability of capturing peripheral vision; and
- We suggested an approach to overcome the limitation: embedding testing stimuli (an influential column in our case).

2 Background

2.1 Multi-attribute Decision Making and InfoVis

2.1.1 Multi-attribute Decision Making

Multi-attribute decision making means making a preference decision over all available alternatives that are characterized by multiple attributes [17]. It can be expressed in a matrix format, where rows represent alternatives, and columns represent the attributes considered [55]. Multi-attribute decision making is intrinsically difficult because multiple attributes can conflict with each other, which often requires trade-off decisions. One also may need to consider his or her own preference regarding different attributes to assign weights to them; furthermore, different attributes usually have different units, which makes the selection task even harder since it is very difficult to come up with an easy equation to calculate the value of each alternative. Using car selection as an example, mileage is measured in miles, gas mileage in miles per gallon, and price in dollars while maker information is marked by brands. A customer often faces cars either low in mileage but high in price or low in price but also low in safety and tries to remember their makers and equipment at the same time. There is no simple algorithm that could help with the decision making; thus, multi-attribute decision making in everyday life often involves high cognitive load and eventually induces a person to make a trade-off between effort and decision quality [2, 7]. Sometimes, decision makers even weigh effort reduction higher than decision-quality maximization [21], which unfortunately lead them to make sub-optimal decisions.

2.1.2 Visualization Techniques

To alleviate such difficulties, various InfoVis techniques have been utilized and have helped with the decision-making process by presenting insights more interpretable and by lowering the cognitive load for decision makers [54]. We narrowed our interest to methods that solved the problem of representing multi-attribute data sets.

Parallel coordinates is an example of a classic approach to projecting hyper-dimensional data onto a 2-D display. The attributes are represented as parallel axes to each other, and a data point with multiple attributes is visualized as a poly line connecting each data dot on each axis [18, 51]. This technique is known to be effective in visualizing large multi-attribute data sets because it provides an overview of the data trend, which may help multi-attribute decision making.

Although parallel coordinates performs well at presenting high-dimensional data, the fact that it lacks a tabular view limits its application for helping with daily decision-making tasks. This is because the tabular form of presenting data with sorting features, such as Microsoft Excel, not only is familiar to general users but also makes each data point visible. The data points in parallel coordinates are initially hidden and require additional interactions in order to retrieve the data.

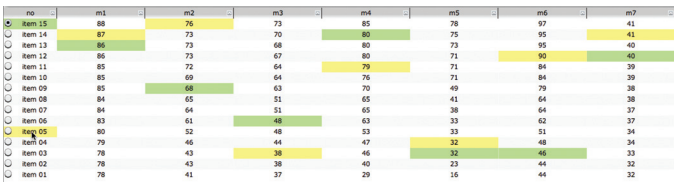
In contrast, a problem with tabular data presentation is one-column sorting; as the sorting is done for a single column, data sequences for other columns are changed accordingly. This brings difficulty for users doing comparisons among alternatives since they keep losing the context of the previous sorting results as they have to sort the columns several times separately. Other visualization tools combining the advantage of parallel coordinates and tabular form offer better solutions for multi-attribute decision making. Parallel bargrams is an example that succeeds a generic tabular form and sorts all the attributes in paralleled rows at the same time [53]. It is designed to help consumers with multi-attribute mechanized purchasing decisions by providing simultaneously comparable attribute values for different alternatives. FOCUS [45], EZChooser [53], and InfoZoom [44] are some visualization tools that apply the idea of parallel bargrams.

2.1.3 SimulSort

SimulSort (SS) is another visualization technique that also aims to help with daily multi-attribute decision making [16]. As shown in Figure 2(a), SS presents all columns sorted simultaneously so that one can see the relative values or utilities (i.e., pros and cons) of an alternative over multiple attributes. This visual representation is expected to avoid the constant shuffling of rows induced by sorting a column in the TS interface (i.e., Figure 2(b)) and to offer insights to users by presenting the trend of the data at a glance; however, since it preserves the tabular form that reveals the values, there is a limitation of the number of alternatives and attributes that could be represented on the screen without additional interaction techniques (e.g., zooming).

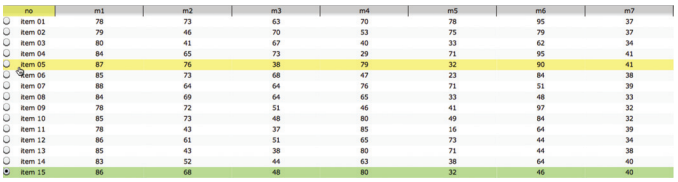
A controlled laboratory study successfully demonstrated that participants using SS made higher-quality decisions (see Section 6.5 for the definition of "decision quality") [15]; however, as discussed in the Introduction section, the controlled lab study failed to clearly show how the two representations caused the differences. A post-task survey was given to ask what kinds of strategies the participants used during the task, but the descriptions of their strategies were unfortunately vague. We felt that a more systematic investigation was necessary to better understand the influence of different visual representations on decision-making quality.

Note that the versions of SS and TS used in the present study were much simpler than the original version [16] because additional features (i.e., horizontal bargram, multiple selection, filtering, and zoom in/out) actually became compounding factors in pilot studies; therefore, they were disabled for this study.



no	m1	m2	m3	m4	m5	m6	m7
item 15	88	76	73	85	78	97	41
item 14	87	73	70	80	75	95	41
item 13	86	73	68	80	73	95	40
item 12	86	73	67	80	71	90	40
item 11	85	72	64	79	71	84	39
item 10	85	69	64	76	71	84	39
item 09	85	68	63	70	49	79	38
item 08	84	65	51	65	41	64	38
item 07	84	64	51	65	38	64	37
item 06	83	61	48	63	33	62	37
item 05	80	52	48	53	33	51	34
item 04	79	46	44	47	32	48	34
item 03	78	43	38	46	32	46	33
item 02	78	43	38	40	23	44	32
item 01	78	41	37	29	16	44	32

(a) SimulSort interface



no	m1	m2	m3	m4	m5	m6	m7
item 01	78	73	63	70	78	95	37
item 02	79	46	70	53	75	79	37
item 03	80	41	67	40	33	62	34
item 04	84	65	73	29	71	95	41
item 05	87	76	38	79	32	90	41
item 06	85	73	68	47	23	84	38
item 07	88	64	64	76	71	51	39
item 08	84	69	64	65	33	48	33
item 09	78	72	51	46	41	97	32
item 10	85	73	48	80	49	84	32
item 11	78	43	37	85	16	64	39
item 12	86	61	51	65	73	44	34
item 13	85	43	38	80	71	44	38
item 14	83	52	44	63	38	64	40
item 15	86	68	48	80	32	46	40

(b) Typical Sorting interface

Fig. 2. Example of the two interfaces, SimulSort and Typical Sorting, comparing two alternatives; item 15 is highlighted in green, and item 5 is highlighted in yellow. (a) SimulSort: The comparison of the two items can be done by comparing the vertical positions of the highlighted cells. (b) Typical Sorting: The comparison of the two items can be done by reading the face values of all of the cells.

