Evaluating Multivariate Visualizations as Multi-Objective Decision Aids

Meirav Taieb-Maimon[1], Lior Limonad[2], David Amid[2],
David Boaz[2], Ateret Anaby-Tavor[2]

[1]Information Systems Engineering, Ben-Gurion University, Israel
meiravta@bgu.ac.il
[2]IBM Research – Haifa, Israel
{liorli, davida, davidbo, atereta}@il.ibm.com

Abstract. When choosing a solution, decision makers are often required to account for multiple conflicting objectives. This is a situation that can result in a potentially huge number of candidate solutions. Despite the wide selection of multivariate visualizations that can potentially help decide between various candidates, no designated means exist to assess the effectiveness of these visualizations under different circumstances. As a first contribution in this work, we developed a method to evaluate different types of multivariate visualization. The method focuses on the visualization's ability to facilitate a better understanding of inter-objective trade-offs as a proxy to more sensible decision making. We used the method to evaluate two existing visualization aids: Parallel-Coordinates and an adaptation of Self Organizing Maps (SOM). Both visualizations were compared with tabular data presentation. Our results show that the first visualization is more effective than a plain tabular visualization for the purpose of multi-objective decision making.

Keywords: Multi-criterion decision making; Multivariate visualizations; Information Visualization; Usage experience evaluation.

1   Introduction

There are many circumstances where a person has to decide among competing alternatives. From a consumer trying to choose the right car, house, or cell phone to a business executive who must decide upon a new portfolio of product offerings, and on to elected representative voting on national health policies—all of them must choose among multiple, often competing, alternatives. Many studies in the area of *multi-criterion decision making* (MCDM) attempt to help decision-makers reach better conclusions in a more efficient way [18]. The multi-criterion decision making process typically examines various alternatives with respect to their values for each criterion. In most practical scenarios, the number of options is too large to be examined by a human, and decision makers aspire to examine a limited set of options. Towards this end, the MCDM research community has identified two major challenges: (a) reducing the number of options by means of an optimization process that yields a smaller

set of optimal solutions, termed Pareto Frontier [6], and (b) effectively visualizing these solutions (namely, the solution space) to help users select the solutions that best satisfy their subjective criteria. The task of visualizing the Pareto Frontier is based upon multivariate visualization, and is generally considered a hard problem for more than three objectives [19].

The decision making process becomes even more challenging when some of the objectives being considered conflict with each other. In this conflicting relationship, the performance in one objective is seen to deteriorate as performance in another is improved. In such settings, actual choices should ideally reflect the tradeoffs that are in keeping with the decision maker's priorities. For example, given a certain budget, a person who travels frequently and wants to buy an new laptop would tend to give weight a higher priority over performance, while a "gamer" would aim for increased performance and settle for a heavier weight.

We distinguish between *sensible* and *non-sensible* choices. Sensible choices faithfully correspond to the decision-maker's subjective perception of inter-objective tradeoffs and non-sensible choices do not. According to Lotov et al. [22], "Tradeoff information is extremely important for the decision maker since it helps to identify the most preferred point along the tradeoff curve." Hence, it is expected that the better the decision-maker's understanding of the various tradeoffs in the solution space, the more sensible their choices will be. Correspondingly, we introduce the following as an underlying premise:

Assumption 1: A better understanding of the inter-objective conflicts results in more sensible decision making.

Derived from the above assumption, our main objective in this work is *to provide decision makers with a concrete means to improve the understanding of inter-objective conflicts.*

Ideally, identifying a sensible choice should rely on first discovering the perceived trade-offs the decision maker may have with regard to a given set of objectives [10, 11, 26]. Pragmatically, extracting even a single trade-off function seems to be highly cumbersome, and in some cases very inaccurate [12]. Furthermore, in some cases, the decision maker may become aware of subjective preferences only after confronting the actual solution space [12].

