Eurographics Conference on Visualization (EuroVis) 2014
H. Carr, P. Rheingans, and H. Schumann
(Guest Editors)

Volume 33 (2014), Number 3

Evaluating the Impact of User Characteristics and Different Layouts on an Interactive Visualization for Decision Making

C. Conati, G. Carenini, E. Hoque, B. Steichen, D. Toker
University of British Columbia, Vancouver, Canada

Abstract

There is increasing evidence that user characteristics can have a significant impact on visualization effectiveness, suggesting that visualizations could be designed to better fit each user's specific needs. Most studies to date, however, have looked at static visualizations. Studies considering interactive visualizations have only looked at a limited number of user characteristics, and consider either low-level tasks (e.g., value retrieval), or high-level tasks (in particular: discovery), but not both. This paper contributes to this line of work by looking at the impact of a large set of user characteristics on user performance with interactive visualizations, for both low and high-level tasks. We focus on interactive visualizations that support decision making, exemplified by a visualization known as Value Charts. We include in the study two versions of ValueCharts that differ in terms of layout, to ascertain whether layout mediates the impact of individual differences and could be considered as a form of personalization. Our key findings are that (i) performance with low and high-level tasks is affected by different user characteristics, and (ii) users with low visual working memory perform better with a horizontal layout. We discuss how these findings can inform the provision of personalized support to visualization processing.

Categories and Subject Descriptors (according to ACM CCS): H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

1. Introduction

As visual analytics is increasingly applied in a large number of domains, there is a pressing need to design interactive visualizations that can support users with a variety of backgrounds and abilities. This endeavor requires understanding how different backgrounds and abilities impact the processing of highly interactive visualizations that support a variety of visualization manipulation methods. By visualization manipulation methods we mean conceptual methods to solve a visualization task, as defined in [BM13] (e.g., *navigation, selection, filter, aggregate*), whereas interaction techniques are the actual interface actions that allow performing specific manipulations (e.g., *zooming, details-on-demand*). Understanding the impact of individual differences for such highly interactive systems can help create personalized visualizations, either through system-driven adaptation or by providing the user with ways to customize the visualization. ([Gra06, GF10]).

However, most previous work on the impact of user differences on visualization processing has focused on basic visualizations that are used in a non-interactive manner (e.g., bar and radar charts), while results on more complex interactive visualizations are still limited. More specifically:

- There is no study on interactive visualizations examining the impact of a wide spectrum of user characteristics, ranging from cognitive abilities to personality traits and expertise.
- Interaction techniques in previous studies only support the conceptual methods of *navigation* and *selection* to solve a visualization task [BM13].
- Prior studies only consider either low-level tasks (e.g., value retrieval) or high-level tasks (in particular *discovery* [CMS99]), not both. Moreover, no study considers the high-level task of *decision making*.

With the long-term goal of devising interactive visualizations personalized to each user's individual need, this paper contributes to fill some of these gaps with a user study that investigates the impact of a large set of user characteristics on both low and high-level tasks that can be performed with a variety of visualization manipulation methods. We focus on visualizations that are designed to *support decision making through facilitating the analysis of multi-attribute rankings*, for two reasons. First, these visualizations for decision making are becoming increasingly popular (e.g., [CL04, GLG13]). Second, they support multiple visualization manipulation methods using a variety of interactive techniques, which allow users to execute a range of low and high level visualization tasks related to decision making. Hence, they represent a suitable test-bed for our purposes. To keep the number of study conditions manageable, the study presented in this paper targets one specific visualization, ValueCharts [CL04]. However, our results may possibly apply to other similar visualizations (e.g. [GLG13]) because they support analogous tasks via very similar visualization manipulation methods. Furthermore, ValueCharts (or visualizations based on ValueCharts) have already been used in practical settings (e.g., [WPTMS12, CCb*13]), so finding ways to improve their effectiveness for users with different needs can have a practical impact.

There are two different versions of ValueCharts: one integrates the relevant visualization components *horizontally*, the other one *vertically*. Previous ValueChart studies reported differences in performance with these two layouts, but did not explain them [BC08]. Thus, our study investigates whether an explanation may be the impact of user characteristics, which would indicate that visualization layout could be considered as a form of personalization. The user characteristics we investigate include *cognitive abilities* (perceptual speed, verbal and visual working memory), the *personality trait* known as locus of control, and several measures of *visualization* and *domain expertise*. The study research questions are as follows:

1. Do these user characteristics impact user performance in *low-level* tasks with ValueCharts?
   a) are these effects, if any, mediated by the type of low-level task (e.g. retrieve value, sort)?
2. Do these user characteristics impact user performance in a *high-level* decision-making task with Value Charts?
3. Are the effects for 1 & 2, if any, mediated by the visualization layout (e.g., horizontal vs. vertical)?

In the rest of the paper, we first review related work. Next, we describe ValueCharts, followed by the description of the user study. Then, we present the study results, followed by a discussion of how they relate to the above research questions, and of implications for personalization.

2. Related Work

2.1 Impact of Individual User Differences on InfoVis

While there is increasing interest in examining the role of user differences in InfoVis, most studies are primarily focused on static visualizations. Velez et al. [VST05] found correlations between five spatial abilities and proficiency in a visualization task involving the identification of the correct 3D object from its orthogonal projections. Conati & Maclaren [CM08] investigated the impact of a number of cognitive measures on a variety of low-level tasks with two alternative visualizations (radar graph and Multiscale Dimension Visualizer). They found that a user's perceptual speed was a predictor of which of the two visualizations would be most effective for one of the tasks. Toker et al. [TCCH12] performed a similar study that compared radar graphs with bar graphs, and also found an impact of perceptual speed on task performance. In addition, they found impacts of visual and verbal working memory on measures of user satisfaction.