2.2 Two Potential Explanations

Based on our observations of how the participants used SS and TS, we came up with two general explanations for the performance difference between the two representations, (1) efficient browsing behavior and (2) different decision strategies, each of which is further described in subsequent sections.

2.2.1 Efficient Information Browsing

A potential benefit of SS (also found in other InfoVis tools [12, 50]) is that users can get an overview of multiple alternatives and their associated attributes through instantly grasping visual patterns without reading detailed information. Pattern detection includes detection of data distribution, trends, outliers, or other structures of a data set [54]. This pattern detection could happen within peripheral vision, especially for an interface that includes visual presentation rather than just text because pictorial presentation can convey more useful visual information and increase the radius of what can be captured by the peripheral vision [33]. Acquiring information using peripheral vision and pattern detection can help participants to grasp information quickly and to guide their focus to more worthy information [33, 54].

Taking SS as shown in Figure 2(a) as an example, one can quickly sense that item 5 (highlighted in yellow) is better than item 15 (highlighted in green) without reading the actual values of all of the corresponding cells because the yellow cells are generally located above the green cells. In contrast, while using TS, one needs to read numbers on cells for items 5 and 15 and to compare them as shown in Figure 2(b). This could explain why the SS participants made faster decisions than the TS participants; however, it cannot fully explain why the SS participants made higher-quality decision outcomes. One might argue that the SS participants who relied on visual trends read less accurate information than the TS participants, who relied on actual numbers. SS even induces distortion between two values; for example, the green cell and the yellow cell in the m4 column in Figure 2(a) are two cells apart even though the difference in value is just one. Thus, if efficient information browsing were the only advantage of SS over TS, the decisions of the TS participants should have been more accurate than or equivalent to those of the SS participants, but our previous experiment showed the opposite results [15]; therefore, we need to find another explanation.

2.2.2 Different Decision Strategies

Another potential advantage provided by visualization tools for decision making is that these tools promote better decision-making strategies, which associated with how an individual would process given information to make a choice [31].

In order to understand different decision strategies, studies in decision science should be briefly reviewed. Over the past 30 years, several researchers from decision science and behavioral economics have researched how people actually make decisions. There have been several strategies introduced in the context of multi-attribute decision making [30, 31, 39], and they are categorized into the following two groups: *compensatory strategies* and *non-compensatory strategies*.

Compensatory Decision Strategies. If a decision maker applies compensatory strategies, it means that by considering all of the attributes, a low value of one attribute can be compensated by a high value of another attribute; therefore, the final value of an alternative is calculated based on the trade-off among all of the attributes. Compensatory strategies are known to be closer to the normative approach, which leads to higher accuracy as all of the relevant information is processed. For example, a Weighted Additive (WADD) strategy would calculate the value of an alternative by the sum of each attribute multiplied by the weight given to that attribute. An Equal Weighted (EQW) strategy is a particular case in which all of the attributes are equally weighted.

Non-compensatory Decision Strategies. In contrast, applying non-compensatory strategies means that an alternative can be eliminated from a set of candidates simply because one of its attributes has a lower value even though values in the other attributes might have high values. For example, Elimination by Aspects (EBA) begins by determining important attributes and eliminating alternatives that do not fulfill the cutoff values for the attributes [49]; therefore, not all of the attributes or alternatives get attention, which often leads to a relatively lower decision quality.

Although compensatory strategies lead to better decision outcomes, researchers found that people often deviated from compensatory strategies. An individual decision maker has a limited cognitive capacity that restrains information processing at a certain level. Because compensatory strategies are more cognitively demanding, people take mental shortcuts to reduce the amount of data to process [9, 42]. Some of these shortcuts are non-compensatory strategies, where the attributes to be considered are selectively chosen, leading people to consider fewer attributes and alternatives.

Given the two general categories of decision strategies (i.e., compensatory and non-compensatory), we suspected that the two visual aids (SS and TS) may lead to choose one or the other decision strategy. In order to test this hypothesis, we employed a measure of information search process, called "depth of search," [52]; therefore, the decision strategies could be examined according to the different amounts of information searched. Depth of search reflects the amount of information processed, which could be interpreted as the number of attributes considered. In order to select the optimal solution, one has to consider the trade-off among all of the attributes; therefore, the number of attributes considered could be an indirect metric of the likelihood of employing compensatory strategies: the more attributes considered, the more likely compensatory strategies were applied.

2.3 Process Tracing

In order to verify the two potential explanations (i.e., Efficient Information Browsing and/or Different Decision Strategies), we needed to employ process-tracing techniques:

First, verbal protocol (think-aloud) analysis was briefly considered; however, our pilot study showed that generating concurrent or retrospective verbal reports either interfered with visually taxing decision making or were not reliable, as was reported in previous literature [4].

Second, information board (or its electronic version, Mouselab) [20], which records the acquisition of information represented in a matrix form, is widely used among decision scientists. In Mouselab, the value of each cell is presented in a box that is first hidden. The value is revealed when the mouse cursor is over the cell and covered again when the cursor leaves the cell. Although it captures all of the data acquisition behavior explicitly, it does not seem to be appropriate because it allows a participant to retrieve only one value at a time, which defies the main purpose of visualization: allowing one to see the overall trends. It has also been shown that the information board method actually affects people to exert a certain strategy [23].

Thus, we ended up selecting the eye-tracking method because it allowed us to capture more natural behaviors [23]. Although it heavily relies on the eye-mind hypothesis [34], Glaholt Reingold [10] demonstrated that eye movements in decision-making processes could reflect the screening and evaluation of different alternatives. Previous work also has shown that eye tracking can identify the visual exploration behaviors in different visualizations [6, 11].

Among the various metrics employed in eye-tracking studies (e.g., gaze, fixation, and pupil dilation), we particularly employed fixation since it has been widely used to assess participants' cognitive processes [19, 26]. Fixation duration reflects the task difficulty and information complexity [28, 33]. A longer duration indicates that the task complexity and difficulty is higher. We also employed visit as another metric, which is defined as the aggregated fixations and saccades of an individual visit to an Area of Interest (AOI) [47]. Fixation and visit are slightly different, as shown in Figure 3. Fixation duration within an AOI is the sum of the duration of all of the individual fixations in the corresponding AOI, while visit duration comprises all of the fixations that occurred in one visit within the AOI as well as the saccadic duration among those fixations within that AOI until fixation is placed outside of the AOI. With eye-tracker data based on AOIs, we could examine the interaction between the participants and the attributes.