Since a purely analytic approach is deemed unfeasible, we chose to focus on the area of 'visual analytics,' in which visual interfaces are used to facilitate interactive reasoning [14]. Having decision makers closely involved in data exploration and processing cycles eliminates the need to extract and formalize their subjective preferences. Despite the plethora of multivariate visualization aids that exist in this domain (e.g., SOM [25], Interactive Decision Maps [21], Parallel Coordinates (PC) [13]), we found no prior work on user-experience evaluation to direct our selection of a designated visual interface that would facilitate multi-objective decision making. As a result, we decided to investigate and test the effectiveness of existing visual interfaces in promoting the selection of sensible choices. The evaluation of information visualization is an area in which the development of designated metrics and benchmarks has been identified as invaluable [28]. Our work is also unique in tackling the special

circumstances in which any choice made is deemed acceptable since all competing alternatives are quantitatively equivalent (all being on the pareto frontier). It is the sole (unknown) preference of each decision maker that can determine the subjective ranking of the solution space.

We consider our contribution as twofold: (1) providing a pragmatic method and a carefully adapted measurement scale for testing the effectiveness and usage-experience of multivariate visualization aids for MCDM, and (2) providing a first glance into the actual performance of two existing visualization techniques for promoting effective MCDM. In the following sections we present the preliminary process of adopting and refining our experimental instruments, followed by a complete report of our experimental settings and results.

2 The Effectiveness of Visualization Aids

The presented work is driven by the growing desire evident in the HCI literature for the development of alternative methods to evaluate visualization and encourage more widespread adoption of visualization [31]. Specifically, we focused on the unique nature of interaction with multivariate information visualization to more effectively benefit from and measure insights discovered [30] rather than extracting efficiency measures such as time to complete a task. In our experimental configuration, we did include a measurement of task completion time for which no significant differences had been found across the different visualization types[1]. Hence, we considered effectiveness according to the following proposition:

Proposition 1: A (multi-dimensional) *visual representation* of the solution space promotes better *decision making sensibility* than a *plain representation* of the solution space.

For the purpose of a *visual representation*, we used two alternative techniques: Parallel Coordinates (PC), which is considered the prevalent technique for the visualization of multi-objective data [1, 7, 32], and Self-Organizing Map for Multi Objective (SOMMOS), a recent adaptation of a Self-Organizing Map (SOM) [17] visualization being developed by IBM as a proprietary visualization for exploring and visualizing a Pareto Frontier.

Figure 1 illustrates an example for the two visualizations, showing three different apartments being considered for rent. Each alternative is measured based on three objectives: number of roommates, price, and distance from the university. On the left hand side, the PC visualization is depicted with each axis assigned a corresponding objective. Along each axis, an arrow indicates the aspiration of the decision maker to either minimize or maximize the specific objective. In Figure 1, the decision maker's goal is to minimize all three objectives. Each choice is depicted as a line that intersects every axis in accordance with the associated objective's value. The decision maker can observe the extent of conflict between objectives by searching for a cross

---

[1] p=.186 and p=.144 for the two experimental tasks RENT and TAM.

(an 'x' pattern) between axes. For instance, based on the cross pattern apparent between the 'Roommates' axis and the 'Price' axis, the decision maker can infer a strong conflict between the corresponding objectives. In contrast, by examining the lines between the 'Price' axis and the 'Distance from University' axis, the decision maker can conclude that there is no conflict between the corresponding objectives based on the absence of any crossing.
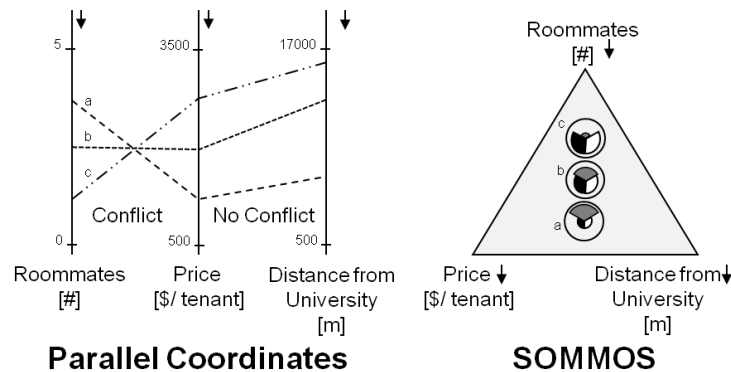