Results on the impact of user differences on interactive visualizations are still quite limited. They mostly consider either cognitive abilities or personality traits, and only pertain to either low-level tasks or high-level *discovery* tasks. Moreover, the various visualizations in previous studies only support limited visualization manipulation methods and related interactive techniques. For example, Büring et al. [BGR06], Chen & Czerwinski [CC97], and Allen [All00] each examined the role of *cognitive abilities* (spatial ability [BGR06, CC97], spatial scanning [All00], perceptual speed [All00]) in visualizations that support the manipulation methods of *selection* and *navigation* through the interaction techniques *detail-on-demand* [BGR06, CC97, All00] and *zooming* [BGR06, CC97]. Results for low-level retrieval and comparison tasks [BGR06], as well as high-level discovery tasks ([CC97, All00]) showed that all three cognitive abilities had an effect on user performance. More recent work has explored the impact of *personality traits* on interactive visualizations. Locus of control has been shown to impact relative task completion times for alternative visualizations [ZCY*11]. In their experiment, each visualization allowed users to *select* and *navigate* by performing *detail-on-demand* and *collapse/expand* interactions. A similar study by Green and Fisher [GF10] confirmed the influence of personality traits, with locus of control and anxiety-based traits being most influential.

In contrast with this previous work on interactive visualizations, we investigate within the same experiment a more comprehensive range of user characteristics, namely cognitive abilities, personality traits, and expertise. Furthermore, we study the impact of these user characteristics on both low-level tasks as well as a high-level, open-ended decision making task, something that has never been done before. Performing these tasks requires a broader array of visualization manipulation methods than the ones examined in previous studies, namely *select, navigate*, *arrange* (reorder visual elements), and *change* (modifying the visual encoding) [BM13]. Lastly, while previous work compared visualizations that differ in terms of functionalities, our study examines two visualizations that only differ in terms of layout, which is a basic design choice for visualizations supporting the analysis of multi-attribute rankings.

2.2 Personalization in Decision Support and in InfoVis

Complex decisions can often be framed as preferential choices, i.e., the process of selecting the best option out of a possibly large set of alternatives based on multiple attributes (e.g., select a house to buy, identify a site for a new airport). Since preferential choice may require the user to explore and analyze a large amount of information, several interactive visualization tools have been developed to support this decision task, in-

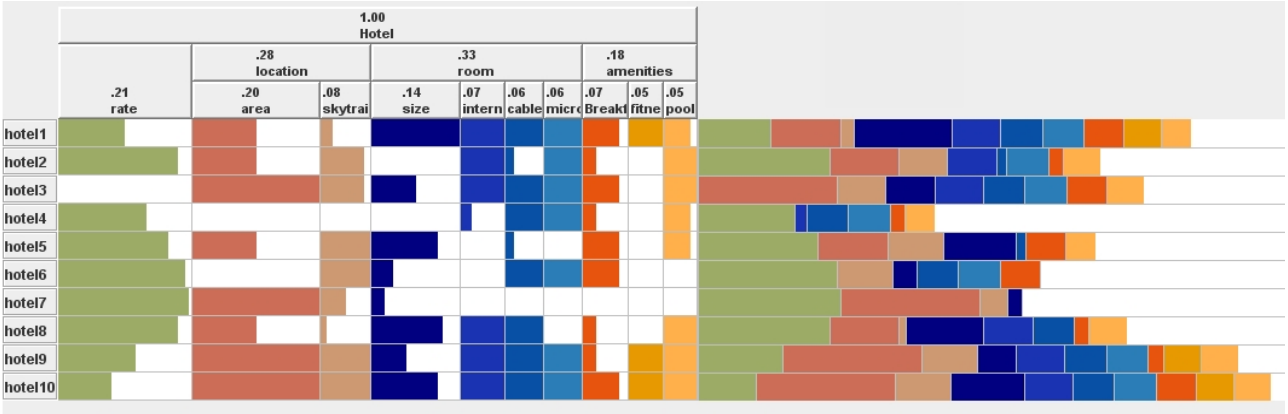

Figure 1: *ValueChart using horizontal layout (VC-H), in the sample domain of hotel selection.*

cluding the ValueCharts studied in this paper (e.g., [ATS95, AA03, Val, CL04, GLG13]).

Some researchers have moved one step further and started investigating how systems that support preferential choice can be personalized to the specific needs of each user. Some works [Lia87, SNFB95, CM06] have looked at personalization based on a user's domain knowledge, preferences, experience, and current task, but have not considered cognitive abilities or personality traits. In the vast majority of work, personalization is primarily related to *what* information should be presented to the user, as opposed to our long-term goal of understanding *how* to visualize information. The one exception to this is Liang's work [Lia87] which focused on extremely simple non-interactive visualizations.

In terms of possible personalizations in a visualization tool, [GF10, Gra06] demonstrated systems that can suggest alternative visualizations based on specific user or task features, whereas another possibility could be to improve the effectiveness of a target visualization via dynamic visual prompts. Also note that, in order to adapt to user characteristics, a system needs to acquire these characteristics as unobtrusively as possible. [SCC14] presents promising results on how this can be done from analysis of eye-tracking data in real-time.

3. ValueCharts

ValueCharts [CL04] is a set of visualizations and interactive techniques intended to support decision-makers in preferential choice; more specifically, in inspecting linear preference models created to select the best option out of a set of alternatives. Linear models are popular decision-making tools designed to help the decision-maker to *rank* the available alternatives according to multiple attributes. However, as models and their domain of application grow in complexity, the analysis of the resulting rankings becomes very challenging. Systems like ValueCharts [CL04] and LineUp [GLG13] are intended to help decision makers deal with this complexity. The effectiveness of ValueCharts has been shown in two analytic evaluations (i.e., with respect to a task model) [BC06, Yi08], and in

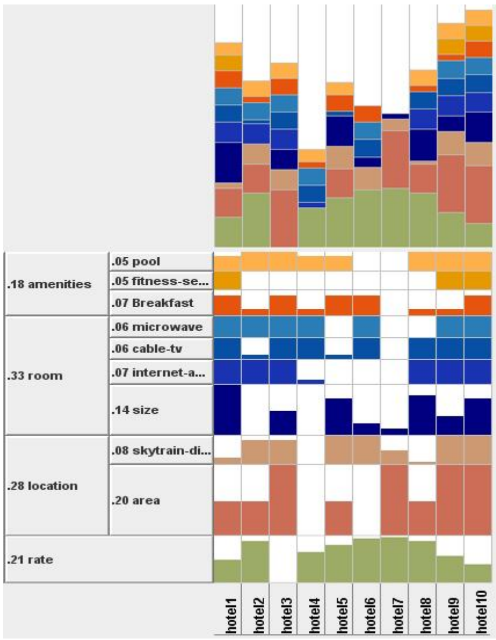

Figure 2*: ValueChart using vertical layout (VC-V).*

two user studies [BC08, PBW*12], as well as several applications (e.g., [WPTMS12]).