We defined AOIs by columns, as shown in Figure 4, because in the TS setting, the rows were shifting throughout the process as the participants tried to sort the different columns. With the AOIs fixed and the rows changing, it was challenging to capture gaze data within certain alternatives. Even though one could trace programmatically or manually the locations of the dynamically moving cells, detecting which cell was gazed at at any given moment was challenging due to the short

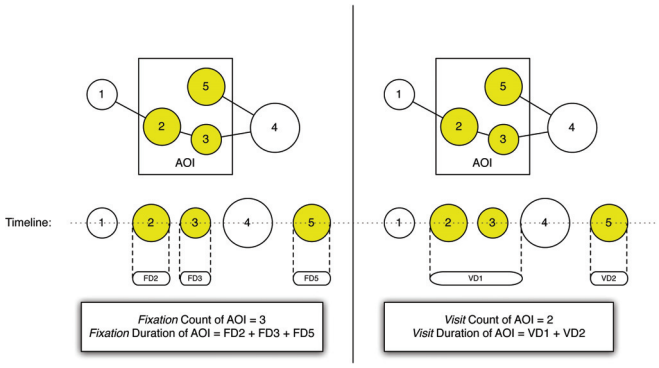
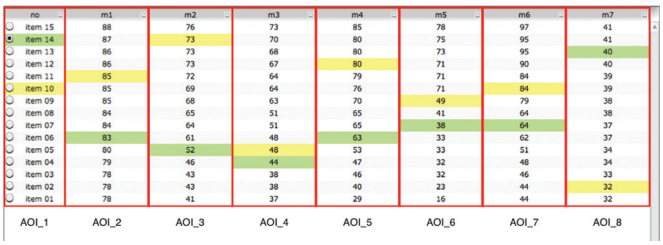


Fig. 3. Definition of fixation and visit in the context of an AOI.

height of each cell (i.e., 20 pixels or 6 mm). In order to overcome such a limitation, we considered and captured mouse hovering data on all of the cells as another process tracing data source; however, we ended up ignoring the mouse hovering data because the SS participants were forced to hover over a cell to see what the associated attributes were while the TS participants were not (they used mouse hovering voluntarily for marking purposes).



no	m1	m2	m3	m4	m5	m6	m7
item 15	88	76	73	85	78	97	41
item 14	87	73	70	80	75	95	41
item 13	86	73	68	80	73	95	40
item 12	86	73	67	80	71	90	40
item 11	85	72	64	79	71	84	39
item 10	85	69	64	76	71	84	39
item 09	85	68	63	70	49	79	38
item 08	84	65	51	65	41	64	38
item 07	84	64	51	65	38	64	37
item 06	83	61	48	63	33	62	37
item 05	80	52	48	53	33	51	34
item 04	79	46	44	47	32	48	34
item 03	78	43	38	46	32	46	33
item 02	78	43	38	40	23	44	32
item 01	78	41	37	29	16	44	32
AOI_1	AOI_2	AOI_3	AOI_4	AOI_5	AOI_6	AOI_7	AOI_8

Fig. 4. A screenshot of SS, where all of the columns are sorted simultaneously. Each highlighted color (i.e., green and yellow) corresponds to one item (i.e., items 14 and 10, respectively). The eight red boxes show how the AOIs are defined.

3 Hypotheses

Based on the above discussion of the two potential explanations for why SS has performed better than TS, we investigated browsing behavior and decision strategies, respectively, using eye-movement data. If the participants applied peripheral vision to gain the trend information, then we expected to see shorter fixation duration and higher fixation counts as trend reading requires a lower cognitive load than reading and comparing numbers and promotes more scanning through data. If one applied the optimal compensatory strategy, then he or she should have evaluated each alternative, considering all of the attributes. We hypothesized that SS would promote the compensatory strategy with higher depth of search and less effort. Having adopted this compensatory strategy, the visit count should have been higher and uniformly distributed among the attributes.

H1 Efficient browsing behavior would appear more in SS than in TS.

- H1a Fixation durations in SS would be shorter than those in TS.
- H1b Fixation counts in SS would be higher than those in TS.

H2 Compensatory strategies would be promoted more in SS than in TS.

- H2a Visit count would be higher in SS than in TS.
- H2b Visit count would be more uniformly distributed in SS than in TS.

4 Experiment 1: Eye-tracking Study

The goal of the eye-tracking study was to understand the underlying differences in cognitive processes while using different interfaces to make decisions. We replicated a controlled laboratory experiment done by Hur et al. [15] with the two interfaces: SS and TS. We also chose the same between-subject design to minimize the influence of extraneous factors. Within-subject design would have the risk of carryover effects, where the first interface could influence the strategies used with the following interface.

4.1 Participants

A total of 20 participants (11 females and 9 males) were recruited from undergraduate and graduate students at Purdue University. All of the participants had not participated in the previous studies [15, 16]. Seventeen of them were from engineering, two from science, and one from finance. Their average age was 24.2 years, ranging from 18 to 30 years. The participants earned $13.16 on average, depending on their performances. This will be explained more in Section 4.6.

4.2 Apparatus

An eye tracker, Tobii X60, was used to track the participants' eye movements. Eye gazes on the screen were recorded at a sampling rate of 60Hz. The participants sat in front of a 19-inch computer screen at an approximate distance of 65 cm.

4.3 Procedure

Upon arriving, the participants were given instructions on how to use the interface and the goal of the given task. After instruction, the participants were asked to perform a nine-dot calibration with the eye tracker. Then, the participants completed a total of 15 rounds of tasks with the eye tracker monitoring their eye movements. To minimize the learning effect shown in the previous laboratory experiment, the first five rounds were regarded as practice and were not included in the analysis; however, in order to promote serious participation, the participants were not told that the first five rounds were for practice. We stopped the participants after every five rounds to let them take a rest and to recalibrate the eye tracker. This was to avoid the participants' fatigue after the intense experiment sessions and to ensure the quality of the eye-tracker data. An open-ended interview was conducted after the participants finished all of the 15 rounds of tasks to better understand the strategies they used to complete the tasks as well as their opinions toward the assigned interface (i.e., TS or SS). At the end of the experiment, the participants filled out a simple demographic survey regarding their age, gender, and education information.