Fig. 1. - The PC and SOMMOS visual representations evaluated

On the right hand side, the same solution space is depicted using SOMMOS. SOMMOS is a variation of the well-known self-organizing map (SOM) algorithm [17] adapted for visualizing Pareto Frontiers. The SOMMOS visualization comprises three pillars: a map layout, a visual representation of individual choices, and interactive capabilities. The map layout enables the user to explore and draw insights based on its structure. The objectives are represented by the vertices of the map. For example, the three objectives in Figure 1 form a triangular layout. One of SOMMOS' virtues is the ability to navigate through the map. Specifically, the decision maker can expect that the closer the choice is to a vertex of an objective, the closer its value is towards the goal of that objective. Each choice on the map is encoded using polar area chart glyphs. Each objective is assigned a corresponding colored slice. By noting the level of colored filling, decision makers are able to understand the nature of the compromise in an area, spot an area of choices that maximize\minimize one of the objectives at the expense of the other, and identify cases in which objectives are correlative. In Figure 1, the decision maker can clearly identify that the Roommates objective is in conflict with both Price and Distance from University such that as the slice filling for Roommates decreases, the slice filling for Price and Distance increases. The decision maker can also observe that across all choices, the filling pattern for both Distance from University and Price is consistent. Hence, there is no conflict between the two.

In addition to testing each of the visualizations, we suspected that the two visualization types may complement one another. Because each has its unique strengths and weaknesses, allowing decision makers to simultaneously interact with both could potentially yield better results. Nevertheless, it was unclear whether the cognitive

effort associated with toggling between two visualizations would impair the potential impact of combining the two. As a result, our experimental settings included a visual interface that combined PC and SOMMOS.

We used a simple table for the purpose of a *plain representation*, being synonymously referred to as the "no visualization" in this work. We structured the table with columns designating objectives and rows designating possible choices. Table 1 shows a plain representation of the same data underlying the solution space illustrated in Figure 1.

Table 1. - Plain representation of the solution space in Figure 1

	Roommates (#)	Price ($/tenant)	Distance from University (m)
a	4	400	700
b	3	600	1500
c	2	800	2300

As an antecedent to the notion of *decision making sensibility*, we developed a designated scale (see next section) to measure *subjective conflict understanding*. This helped us adhere to Assumption 1, according to which sensible choices rely first and foremost on the decision maker's ability to understand the nature of trade-offs between the objectives.

Correspondingly our experimental hypothesis is as follows:

H1: A (multi-dimensional) *visual representation* of the solution space promotes better *subjective conflict understanding* than a *plain representation*.

In addition to the measures that relate directly to the above hypothesis, we considered several background factors that may interfere with the effectiveness of the visualization being used. We extended our research model as illustrated in Figure 2 to consider three types of moderating factors:

Subjective usage experience – Several measures were employed to reflect the way users feel about using the visualizations. Such experience may be a potential moderator to the main effect hypothesized above. However, the secondary effect of each of the visualizations on usage experience was in itself unclear. Hence, this led to a second hypothesis:

H2: A (multi-dimensional) *visual representation* of the solution space promotes better *usage experience* than a *plain representation*.

We adopted a set of concrete preliminary measures from previous literature and further refined them as part of a card sorting procedure, explained in the next section.

Problem characteristics – The particular *problem domain* and *complexity* may also be a factor affecting the degree to which the visualization means can facilitate better objective trade-off understanding. The complexity, for example, may affect its capacity to scale in size of data [33]. As a result, we decided to manipulate these factors to account for their influences.

Participant factors – As noted by Gemino et al. [9], individual factors can affect the effectiveness of a visualization. To account for potential interference, we included questions to measure participants' familiarity with the corresponding problem domain and their experience. The latter included general questions about experience using visual analytics and specific questions regarding experience with multi-objective decision making.
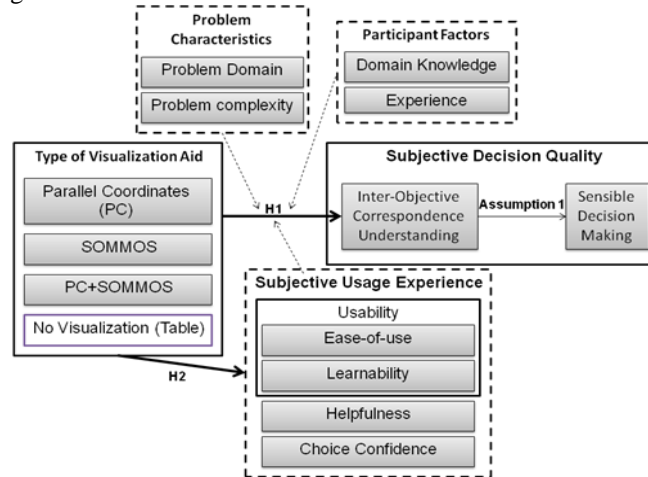