Based on its success, we chose ValueCharts as a suitable testbed for examining the impact of individual characteristics on an interactive visualization tool that supports multiple manipulation methods for low-level tasks as well as the high-level task of decision making/preferential choice. Figure 1 shows an example of a ValueChart for the simple preferential choice of selecting a hotel when traveling to a new city, out of ten available alternatives. For the sake of simplicity, here we just describe the key features of ValueCharts[1]. The relevant hotel attributes or objectives (e.g., area, skytrain distance, internet access, etc.) are arranged hierarchically (in the so-called *objective tree*) and are represented in the top-left quadrant of the figure, forming the columns in the ValueChart

[1]The video demo: www.cs.ubc.ca/group/iui/VALUECHARTS

display. The width of each column indicates the relative weight assigned to the corresponding objective (e.g., sky-train distance is less important than area). The available alternatives (i.e., hotels) are represented as the rows in the display. The cells in each row specify how the corresponding alternative fares with respect to each objective, indicated by the amount of filled color in the cell. For instance, hotel1 is far from the sky-train, but it has excellent internet access. In the rightmost quadrant, all values for each alternative are accumulated and presented as horizontal stacked bars, displaying the overall value of each alternative (e.g., in Figure 1, hotel10 is the best alternative).

ValueChart provides interactive techniques that support multiple visualization manipulation methods to enable the inspection of the preference model. For instance, users can inspect the specific domain value of each objective (e.g., actual distance from the sky-train of hotel1), which is an instance of the *select* and *navigate* manipulation methods according to [BM13]. This can be performed via the *detail on demand* action (e.g., clicking on an alternative). For instance, clicking on hotel1 will show the value of each objective for this alternative. Double-clicking on a column heading *sorts* the alternatives according to how valuable they are with respect to the corresponding objective (*arrange*, according to [BM13]). Also, the position of an objective can be interchanged with another objective position through *swap* (an instance of *change*, according to [BM13]), performed by dragging. For example, if the user wants to see the aggregated weight of all alternatives based on two objectives (location and rate), she can drag rate to bring it adjacent to location in the objective tree. This will cause the related colored bars to be stacked adjacently in the stacked bar charts representing the overall values of the available alternatives. Finally, sensitivity analysis of objectives' weight is enabled by allowing the user to change the width of the corresponding column. This can be performed using the *pump* action (another instance of *change*), where the user clicks on an objective to change it by a certain increment, which changes all other objectives accordingly.

4. User Study

4.1 Individual Characteristics Explored in the Study

Our study investigates the impact of user characteristics on the effectiveness of the two different ValueChart layouts when users perform a variety of tasks related to making preferential choices. The individual characteristics we investigate include three cognitive abilities, one personality trait, and five measures of user expertise.

For cognitive abilities, we selected *perceptual speed* (PS - a measure of speed when performing simple perceptual tasks), *Visual Working Memory* (Visual WM - a measure of storage and manipulation capacity of visual and spatial information), and *Verbal Working Memory* (Verbal WM - a measure of storage and manipulation capacity of verbal information). We selected these three because they have repeatedly been shown to influence the effectiveness of (non-interactive) bar graphs [CM08, TCCH12], and ValueCharts are build on this visualization. For personality traits we selected *locus of control* because it is the personality trait that has been most reliably linked to user visualization performance so far, particularly for its impact on relative performances between simple visualization design alternatives [GF10, ZCY*11].

For expertise, we selected three measures relating to *visualization expertise*, and two relating to *domain expertise*. We look at visualization expertise because previous research has linked it to user satisfaction during simple tasks with static visualizations [TCCH12]. In this paper, we are interested in seeing whether this impact also extends to more complex, interactive visualizations. We measure *visualization expertise* in terms of familiarity with different types of bar graphs, because ValueCharts are built on these basic visualizations: familiarity with simple bar graphs with few data series and samples (*expertise-simple*); familiarity with complex bar graphs containing many series and samples (*expertise-complex*), and stacked ones (*expertise-stacked).* For *domain expertise* we include a measure of familiarity with decision making, i.e. how often a user makes preferential choices (*pref-choice-frequency),* and one that gauges how often one uses visualizations in order to make preferential choices (*use-viz-pref-choice).* We include these last two measures as domain expertise often influences performance (e.g., [LS89, Dil00]).

4.2 Visualization Layout

We tested two different ValueChart layouts in the study. The first (VC-H) uses a horizontal layout to show the various components of the decision making problem (see Figure 1), while the second (VC-V) displays the same information using a vertical layout (see Figure 2). These two layouts were originally designed because they each have pros and cons in visualizing the relevant information. For instance, in VC-H the number of objectives affects the ability to show the full names of the objectives in the tree. This is less of a concern in VC-V, because with this layout the visibility of the objectives names is only affected by the depth of the tree. On the other hand, VC-C requires the labels for alternatives to be displayed vertically (or slanted), making them harder to read. We compare these two layouts because previous studies with ValueCharts suggest that they may not be equivalent in terms of user performance [BC08]. In particular, it was found that subjects performed better on the VC-V than VC-H on low-level tasks (e.g., retrieve value & sort**),** whereas there were only non-significant trends for the high-level decision making tasks. This previous study did not account for user characteristics, thus in our study we investigate whether some of these characteristics may play a role in layout effectiveness.