4.4 Tasks

For each round, the participant was asked to select the alternative with the highest utility out of 15 alternatives (i.e., rows) after considering seven attributes (i.e., columns). The term *utility* used here is borrowed from economics and decision science, where utility is a measure of satisfaction and is different from the face value of a cell. For example, the same face value of 40 means completely different things depending on which attribute it belongs to; 40 miles per hour is relatively high fuel efficiency (i.e., high utility), but 40 horsepower is relatively low horsepower (i.e., low utility) according to the standards of modern vehicles. To mimic such a reality, the attribute-wise utility of a face value of a cell was normalized within that attribute between the maximum and the minimum values of the attribute. The utility of an alternative was the summation of the attribute-wise utilities for the seven attributes, as shown in Equation 1:

$$\text{utility}_i = \sum_{j=0}^{7} \text{utility}_{ij} = \sum_{j=0}^{7} \frac{T_{ij} - \min T_{.j}}{\max T_{.j} - \min T_{.j}}, \qquad (1)$$

where utility_i is the ith alternative's utility, utility_{ij} is the jth attribute-wise utility of the ith alternative, and T_{ij} is the face value in the jth attribute of the ith alternative in data set T. Due to the normalization, the range of the attribute-wise utility was $[0, 1]$, and the range of utility

was theoretically $[0,7]$; however, the maximum utilities of the alternatives of each round were different from each other and generally lower than 7.0 because the data sets were randomly generated in a way that there was no obvious best alternative (more details in Section 4.5). Although this calculation sounds intimidating, it turned out to be a realistic task; the participants needed to select an alternative that had as high an attribute-wise value as possible.

We also removed any contextual information from the table because it could bring in the participants' biases in considering a certain column to be more important than another (e.g., fuel efficiency could be more important than price), which would make it difficult to compare different individuals' data. This kind of context-free experimental task has been widely used in other decision science and economic studies (e.g., [41]). Each task had a time limitation of 3 minutes.

4.5 Data Sets

The same 15 data sets used in the previous study [15] were used for this study. Each data set had 15 alternatives (i.e., rows) and seven attributes (i.e., columns) with a two-digit numerical value from 10 to 99. The difficulty of each round was restricted to be similar by controlling the Average of Inter-attribute Correlations (AIAC) value [24]. The AIAC was obtained by computing the average of the correlations among all of the combinations of the two columns out of all of the columns. When the AIAC is lower, selecting the best alternative is more difficult because it leads to considering more trade-off due to negative correlations between columns. In contrast, when the AIAC is higher (maximum = 1), selecting the best alternative is easier because there are fewer trade-off situations, and the best option becomes obvious [8, 13]; thus, data sets with high AIAC would make the decision-making task over-easy, which would bias our study by making non-compensatory strategies sufficient for solving all of the problems. Therefore, in our study, the AIAC value was controlled around 0.01 to generate the appropriate level of difficulty [15]. Because of this, if one did not consider all of the columns, it would decrease the overall decision quality.

4.6 Rewards

The participants' earnings were proportional to the utilities of the final alternatives selected, and the theoretical maximum earning for each round was $7.00. In reality, the participant was paid based on the utility of two randomly selected rounds out of the 15 rounds. This quota scheme payment was known to motivate the participants by increasing the perceived benefit for each round [3].

4.7 Measurements

In order to examine the depth of search while using the two interfaces, we chose to see how the participants interacted with the alternatives and the attributes. The eye-tracker data were used to capture their interactions with different attributes. Since we defined each column as a separate AOI (i.e., each red box, as shown in Figure 4), the analysis of the fixation and visit data based on the AOI could be used to examine how the participants perceived the attributes.

5 Results and Discussion

5.1 Fixation Duration and Count

A fixation occurs when the foveal attention is focused on a particular object. Fixation duration and fixation frequency are important metrics for revealing the cognitive load of users and the perceived importance of interface elements. We looked at the fixation count and fixation duration that occurred in the AOIs to examine the visual interaction patterns during the tasks.

We employed a repeated measures ANOVA test with a within-subject factor (i.e., AOIs) and a between-subject factor (i.e., visualizations: SS vs. TS), and the results are depicted in Figure 5. First, the mean of fixation duration for SS was 0.36 seconds and 0.43 seconds for TS. The fixation duration for SS was significantly shorter than that for TS $(F(1,18) = 5.28, p = 0.0338)$. Second, the mean average fixation count for SS was 47.1 and 36.1 for TS. The fixation count for SS was significantly higher than that for TS $(F(1,18) = 6.17, p = 0.0231)$. The significant difference of fixation duration and fixation

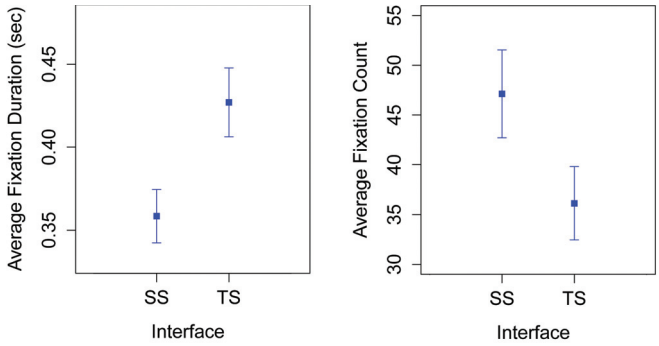


Fig. 5. Interval plot of average fixation duration (left). Interval plot of fixation count (right).

count between SS and TS suggests that the participants had quite different reading patterns for these two interfaces. With lower fixation duration (i.e., H1a confirmed) and higher fixation count (i.e., H1b confirmed), we can assume that while using SS, the participants spent less time retrieving information from each data point but visited more data points throughout the process.

Decision scientists suggest that decision processes can be decomposed into a sequence of events, such as reading the values of two alternatives of an attribute, comparing the two values, calculating the difference, adding the values for the attributes, and so forth [21]. During the decision-making process, the participants went through a series of such sub-tasks that had different cognitive loads. We assumed that the two interfaces associated with these sub-tasks comprised different patterns. Combining the participants' interview data, we found that the SS participants gained maximum information by scanning the relative positions of the green and yellow cells within one column and by reading the actual numbers when they wanted to acknowledge the numerical difference; therefore, the visual bars helped the participants in reading the values of the two alternatives and in comparing them by the participants just looking at the visual difference. The TS participants were likely to read and to remember the actual numbers when they compared attribute values between two alternatives. As people have limited cognitive capacity, the visualization could have helped unburden some subtasks, which would leave some cognition room for processing more information. In the task in which comparison was important, browsing behavior occurred more for SS users, which could be the reason for shorter fixation duration. Eventually, the participants could consider more attributes and alternatives with the SS interface.

5.2 Heatmaps

The overall patterns of eye movements in the SS and TS conditions are shown in Figure 1. The total fixation duration for each interface was aggregated for all of the tasks completed by 10 participants. Figure 1 shows that there were distinctive differences in the eye movement patterns between the two interfaces. In TS, the fixations had longer duration in the top areas of the columns, which might indicate that the participants focused on reading the maximum value of a sorted column. In addition, the total duration of fixations looked stronger in the left columns (i.e., in AOI_2, 3, 4, and 5) than in the right columns (i.e., in AOI_6, 7, and 8), suggesting the participants might have spent more time comparing values within the columns on the left than within those on the right. This could be due to the nature of most of the participants' reading patterns in which they started from the left and continued to right, which also has been shown commonly in reading web content [43]. Additionally, according to satisficing theory [5], an individual tends to make decisions based on information that is immediately perceptible; the options listed first in a sequence have a higher probability of being selected, while the last options listed may not even be observed. In contrast, with SS, the total duration of the fixations in the top areas of the columns was weaker, and, interestingly, the center columns (i.e., AOI_4, 5, and 6) received more fixations than the

other columns even though we expected the seven columns to be more evenly browsed. To better understand the implications of the differences, we analyzed the visit count data from the eye tracker.

5.3 Visit Count

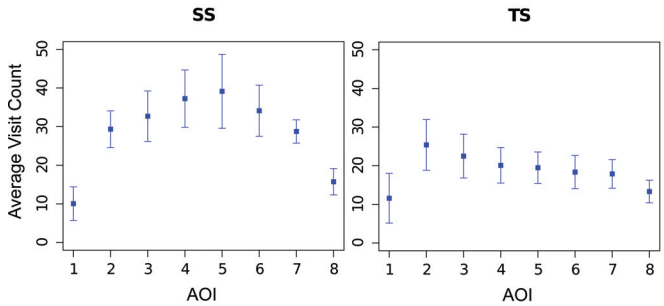


Fig. 6. Interval plot of average visit count for the two interfaces.

Since a visit was defined as one observation within an AOI, and an AOI was defined by a column (Figure 4), visit count was a good metric to examine how users switched their fixations among different attributes. As stated in H2, we expected to see higher visit counts (H2a) with more evenly distributed patterns in SS than in TS (H2b). When we applied repeated measures ANOVA to visit count, both interface type ($F(1,18) = 15.50, p = 0.0010$) and AOI ($F(7,126) = 27.78, p < 0.0001$) came out to be significant main effects. The mean of visit count (for one participant and one task round) for the SS interface was statistically significantly higher with 28.3 for SS and 18.5 for TS. As in Figure 6, we saw that SS had a higher visit count for the seven attributes compared to TS (i.e., H2a was confirmed).

However, the center-column total fixation duration for SS, shown in both Figures 1 and 6, was an unexpected result that disconfirmed H2b. This could be because the SS participants considered the middle columns to be more important and overlooked the columns on the sides, or they could have used their peripheral vision to look at the overall trend while generally fixating on the center of the screen. Simply based on the eye-tracker data, which only tracked the gaze points, we could not determine which explanation was true; thus, we conducted a second experiment to further investigate these possibilities.

6 Experiment 2: Crowdsourcing-based Study

This experiment was conducted to supplement the results from Experiment 1 discussed in Section 5.3, where we could not clearly see whether the SS interface promoted consideration of more attributes. In order to triangulate our confirmation of H2a and to further test whether SS promoted consideration of attributes in a more uniformed way (H2b), the data set was designed to capture the behavior of overlooking certain attributes in certain column positions.

We conducted the experiment through Amazon Mechanical Turk (MTurk), a well-known crowdsourcing platform. The crowdsourcing approach has several advantages over conventional, controlled laboratory studies [22, 37], including recruiting a large number of participants with diverse backgrounds in a more natural environment.

6.1 Participants

6.1.1 Demographic Summary

A total of 176 participants were originally recruited through MTurk, but 57 were identified as outliers and were excluded from the analysis (see Section 6.1.2 for details). The remaining, legitimate participants comprised 58 for the SS interface and 61 for the TS interface. Due to the analysis, we also removed additional data from 19 participants (i.e., 8 from SS and 11 from TS) to conveniently end up with a balanced data set (i.e., 50 for each group).

The remaining 100 participants consisted of 48 females and 52 males with a self-reported age range of 18 to 56, an average age of 28.6 years. They were evenly and randomly assigned to the two conditions (TS and SS), and none of them participated in both conditions. The education levels of the participants were as follows: 4-year college, 34.5%; Master's degree, 31%; and 2-year college, 18%. Their majors were computer and information systems, 25.5%; engineering, 20.5%; and science and math, 13%; followed by business, 12%. The baseline payment for participation was $0.10. An additional bonus reward was a maximum of $0.30 from two randomly selected rounds, as in Experiment 1. The participants earned $0.23 on average, depending on their performance.

6.1.2 Outliers

Previous crowdsourcing studies have shown that there were workers who completed the tasks without paying reasonable attention (e.g., by selecting random responses as quickly as possible to merely earn compensation) [14, 27]. To prevent such outliers from contaminating the present experiment's data, we excluded the participants who responded inconsistently (or randomly, in extreme cases) over the multiple trials. We employed Pearson's χ^2 test to see how close the distribution of the rank of their responses was to the uniform distribution. We assumed that if the p-value of the Pearson's χ^2 test was bigger than 0.02, then the participant's performance was sufficiently close to the uniform (in other words, random) distribution. The threshold of 0.02 was determined according to the data obtained from a controlled lab study.[1] As a result, 57 participants were identified as outliers and were excluded from further analysis. Figure 7 shows the differences in the average time spent on finishing a single task between the two groups; legitimate participants spent an average of 51.3 seconds, while outliers spent an average of 8.3 seconds, a duration too short for finding the best answer.

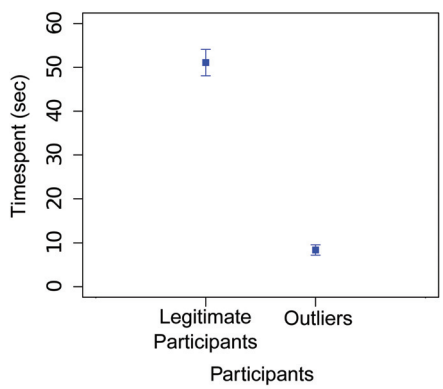


Fig. 7. The mean of time spent for each trial, comparing the participants between nonrandom clickers and random clickers.

6.2 Procedure

Our experiment was posted on the MTurk platform. After the participants read the instructions on the MTurk website, they were redirected to our experimental website, which was separately hosted. Each participant was asked to complete 15 trials, where the first six trials were for practice, and the following nine trials were for actual analysis.