4.3 Experimental Tasks

As was done in [BC08], our study included both a *high* and *low-level* tasks, which are described in detail next. Preferential

Table 1: *Description of the five low-level task types.*

Task Name	Task Definition from [AES05]	Sample task question from study	Conceptual operations involved
Retrieve Value **(RV)**	Given a set of specific cases, find attributes of those cases.	Is the value of 'skytrain-distance' of hotel3 less than hotel6?	• 2 retrieve value • 2 compare value
Find Extremum **(FE)**	Find cases possessing an extreme valued attribute over its range within a data set.	What factor contributes the most towards the overall value of hotel4?	• 10 retrieve value, • 10 compare values • 1 retrieve label of extremum
Sort **(Sort)**	Given a set of cases, rank them according to some ordinal metric.	List the top 3 hotels (in descending order) according to overall value.	• 10 retrieve value • 10 compare value • 1 retrieve labels of highest 3
Compute Derived Value 1 **(CDV-1)**	Given a set of data cases, compute an aggregate numeric representation of those data cases.	For how many homes is the 'rent' less than the 'rent' of home3?	• 10 retrieve value • 10 compare value • 1 generate aggregate
Compute Derived Value 2 **(CDV-2)**		List the top 3 homes (in descending order) according to the aggregated value of 'cost' and space'.	• 20 retrieve value • 10 generate aggregates • 10 compare aggregates • 3 retrieve labels (of highest 3 alternatives)

choice is by definition subjective, so *high-level* tasks in studies targeting this process generally involve subjective *open-ended* decision making activities that may be typically performed with a decision support system. In our study, subjects selected a domain of interest (e.g., choosing a restaurant), out of 3 available, and were then instructed to take some time to interact with the relevant VC and select their preferred item. Because preferential choice is by definition subjective, we have no gold-standard decision that can be used to evaluate our subjects' choices. Thus, as is common in decision analysis, we relied on subjective measures for decision quality (e.g., decision satisfaction & confidence), based on measures used in [BC08, Yi08].

As in [BC08], *low-level tasks* involve having users answer specific questions about the data displayed in a VC. There are five different types of low-level tasks in our study, chosen from a set of low-level data analysis tasks defined by Amar et al. [AES05]. Different task types were selected both to allow participants to experience the key functionalities of VC, and because there is evidence that the influence of individual differences may vary with task type (e.g., [CM08]). Table 1 summarizes the five low-level tasks, and for each task gives *i)* its definition from Amar et al. [AES05], *ii)* a sample question from the study, and *iii)* the conceptual operations involved. As noted in [AES05], a single low-level visualization task might require multiple additional lower-level mathematical and cognitive actions (e.g. *generate aggregate value*, *compare values*). Similarly, low-level tasks are often compounded (e.g., *compute derived value* typically requires multiple preceding *retrieve values*). The number of conceptual operations in Table 1 reflects this additional layer of actions per task type, and provides us with a conceptual measure of complexity (i.e., more steps entailing more complexity). The specific operations in Table 1 are based on our study tasks, which always involved domains with 10 different alternatives and 10 objectives. Figure 3 below shows the actual average number of VC actions performed by users during the study to accomplish the ensemble of conceptual operations for each task type. As shown in Table 1, CDV-2 is the most complex task in terms of the quantity of conceptual operations involved. It also requires a significantly higher number of overall actions to be performed, compared to all other tasks ($p < .001$), as per a 2 (viz layout) by 5 (task type) repeated measures ANOVA with overall interface actions as the dependent measure.

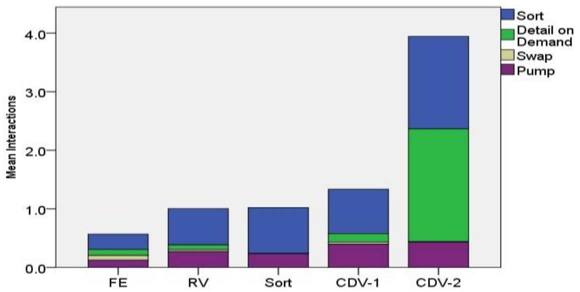

Figure 3: ***VC Interface actions for each low-level task***

4.4 Study Procedure

We had 99 participants (age range 16 to 40, 49 female), who were mostly recruited via dedicated systems at our institution. The number of participants was determined a priori based on a power analysis [EFB96] given the parameters of our experimental design, defined to detect a small effect size of at least $R^2 = .01$ with 0.8 power.

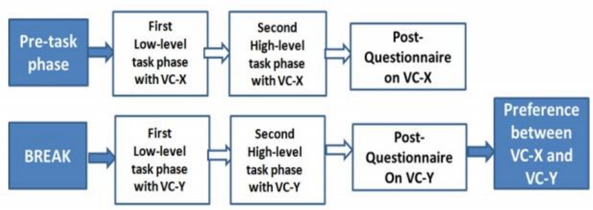

Figure 4: *Experimental Procedure.*

As shown in Figure 4, each session consisted of two parts divided by a break. Each part included a low-level and high-level task phase, during which participants performed tasks with one of the VC-H /VC-V layouts. At the end of each part, users completed a questionnaire about the subjective evaluation of the visualization used and of the quality of the decisions they made in high-level tasks. Visualization layout order was fully counterbalanced to account for learning effects, making layout order a between-subject control variable in our design. It should be noted that the study also included the collection of users' gaze using a display-based non-intrusive Tobii T120 eye-tracker. We do not provide further details on this aspect because we are not discussing gaze data in this paper.

Pre-task phase. Participants began by filling out a pre-study questionnaire asking for demographic information as well as self-reported expertise with simple, complex, and stacked bar graphs. *Expertise-simple* was elicited with the question: *'How often do you look at simple Bar Graphs'*, followed by a basic bar graph with 8 bars from only 1 data series; *Expertise-complex* and *Expertise-stacked* were elicited with similar questions referring, respectively, to a bar graph with 48 bars grouped in 6 data series and to a stacked bar graph consisting of 5 stacked bars. Next, there were two questions to elicit domain expertise on decision making (*use-viz-pref-choice* and *pref-choice-frequency*): 1) *'How often do you need to make a preferential choice (e.g., to make a choice between different smartphones based on their different features)?'*; and 2) *'How often do you use visualization tools to make such preferential choices?'*[2]. All questions had five answer options ranging from *'Never'*, to *'Very frequently (several times a day)'*, with *'Occasionally (several times a month)'* as a mid-point. Users were then given standard tests for locus of control [Rot66], verbal WM [TE89], visual WM [FV09], and perceptual speed [EFHD76]. Finally, users underwent a training phase to expose them to ValueCharts and to the study tasks.

Low-level task phases. In each low-level task phase, participants performed each of the 5 low-level task types shown in Table 1 with one of the two VC layouts. Participants were given no time limit, but were instructed to complete the tasks as quickly and accurately as possible. To account for within-subject variance, we had participants repeat each task type 4 times, which is a well-established procedure in perceptual psychology experiments measuring performance in terms of time and accuracy [PHTW11, VZ00]. This process resulted in a total of 20 trials per participant for each of the two low-level phases, yielding 3960 unique measurements. The tasks for the low-level phases were drawn from two domains: 'selecting a hotel' and 'buying a home'. All low-level tasks involved 10 alternatives and 10 objectives, to avoid potential effects in the variability of the dataset. Trials were fully randomized per participant for task type, task domain, and the objectives and alternatives mentioned in each task question.