6.3 Experiment Design

With the aim of testing whether SS promoted strategies considering all of the attributes evenly rather than focusing on central columns, a testing stimulus (i.e., influential column) was devised for this experiment. The idea behind the influential column was to present a certain column so skewed in its value that the alternative with the highest utility would become too obvious if one paid attention to the column; however, if one failed to pay attention to it, the person would end up selecting a suboptimal alternative. More details on how the influential column was constructed are discussed in Section 6.4. This column

[1] Detailed investigation of this issue will be published in a separate venue.

was placed in three different positions (i.e., A, leftmost; B, center; and C, rightmost), as shown in Figure 8; thus, by comparing the decision qualities when using SS and TS under these three position settings, we could see whether SS or TS helped to avoid skewed consideration of attributes. A total of nine trials (three for each position: A, B, and C) were permuted to avoid ordering effects.

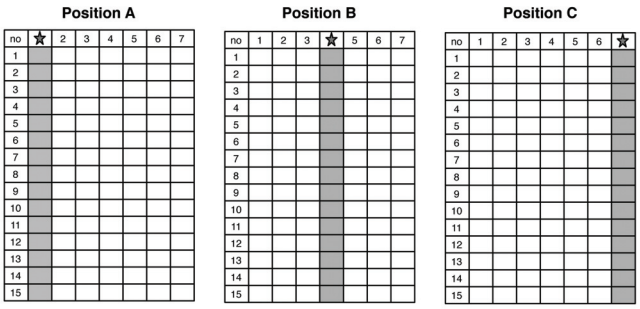


Fig. 8. Sample tables comprising three variations in the position of the influential column among the seven attributes: position A at leftmost, position B at center, and position C at rightmost.

6.4 Data Sets

The data sets used in the first six practice trials were selected from the previous experiment, Experiment 1 (Section 4.5). For the actual data sets used in the analysis, nine data sets were generated to guarantee that no duplicate data sets were presented to a participant. Each data set had two-digit integers ranging from 10 to 99 that were presented as 15 alternatives (i.e., rows) with seven attributes (i.e., columns), the same as the data sets used in Experiment 1. For each data set, we used the following strategies to generate the non-influential columns and the influential column. Figure 9 shows one of the data sets we used in the experiment. First, for the non-influential columns, we maintained the same task difficulties across the data set by controlling the AIAC [24]. The AIAC for the six columns was controlled to be 0 ± 0.001. Second, for the influential column, there was a 70% drop from the highest number to the second highest number; however, the gap between the two adjacent values (sorted by values) was kept at 2.5%.

Item	Attr 1	Attr 2	Attr 3	Attr 4	Attr 5	Attr 6
1	35	83	16	47	33	20
2	67	91	79	47	17	18
3	61	80	92	43	33	12
4	39	85	12	45	35	24
5	56	90	36	42	30	24
6	54	80	48	46	24	21
7	54	81	17	41	14	16
8	46	88	16	46	34	24
9	64	85	30	42	14	15
10	58	80	16	44	14	22
11	35	90	74	42	15	22
12	65	87	71	47	17	23
13	38	85	79	47	33	16
14	45	84	89	41	32	22
15	64	80	32	47	31	21

(a)

Attr
80
35
34
33
32
30
29
28
26
25
24
22
21
20
19

(b)

Fig. 9. An example of (a) six non-influential columns with an AIAC = 0.00027 and (b) an influential column.

After both the non-influential columns and the influential column were generated, we merged the two matrices to form a 15×7 table according to the following steps:

1. We computed the overall utility value for each row of the six non-influential columns;
2. We found the lowest possible utility value that could be raised to the best alternative after adding 0.7 to the overall utility value; and
3. We assigned the highest value in the influential column to this row as well as randomized the assignments of the other values in the influential column to the other rows.

This strategy aimed to reduce the possibility of the best alternative being selected by the participants if they overlooked the influential column.

6.5 Measurement

As the data sets were generated to be different, the highest valued item in each set changed for each data set; therefore, we followed the measure decision quality used in the previous study [15], calculated as follows:

$$Decision\ quality = \frac{utility_i - \min(utility.)}{\max(utility.) - \min(utility.)} \tag{2}$$

where $utility_i$ is derived from Equation 1. This normalized the performance from 0 to 1, where 1 was the maximum decision quality when the highest item was selected. The reason to introduce yet another metric over utility was that the maximum utility per each round was different from each other because each data set was randomly generated.

Other than introducing the influential column and using the crowdsourcing approach, the remaining experiment design stayed the same.

7 Results and Discussion

7.1 Decision Quality

We employed a repeated measures ANOVA test on the decision quality of the interface type, which had a significant main effect ($F(1,98) = 5.60, p = 0.0199$). On average, the decision quality with the SS interface (0.91) was higher than with the TS interface (0.88), as shown in Figure 10. This finding is consistent with an earlier controlled laboratory study [15].

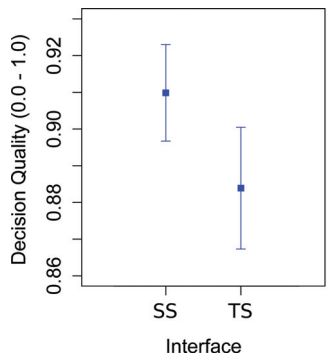


Fig. 10. The decision quality with the SS interface and the TS interface.

As our experiment was designed to penalize decision makers who overlooked the influential column, we could see whether more columns were likely to be considered in SS, which could have been due to the employment of compensatory strategies.

7.2 Column Order Effect

To see whether the three influential columns' positions affected the performances with the different interfaces, we analyzed the data separately for SS and for TS. According to the data collected from SS, the repeated measures ANOVA did not show any main effect from influential column position ($F(2,98) = 0.82, p = 0.4455$); however, according to the data collected from TS, the column position had significant effects on the decision quality ($F(2,98) = 8.81, p = 0.0003$). This is also shown in Figure 8, where, for TS, the decision quality dropped

when the influential column was positioned toward the right side as in positions B and C; for SS, there was no significant difference among the different position settings.

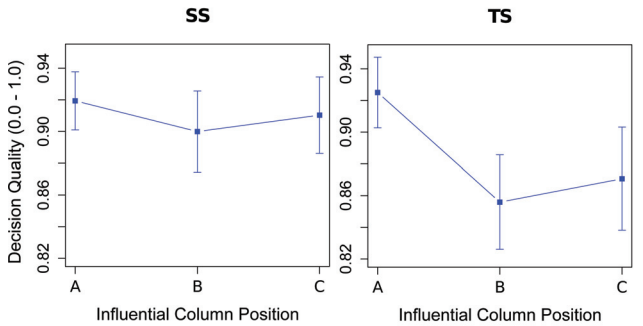


Fig. 11. The decision quality for SS and TS with the three different influential column positions.