[2] For instance ConsumerReports© provides multi attribute visualizations of product rankings.

High-level task phases. In each high-level task phases, participants began by first selecting a domain among three available: 'buying a smartphone', 'selecting a university', and 'choosing a restaurant'. They then proceeded to explore this domain as needed to select their preferred item. In the second high-level phase, users could not choose the same domain again. Participants were told to take as much time as they needed to make a good decision. Due to the open-ended nature of the high-level tasks, we did not have repeated trials in this phase. As for low-level tasks, all high-level tasks involved the same number of elements (10 alternatives, 10 objectives).

Post-Questionnaires. After each sequence of low-level and high-level phases, users completed a questionnaire to evaluate various components of VC-H and VC-V, as well as their decision making experience with each visualization in terms of decision confidence and satisfaction. For the latter, as it is commonly done in experiments on decision support (e.g., [CM06]), they provided agreement ratings on the following statements on a 5-point Likert scale: '*I am satisfied with the decision I made*', & '*I am confident about the decision I made*'.

5. Results

Table 2: *Summary statistics of user characteristics data collected from the study.*

	Min.	Max	Mean	Std. Dev.
Perceptual Speed	25	67	46.7	7.1
Visual WM	.00	5.2	2.25	1.16
Verbal WM	3	6	4.7	.9
Locus of Conrol	3	23	12.1	3.7
Expertise-Simple	2	5	3.1	.9
Expertise-Complex	1	5	2.3	.9
Expertise-Stacked	1	3	1.6	.6
Pref-Choice-Frequency	1	5	3.0	.9
Use-Viz-Pref-Choice	1	5	2.3	1.0

A correlation analysis over our 9 user characteristics (see Table 2) showed that *expertise-simple, expertise-complex*, and *expertise-stacked* were strongly correlated ($r > 0.54$, $p < .01$). Thus, we retain only *expertise-complex* as a measure of visualization expertise, given its higher variance and range. This leaves 7 user characteristics for data analysis.

5.1 Performance with Low-level Tasks

For low-level tasks, we look at both *task completion time* (*time* from now on) and task *accuracy* as performance measures. Since the low-level tasks involved multiple trials (e.g., each subject performed the same experimental condition four times), a suitable means for analysis is a *Mixed Model* [FH03]. We run separate mixed models for *time* and *accuracy,* because mixed-models do not support multiple dependent measures. We compensate for family-wise error using a Bonferroni adjustment of n=2, with statistical significance reported post-correction at the .05 level. All reported pairwise comparisons are also further corrected with Bonferroni adjustments. We calculate effect sizes (R^2) following [ASB94], and report R^2= .01 as small, R^2= .09 as medium, and R^2= .25 as large.

For each dependent measure, we first run a mixed model on the experimental factors, i.e., a 2 (vis layout) by 5 (task type), with layout-order and trial-order as a between-subject factors. Our data set thus consists of 3960 rows of data (99 participants doing 4 trials for each of the 2 vis layout and 5 task types). Next, we analyze the effects of each of our seven user characteristics separately, by running a mixed model with the study factors and that characteristic as a covariate. Due to the high number of covariates in our study, this approach ensures that we do not overfit our models by including all co-variates at once. Results relating to user characteristics are reported by discretizing these measures using a three-way split. *Low* represents the bottom quartile of the values distribution (lower 25%), *average* the interquartile (middle 50%), and *high* the upper quartile (top 25%).

Table 3 shows a selection of results that were found for time and accuracy, focusing on the effects of individual differences and VC layout. A discussion of these effects follows, organized around the two subgroups of results identified by the bolded line in Table 3. Main effects for task type and user characteristics are not reported as they are all qualified by interactions.

Table 3: *Significant results for time and accuracy on low-level tasks. 'PS' stands for perceptual speed.*

Effect	Time	Accuracy
ExpertiseComplex * Task Type	$F_{(16,2933)}$ =2.59, p < .001, R^2= .06	-
VerbalWM * Task Type	$F_{(12,2765)}$ =4.44, p < .001, R^2= .11	-
VisualWM * Task Type	$F_{(52,2851)}$ =2.46, p <.001, R^2= .08	-
PS * Task Type	$F_{(112,3101)}$ =2.35, p < .001, R^2= .08	$F_{(112,2803)}$ =1.48, p < .001, R^2= .05
VisualWM * Vis Layout	$F_{(13,2725)}$ =3.31, p < .001, R^2= .04	$F_{(13,2627)}$ =2.03, p < .05, R^2= .02

5.1.1 Interactions between User Characteristics & Task Type

As Table 3 shows, four of our seven co-variates have an interaction effect with task-type, namely the three cognitive abilities and the retained measure of expertise (*expertise-complex*). It should be noted that, although some of the effect sizes are low, they are per-task effects and they would be compounded if a user performs many low-level tasks, as typically done in a prolonged visual analytics session. For the **interaction of expertise-complex with Task Type on *time***, pairwise comparisons show that for the CDV-2 task, participants with high expertise are significantly faster (M=33.8s, SD=22.0) compared to both low expertise (M=43.5s, SD=25.2) and average expertise (M=42.3s, SD=28.3). Recall that CDV-2 tasks are the most complex among the low level tasks, in terms of both number of conceptual operations and VC actions (see Figure 3 and Table 1). Hence, this result may indicate that differences in visualization expertise start having an effect when tasks get more demanding. However, it may also be the case that this difference is due to the increase in aggregate manipulation operations required for CDV-2 (i.e. 'generate aggregates' and 'compare aggregates'). Nonetheless, to the best of our knowledge, this is the first work to directly link *visualization expertise* (as opposed to domain expertise) to task performance. Others have tested this user trait [TCCH12], but have failed to find an effect possibly because their tasks were not sufficiently complex or did not involve enough aggregate operations.