With the different effects that the column order had on the interfaces, we could see that the interfaces promoted different decision-making strategies. For SS, as there was no difference among the influential column positions, we saw that the participants considered all of the columns without paying differing degrees of attention (i.e., H2b was confirmed). Our hypothesis H2a, that SS would promote the participants to employ compensatory strategies by considering more attributes, was also indirectly explained. The strategies collected from the participants during the crowdsourcing study also confirmed that they used one of the compensatory strategies, the Majority of Confirming Dimension (MCD), by doing a linear search among all of the alternatives:

> [I] simply looked for the most top highlighted boxes. I selected the first selection (item) and visually compared it to [the] following selections until I found one that appeared to rank higher. Then, [I] selected [a] new selection (item) and repeated [this] until I finished the list.
>
> I would select one option first and then compare the others against the selected one, one by one, down the list. At any one point in time, I would just be comparing two options.

Moreover, as mentioned in the previous literature, this compensatory strategy should be highly demanding on information processing [31]; however, the cognitive load partially inferred from the fixation in Section 5.1 showed that the participants using SS had less cognitive pressure. Eventually, the visualization helped to promote optimal strategies for achieving higher accuracy with less effort.

For TS, the results also corresponded with those of the heatmap in Figure 1. The participants considered the leftmost column more than the other columns. We believe this resulted not because they considered the left columns to be more important; rather, they started from the left due to their reading pattern [43]. While reading or comparing values toward the right, the information processing demands got higher, and the focus dropped drastically; therefore, a participant could have chosen less effort over accuracy, which would have led to lower decision quality [21]. This was also shown when participants selectively chose a few attributes to consider during the process.

The quotations collected from the participants using the TS setting also confirmed this strategy:

> I opted for the second or third highest number in the first two columns.
>
> It was helpful to look for columns with large variance and to try to narrow it down to two to three candidates.

With this crowdsourcing experiment, we believe that the longer duration of fixations shown in the middle columns in the heatmap (Figure 1) and the visit count in Section 5.3 did not result due to the fact that the participants considered the center columns to be more important; rather, it was because the eye tracker could not capture clearly the visual scans of their peripheral vision.

8 Conclusions

We conducted an eye-tracking experiment and a crowdsourcing-based experiment to explore how a visualization technique, SS, could help people in the context of multi-attribute decision making. In the eye-tracker experiment, we analyzed fixation and visit data based on columns. We found that SS and TS resulted in vastly different fixation patterns (Figure 1); moreover, SS rendered higher visit counts, less fixation duration, and higher fixation counts. With these results, we confirmed our hypothesis that SS would help with decision making by providing pattern information for quicker browsing and promoting compensatory decision-making strategies.

However, the longer duration of the fixations in the central columns of SS were unexpected results, and the eye-tracker data did not render a clearer explanation; therefore, Experiment 2, which used the influential column as a testing stimulus, was conducted with a crowdsourcing platform to explore the question. With the influential column placed in three different positions, decision quality was consistently higher in SS, while in TS we observed a drastic drop in decision quality when the influential column was placed in the middle or on the right side. This result further confirmed our finding from the eye-tracker experiment; the participants using SS had a higher depth of search.

Through the study, we verified the theory brought up by Rosen and Rosenkoetter [36] that different stimulus configurations (in our study, different interfaces), rather than the tasks themselves, could affect information-processing strategies when people made choices. More specifically, we demonstrated that SS, as a visualization tool, could help people with multi-attribute decision making by changing their behavior, which enhanced decision accuracy with lowered effort. Through exploring the use of process tracing techniques to study these behavior changes, we saw that quantitative methods could translate the originally illusive concepts into a concrete understanding.

We also showed that the eye-mind hypothesis did not hold in the SS condition probably because SS promoted more information browsing behavior using peripheral vision. We do not believe that this evidence completely nullifies the utility of eye-tracking methods in InfoVis studies; instead, this could be a warning to other researchers who use an eye-tracker. They should be cautious while analyzing collected eye-movement data. Particularly, when an average fixation duration is below 400 milliseconds (which might indicate cursory browsing behaviors, based on our results), the experimenter should make sure whether the eye-mind hypothesis holds. We tested the eye-mind hypothesis using purposefully manipulated data (i.e., the influencing column), which could be applicable to other experimental settings.

We hope that this study will serve as an example showing that eye-movement data could be used to trace cognitive procedures if used carefully. We hope that other researchers join this interesting research direction to evolve the research methods.

9 Future Work

In order to clearly understand how visual aids could effectively promote higher decision quality and efficiency, we had to control several factors. This led to unrealistic settings, such as a primitive visual interface design and small artificial data sets. This was inevitable in eliminating compound factors. Our next step would be to extend the study to a more realistic setting with large-scale and real-world data sets. Different decision strategies could be applied and influence decision quality. We could also compare this performance with existing InfoVis tools, such as parallel coordinates. Another direction could be analyzing the sequential eye-movement data. The measures we used (i.e., fixation and visit) were an aggregation of eye-movement data, which did not reveal sequential patterns. Scanpath data could help detect distinct sequential patterns for different decision strategies. This

would help in understanding more deeply why the visual aids were efficiently helping the decision maker to produce better decisions.

ACKNOWLEDGMENTS

We appreciate Zhicheng Liu for his insightful feedback and suggestions, and we also thank the students and crowdsourcing workers who participated in this study.

REFERENCES

[1] J. Bautista and G. Carenini. An empirical evaluation of interactive visualizations for preferential choice. In *Proceedings of the working conference on Advanced visual interfaces*, pages 207–214, Napoli, Italy, 2008. ACM.

[2] J. R. Bettman, E. J. Johnson, and J. W. Payne. A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, 45(1):111–139, Feb. 1990.

[3] S. E. Bonner, R. Hastie, G. Sprinkle, and S. M. Young. A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, pages 19–64, Jan. 2000.

[4] D. M. Boush and B. Loken. A Process-Tracing study of brand extension evaluation. *Journal of Marketing Research*, 28(1):16–28, Feb. 1991.

[5] T. Brinck, D. Gergle, and S. Wood. *Designing Web Sites That Work: Usability for the Web*. Morgan Kaufmann, 2002.

[6] M. Burch, N. Konevtsova, J. Heinrich, M. Hoeferlin, and D. Weiskopf. Evaluation of traditional, orthogonal, and radial tree diagrams by an eye tracking study. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2440–2448, 2011.

[7] E. H. Creyer, J. R. Bettman, and J. W. Payne. The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making*, 3:1–16, Jan. 1990.

[8] B. Fasolo, G. H. McClelland, and P. M. Todd. Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory*, 7(1):13–26, Mar. 2007.

[9] G. Gigerenzer and W. Gaissmaier. Heuristic decision making. *Annual Review of Psychology*, 62(1):451–482, Jan. 2011.