For the **interaction effect of Verbal WM with Task Type on *time***, pairwise comparisons reveal that the only significant result is that low Verbal WM users are slower than the others during Sort tasks (Sort and CDV-2), as shown in Figure 6 below. This result could be due to the fact that these two tasks require sorting the available alternatives by one or more objectives. To do so, the user first needs to identify the relevant objective(s) by scanning through all the textual labels in the objective tree, which may be more taxing for a user with low Verbal WM. This explanation is in line with previous findings based on gaze data analysis, showing that low Verbal WM users tend to spend more time looking at the textual elements of a visualization (labels, question text) compared to high Verbal WM users [TCCH12].

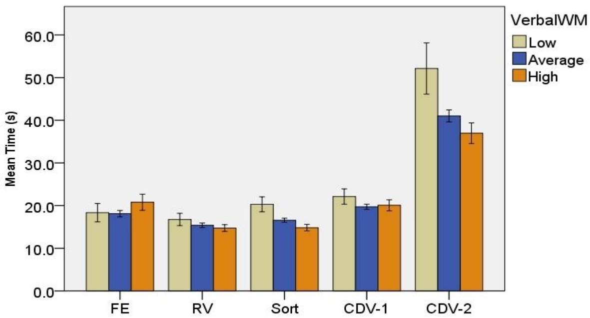

Figure 6: *Interaction with verbal WM & task type for time.*

For the **interaction effect of Visual WM with Task Type on *time***, pairwise comparisons revealed that Visual WM only has an effect during the Sort task, with low Visual WM users being slower than average Visual WM users. We currently do not have an explanation for this result, but we expect to gain more insights from the analysis of eye gaze data that we collected during the study, following the approach proposed in [TCSC13].

For the **interaction effect of perceptual speed (PS) with Task Type on time,** pairwise comparisons indicate that this trait has an effect during almost all tasks except for *Sort* (see Figure 7 below). High PS users are significantly faster than the others for FE, CDV-1, and CDV-2, and both high and average PS users are significantly faster than the low PS users for RV tasks. These results show that the impact of perceptual speed on task performance (previously uncovered for static visualizations [TCCH12, CM08]) extends to Value Charts, a more complex interactive visualization.

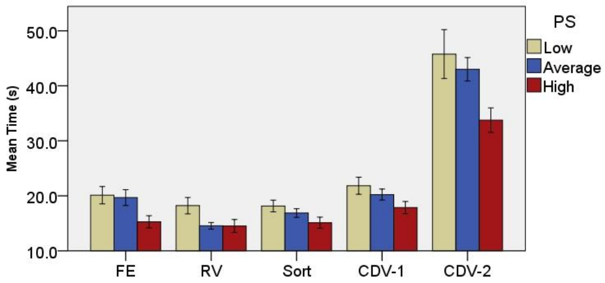

Figure 7: *Interaction between PS and task type on time.*

5.1.2 Interaction effects of VisualWM and VisLayout

We found significant interactions of Visual WM and Vis Layout on both time and accuracy, as illustrated in Figure 8 below. For completion time, pairwise comparisons indicate that both low and average Visual WM users are significantly faster with the horizontal layout (VC-H) compared to the vertical one (VC-V). Furthermore, when working with VC-H, low Visual WM users are faster than high Visual WM users. It should be noted that speed with VC-H for users with lower levels of Visual WM does not appear to come at the expense of accuracy, as there is no significant difference among the accuracy of the three groups when working with VC-H (Figure 8, right).

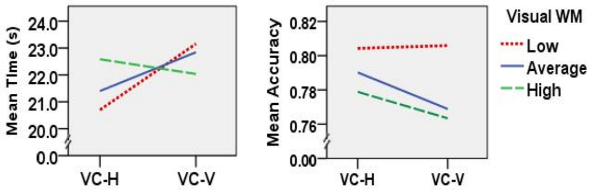

Figure 8: *Interaction effects of Visual WM with Vis Layout for completion time (left), and task accuracy (right).*

Although the actual difference in time performance between low and high Visual WM users with VC-H is rather small, these results are quite strong because previous findings linking Visual WM with performance during visualization tasks so far have shown that users with lower Visual WM are at a disadvantage [CCH*14]. The finding that in our study this is not the case when these users work with VC-H, contributes evidence to the idea that giving users the appropriate visual artifacts for their cognitive abilities (e.g., a VC-H for users with low to average levels of Visual WM) can compensate for limitations in these abilities. A similar result was found by Conati & MacLaren [CM08], but involved changing visualization altogether for users with high vs. low PS, and related to only one of the several tasks addressed in that study (a CDV-type task). Our result, in contrast, involves a simpler change (visualization orientation), and it is not necessarily qualified by task type.

5.2 Performance with High-level tasks

Recall that high-level tasks involve using a VC to explore a set of alternatives in a given domain and then select a preferred item. Thus, there is no notion of task accuracy and the only objective performance measure that we can consider is time taken to make a decision. Since there was only one trial in each high-level phase, we use a GLM repeated measures for this analysis. As we did for performance with low-level tasks, we first run a 2 (Vis Layout) by 2 (Layout Order) GLM to investigate the effects of the experimental factors, followed by additional models to investigate the effects of each of the user characteristics as a co-variate. The only interesting finding from this analysis is an interaction effect between *VisLayout* and *UseVizPrefChoice* ($F_{1,96}$=13.52, p < .001, R^2= .14), namely the self-rated frequency of using visualizations for making preferential choices (see Figure 9 below).

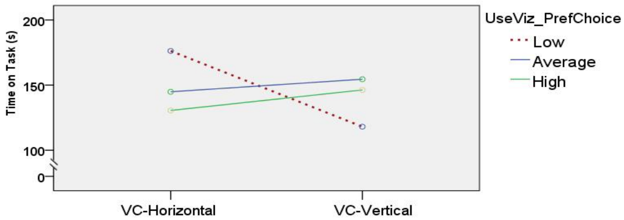

Figure 9*: Time spent for high-level tasks split by reported frequency of visualization use for making preferential choices.*

Pairwise comparisons indicate that participants with low *UseVizPrefChoice* took significantly longer to make their decision with VC-H than with VC-V. Yet, decision quality does not only depend on speed, but also on whether the decision maker is satisfied and confident with her decision. So, VC-V would be better than VC-H for users with low *UseVizPrefChoice* only if VC-V supported decisions in which these users were equally or more confident and satisfied than with VC-H. To verify if this was the case, we performed a GLM repeated measures analysis with the two ratings for decision confidence and satisfaction that we collected, as dependent measures (analogous to the one we ran for performance measures). This test yielded no significant results, however lack of significance does not necessarily indicate that there is no difference, thus we ran a follow-up equivalency test on *UseVizPrefChoice* using the methodology described in [RL11]. Our results show that decision confidence and decision satisfaction with respect to the layout are in fact equivalent for users with low *UseVizPrefChoice*. This result, coupled with the finding that these users perform significantly faster with the vertical layout, indicate that they would benefit from such a layout. More generally, this result is another indication that VC layout can affect performance for users with different abilities and background, and thus should be further investigated for our long-term goal of providing personalized support.

It may seem surprising that cognitive abilities showed an impact on performance with low-level visualization tasks, both in this paper and in previous work, and yet we found no effects of these abilities on our high-level decision making task. A possible explanation is that with high-level tasks, participants were able to compensate for lower levels of cognitive abilities with other abilities relevant to decision making, including familiarity in using visualization tools for making preferential choices. In future work, we plan to investigate other abilities previously linked to performance in decision making, such as *spatial* [Gon05] and *reasoning* abilities [SVV99].

6. Summary and Discussion

This paper presented a study to investigate the impact of user characteristics on performance with ValueCharts, an interactive visualization for preferential choice-making. We extend existing work on the effect of individual characteristics on interactive information visualizations by: (i) investigating a more comprehensive range of user characteristics, i.e., cognitive abilities, personality traits, and expertise; (ii) studying their impact on both low-level tasks as well as a high-level decision making task; and (iii) including a broader array of visualization manipulation methods than considered in prior studies. Lastly, while previous work compared visualizations that differ in terms of functionalities, our study examines two visualizations that only differ in terms of layout.

The study research questions were as follows: (1) Do user characteristics impact performance on low-level tasks with ValueCharts and (1a) Are these effects mediated by task type? (2) Do user characteristics impact performance on a high-level decision making task with ValueCharts? (3) Are the effects for 1& 2, mediated by the visualization layout (e.g., horizontal vs. vertical)? We summarize our results in light of our long term goal of providing support for visualization processing, tailored to the relevant user characteristics, as identified in the study.

For question 1 and 1a, results were found attributing performance differences to task type and all three cognitive measures tested (PS, Visual WM, and Verbal WM), as well as expertise in using complex bar charts. The effect of expertise shows that differences in this trait affects performance in more complex low-level tasks, or at least in tasks that contain significantly more aggregate operations, suggesting that personalized support should be available to non-experts for such tasks. The results for the other cognitive measures are more nuanced. For instance, users with lower measures of PS performed worse than their high measures counterparts in all tasks except one. Users with lower measures of Verbal WM performed worse than their higher measures counterparts in sorting-related tasks (Sort and CDV-2). Whereas these findings provide insights on which combinations of cognitive abilities and task types may warrant personalized help, they do not indicate how this help should be provided. We plan to investigate this issue by analyzing the gaze data that we collected during the study, following the approach in [TCSC13]. The goal is to understand which specific features in a given task hinder performance if one has lower measures of a specific trait (e.g., we may find that users with low measures of Verbal WM may have problems processing the labels in the objective tree during sort-related tasks). This information can then be used to design personalized interventions that alleviate the identified problem (e.g., find ways to facilitate the processing of the VC objective tree for low Verbal WM users).

We also found two results indicating that visualization layout mediates the effects of individual differences on performance, providing a positive answer for question 3.

The first result is that, for low-level tasks, there was an interaction between the effects of Visual WM and layout: users with lower values of Visual WM perform significantly faster with VC-H than with VC-V, and with VC-H they actually perform better than high Visual WM users. Together with previous results showing that low Visual WM can negatively affect performance, this finding supports the idea that adapting visual artifacts to users' cognitive abilities (e.g., using a horizontal instead of a vertical layout for lower Visual WM users) can compensate for limitations in these abilities. This finding also suggests that, if a vertical layout has to be used because of other constraints, additional support should be provided for lower Visual WM users. Gaze data analyses may, once again, generate insights on which aspects of a vertical layout is actually slowing these users down, informing how to design this support.

The second result related to question 3 is that participants with low frequency in using visualizations for preferential choice spent significantly more time making their decision with VC-H than with VC-V, with similar levels of subjective decision confidence and satisfaction. This result also provides a positive answer for our research question 2, i.e., user characteristics impact user performance on high-level decision making tasks, and suggests that personalization based on VC layout should be further investigated for these tasks. To do so, we plan to run additional studies focused specifically on high-level decision making tasks, and including other user characteristics that have been shown to impact decision making (e.g. spatial and reasoning abilities).

Acknowledgements

This work is sponsored by NSERC Canada and GRAND NCE Canada.

References

[AA03] ANDRIENKO N., ANDRIENKO G.: Informed spatial decisions through coordinated views. Information Visualization 2, 4 (2003), 270–285. 3

[AES05] AMAR R., EAGAN J., STASKO J.: Low-level components of analytic activity in information visualization. In . IEEE Symp. Information Visualization (2005), pp. 111–117.5

[All00] ALLEN B.: Individual differences and the conundrums of user-centered design: Two experiments. Journal of the american society for information science 51, 6 (2000), 508–520. 2

[ASB94] AB SNIJDERS T., BOSKER R. J.: Modeled variance in two-level models. Sociological Methods & Research 22, 3 (1994), 342–363. 7

[ATS95] ASAHI T., TURO D., SHNEIDERMAN B.: Visual decision-making: using treemaps for the analytic hierarchy process. In Conference companion on CHI (1995), pp. 405–406. 3

[BC06] BAUTISTA J., CARENINI G.: An integrated task-based framework for the design and evaluation of visualizations to support preferential choice. In Proc. Advanced visual interfaces (2006), pp. 217–224. 3

[BC08] BAUTISTA J., CARENINI G.: An empirical evaluation of interactive visualizations for preferential choice. In Proc. Advanced visual interfaces (2008), pp. 207–214. 2,3,4,5

[BGR06] BÜRING T., GERKEN J., REITERER H.: Usability of overview-supported zooming on small screens with regard to individual differences in spatial ability. In Proc. Advanced visual interfaces (2006), pp. 233– 240. 2

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. IEEE Trans. Vis. & Comp. Graphics 19, 12 (2013), 2376–2385. 1,2,3