[10] M. G. Glaholt and E. M. Reingold. Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, 4(2):125–146, May 2011.

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

[12] V. Gonzalez and A. Kobsa. Benefits of information visualization systems for administrative data analysts. In *Proceedings of the Seventh International Conference on Information Visualization*, IV '03, pages 331–336. IEEE Computer Society, 2003.

[13] C. P. Haugtvedt, K. A. Machleit, and R. Yalch. *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in The Virtual World*. Psychology Press, Jan. 2005.

[14] J. J. Horton and L. B. Chilton. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce*, EC '10, pages 209–218, New York, NY, USA, 2010. ACM.

[15] I. Hur, S.-H. Kim, A. Samak, and J. S. Yi. A comparative study of three sorting techniques in performing cognitive tasks on a tabular representation. *International Journal of Human-Computer Interaction*, 2012. doi:10.1080/10447318.2012.713802.

[16] I. Hur and J. S. Yi. SimulSort: multivariate data exploration through an enhanced sorting technique. In J. Jacko, editor, *Human-Computer Interaction. Novel Interaction Methods and Techniques*, volume 5611 of *Lecture Notes in Computer Science*, pages 684–693. Springer Berlin / Heidelberg, 2009.

[17] C. Hwang and K. Yoon. *Multiple Attribute Decision Making: Methods And Applications: A State-of-the-art Survey*. Springer-Verlag, 1981.

[18] A. Inselberg and B. Dimsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization '90*, VIS '90, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

[19] R. J. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3):4, 2003.

[20] E. Johnson, J. Payne, J. Bettman, and D. Schkade. Monitoring information processing and decisions: The mouselab system. Technical report, DTIC Document, 1989.

[21] E. J. Johnson and J. W. Payne. Effort and accuracy in choice. *Management Science*, 31(4):395–414, Apr. 1985.

[22] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 453–456, Florence, Italy, 2008. ACM.

[23] G. Lohse and E. Johnson. A comparison of two process tracing methods for choice tasks. In *System Sciences, 1996., Proceedings of the Twenty-Ninth Hawaii International Conference on,*, volume 4, pages 86–97, 1996.

[24] N. Lurie. Decision making in information-rich environments: The role of information structure. *Journal of Consumer Research*, 30(4):473–486, 2004.

[25] N. H. Lurie and C. H. Mason. Visual representation: Implications for decision making. *Journal of Marketing*, 71(1):160–177, 2007.

[26] S. P. Marshall. Identifying cognitive state from eye metrics. *Aviation, space, and environmental medicine*, 78(Supplement 1):B165–B175, 2007.

[27] W. Mason and S. Suri. Conducting behavioral research on amazon's mechanical turk. *SSRN eLibrary*, pages 1–23, Oct. 2010.

[28] B. Pan, H. Hembrooke, G. Gay, L. Granka, M. Feusner, and J. Newman. The determinants of web page viewing behavior: An eye-tracking study. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 147–154, 2004.

[29] A. L. Patalano, B. J. Juhasz, and J. Dicke. The relationship between indecisiveness and eye movement patterns in a decision making informational search task. *Journal of Behavioral Decision Making*, pages 353–368, 2009.

[30] J. Payne. Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational behavior and human performance*, 16(2):366–387, 1976.

[31] J. Payne, J. Bettman, and E. Johnson. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):534–552, 1988.

[32] D. Peebles and P. C. Cheng. Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(1):28–46, Jan. 2003.

[33] K. Rayner. Eye movements in reading and information processing. *Psychological Bulletin*, 85(3):618–660, 1978.

[34] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[35] N. Reisen, U. Hoffrage, and F. W. Mast. Identifying decision strategies in a consumer choice situation. *Judgment and decision making*, 3(8):641–658, 2008.

[36] L. Rosen and P. Rosenkoetter. An eye fixation analysis of choice and judgment with multiattribute stimuli. *Memory & Cognition*, 4(6):747–752, Nov. 1976.

[37] J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: Shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 2863–2872, Atlanta, Georgia, USA, 2010. ACM.

[38] S. Rudolph, A. Savikhin, and D. Ebert. FinVis: applied visual analytics for personal financial planning. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 195–202, 2009.

[39] J. Russo and B. Dosher. Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4):676–696, 1983.

[40] A. Savikhin, R. Maciejewski, and D. Ebert. Applied visual analytics for economic decision-making. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*, pages 107–114, 2008.

[41] C. Seidl and S. Traub. Testing decision rules for multiattribute decision making. In *Current trends in economics: theory and applications: proceedings of the third international meeting of the Society for the Advancement of Economic Theory, Antalya, Turkey, June 1997*, volume 8, page 413. Springer Verlag, 1999.

[42] A. Shah and D. Oppenheimer. Heuristics made easy: An effort-reduction framework. *Psychological bulletin*, 134(2):207–222, 2008.

[43] S. Shrestha, K. Lenz, B. Chaparro, and J. Owens. "F" pattern scanning of text and images in web pages. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 51(18):1200–1204, Oct. 2007.

[44] M. Spenke and C. Beilken. InfoZoom - analysing formula one racing re-

sults with an interactive data mining and visualisation tool. In *in Ebecken, N. Data mining II*, pages 455–464, 2000.

[45] M. Spenke, C. Beilken, and T. Berlage. FOCUS: the interactive table for product comparison and selection. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, UIST '96, pages 41–50, New York, NY, USA, 1996. ACM.

[46] M. Steinberger, M. Waldner, M. Streit, A. Lex, and D. Schmalstieg. Context-Preserving visual links. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2249–2258, Dec. 2011.

[47] Tobii Technology AB. Tobii studio 2.2 user manual.

[48] H. Y. Tsang, M. Tory, and C. Swindells. eSeeTrack - visualizing sequential fixation patterns. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):953–962, Dec. 2010.

[49] A. Tversky. Elimination by aspects: A theory of choice. *Psychological review*, 79(4):281–299, 1972.

[50] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, Apr. 2004.

[51] E. J. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990.

[52] M. Westenberg and P. Koele. Multi-attribute evaluation processes: Methodological and conceptual issues. *Acta Psychologica*, 87(2–3):65–84, 1994.

[53] K. Wittenburg, T. Lanning, M. Heinrichs, and M. Stanton. Parallel bargrams for consumer-based information exploration and choice. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, UIST '01, pages 51–60, New York, NY, USA, 2001. ACM.

[54] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: How do people gain insights using information visualization? In *Proceedings of the 2008 conference on BEyond time and errors: novel evaLuation methods for Information Visualization*, pages 1–6, Florence, Italy, 2008. ACM.

[55] K. Yoon and C. Hwang. *Multiple Attribute Decision Making: An Introduction*. SAGE, Jan. 1995.