[CC97] CHEN C., CZERWINSKI M.: Spatial ability and visual navigation: An empirical study. New Review of Hypermedia and Multimedia 3, 1 (1997), 67–89. 2

[CCb*13] CHAMBERLAIN B. C., CARENINI G., ÖBERG G., POOLE D., TAHERI H.: Applying tests of equivalence for multiple group comparisons: demonstration of the confidence interval approach. IEEE Trans. on Computers, Special Issues on Computational Sustainability 16, 7 (2013). 2

[CCH*14] CARENINI G., CONATI C., HOQUE E., STEICHEN B., TOKER D., ENNS J.: Highlighting interventions and user differences: Informing adaptive information visualization support. In Proc. CHI (2014). 8

[CL04] CARENINI G., LOYD J.: Valuecharts: analyzing linear models expressing preferences and evaluations. In Proc. Advanced visual interfaces (2004),pp. 150–157. 1,2,3

[CM06] CARENINI G., MOORE J. D.: Generating and evaluating evaluative arguments. AI 170, 11 (2006), 925–952. 3,6

[CM08] CONATI C., MACLAREN H.: Exploring the role of individual differences in information visualization. In Proc. Advanced visual interfaces (2008),pp. 199–206. 2,4,5,8

[CMS99] CARD S. K., MACKINLAY J. D., SCHNEIDERMAN B.: Readings in information visualization: using vision to think. 1999. 1

[Dil00] DILLON A.: Spatial-semantics: How users derive shape from information space. Journal of the American Society for Information Science 51, 6 (2000), 521–528. 4

[EFB96] ERDFELDER E., FAUL F., BUCHNER A.: Gpower: A general power analysis program. Behavior Research Methods, Instruments, & Computers 28, 1 (1996), 1–11. 5

[EFHD76] EKSTROM R. B., FRENCH J. W., HARMAN H. H., DERMEN D.: Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational testing service, 1976. 6

[FH03] FIELD A. P., HOLE G.: How to design and report experiments. Sage publications London, 2003. 7

[FV09] FUKUDA K., VOGEL E. K.: Human variation in overriding attentional capture. The Journal of Neuroscience 29, 27 (2009), 8726–8733. 6

[GF10] GREEN T. M., FISHER B.: Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. In Proc. IEEE VAST (2010), pp. 203–210. 1,2,3

[GLG13] GRATZL S., LEX A., GEHLENBORG N., PFISTER H., STREIT M.: Lineup: Visual analysis of multi-attribute rankings. Trans. Vis. & Comp. Graphics 19, 12 (2013), 2277–2286. 1,2,3

[Gon05] GONZALEZ C.: Task workload and cognitive abilities in dynamic decision making. Human Factors: The Journal of the Human Factors and Ergonomics Society 47, 1 (2005), 92–101. 9

[Gra06] GRAWEMEYER B.: Evaluation of erst–an external representation selection tutor. In Diagrammatic Representation and Inference, 2006, pp. 154–167. 1,3

[Lia87] LIANG T.-P.: User interface design for decision support systems: a self-adaptive approach. Information & Management 12, 4 (1987), 181–193. 3

[LS89] LEWANDOWSKY S., SPENCE I.: The perception of statistical graphs. Sociological Methods & Research 18, 2-3 (1989), 200–242. 4

[PBW*12] POMMERANZ A., BROEKENS J., WIGGERS P., BRINKMAN W.-P., JONKER C. M.: Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. User Modeling and User-Adapted Interaction 22, 4-5 (2012), 357–397. 3

[PHTW11] PALMER E. M., HOROWITZ T. S., TORRALBA A., WOLFE J. M.: What are the shapes of response time distributions in visual search? Journal of experimental psychology: human perception and performance 37, 1 (2011), 58. 6

[RL11] RUSTICUS S. A., LOVATO C. Y.: Applying tests of equivalence for multiple group comparisons: demonstration of the confidence interval approach. Practical Assessment, Research & Evaluation 16, 7,(2011),1-6. 9

[Rot66] ROTTER J. B.: Generalized expectancies for internal versus external control of reinforcement. Psychological monographs: General and applied 80, 1 (1966), 6

[SCC14] STEICHEN B., CONATI C., CARENINI G.: Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data. In TIIS (2014), 3

[SNFB95] SANKAR C. S., NELSON FORD F., BAUER M.: A dss user interface model to provide consistency and adaptability. Decision Support Systems 13, 1 (1995), 93–104. 3

[SVV99] SPEIER C., VALACICH J. S., VESSEY I.: The influence of task interruption on individual decision making: An information overload perspective. Decision Sciences 30, 2 (1999), 337–360. 9

[TCCH12] TOKER D., CONATI C., CARENINI G., HARATY M.: Towards adaptive information visualization: on the influence of user characteristics. In UMAP (2012), pp. 274–285. 2,4,7,8

[TCSC13] TOKER D., CONATI C., STEICHEN B., CARENINI G.: Individual user characteristics and information visualization: Connecting the dots through eye tracking. In Proc. CHI (2013), pp. 295–304. 8,9

[TE89] TURNER M. L., ENGLE R. W.: Is working memory capacity task dependent? Journal of Memory and Language 28, 2 (1989), 127–154. 6

[VST05] VELEZ M. C., SILVER D., TREMAINE M.: Understanding visualization through spatial ability differences. In IEEE Visualization (2005), pp. 511–518. 2

[VZ00] VAN ZANDT T.: How to fit a response time distribution. Psychonomic Bulletin & Review 7, 3 (2000), 424–465. 6

[WPTMS12] WONGSUPHASAWAT K., PLAISANT C., TAIEBMAIMON M., SHNEIDERMAN B.: Querying event sequences by exact match or similarity search: Design and empirical evaluation. Interacting with computers 24, 2 (2012), 55–68. 2,3

[Yi08] YI J. S.: Visualized Decision Making: Development and Applications of Information Visualization Techniques to Improve Decision Quality of Nursing Home Choice. PhD thesis, Georgia Institute of Technology, 2008. 3,5

[ZCY*11] ZIEMKIEWICZ C., CROUSER R. J., YAUILLA A. R., SU S. L., RIBARSKY W., CHANG R.: How locus of control influences compatibility with visualization style. In Proc. IEEE VAST (2011), pp. 81–90. 2,4