Fig. 2. - Extended research model

3 Scale Development

Aside from the main effectiveness construct, all other scales have been originally adopted from prior literature having solid foundations in the HCI literature. To ensure scale validity and reliability in the operationalization of our research model, we produced a preliminary set of questions based on existing scales, followed by a *card sorting* procedure [3, 4, 24]. This led us to the scale items detailed in Table 5. Card sorting[2] is an iterative scale development technique in which a panel of judges is asked in several rounds to sort a set of scale items into separate categories, based on similarities and differences among them. This technique was previously used by Davis et al. [3, 4] to assess the coverage of an intended domain of constructs. Later, it was further refined [24] to generally account for scale reliability (i.e., content and construct validity). Prior to conducting the first card sorting round, we populated items for each of the constructs as described next.

As mentioned in the preceding section, we operationalized the high level notion of **subjective decision quality** through the construct of **inter-objective correspondence understanding** (i.e., as an antecedent to sensible decision making). This construct is both innovative to this work and related to the particular decision problem domain being explored. As such, no prior measurement scales were available for adoption.

---

[2] We used OptimalSort by OptimalWorkshop.com as a platform for card sorting.

We produced a set of seven items based on our experience with the notion of inter-objective trade-offs. Specifically, all items were generated in accordance with the general definition by Purshouse et al. [29] according to which inter-objective trade-off understanding is defined as the degree to which an individual believes he or she were able to understand the 'conflict' between the objective. The 'conflict' is interpreted as a relationship between any two objectives in which as performance in one is improved, performance in the other is seen to deteriorate. Each related question was worded such that its most correct answer corresponds to one of the extreme ends in a 1-to-9 Likert scale [23].

To operationalize the high level notion of *subjective usage experience* associated with the decision making task, we adopted a preliminary set of post-test concrete constructs from prior literature as follows:

(1) **Usability** – being conventionally used to capture the general appropriateness of an artifact to a particular purpose [2]. In our study, the artifact being evaluated corresponds to the various visualization interfaces mentioned earlier. The concrete purpose corresponds to the facilitation of multi-objective decision making. We refined the common 10-item SUS scale [2] to match such circumstances. In addition, we eliminated two questions with very similar phrasing to two of the other questions that were intended to measure ease-of-use.

(2) **Ease-of-use** – adopted from Moore and Benbasat [24], originally aimed at measuring the degree to which users believe that using a particular system would be free of physical and mental effort. Correspondingly, we adopted a preliminary set of six items and slightly re-worded the phrasing to match the intended usage purpose.

(3) **Cognitive-load** – the actual cognitive capacity required for the decision making task [27]. The measurement of the actual cognitive capacity is typically considered a multi-dimensional construct. In our study we adopted a corresponding measurement scale from DeLeeuw and Mayer [5], to measure two items reflecting: (a) **mental-effort** – the self reported effort subjects indicate as required to reach a certain decision, and (b) **task-difficulty** – the self reported difficulty subjects indicate as being associated with the decision making task. Stemmed from the basic notion of task difficulty, and the particular nature of the task in this work, we added two additional items. These items take into account the capacity of the visualization to assist in the recognition of similar choices (i.e., clusters), and its capacity to gradually narrow down the set of preferred choices until a single choice is identified.

(4) **Decision-confidence** – in its broader sense reflects the subject's confidence in the certainty of estimates or predictions of future performance based on their perceptions of their knowledge and experience. Adopted from the work by Kidwell et al. [16], this construct was measured by a unitary item which we slightly re-worded to reflect the confidence between the choice that is made and subjective preferences.

We operationalized the two remaining construct groups: *problem characteristics* and *participant factors* in a more straightforward manner. We incorporated problem characteristic factors into the experimental design by manipulating the problem's domain and the problem's complexity. Its exact levels of manipulation are further explained in the following section. We incorporated participant factors by including three questions on background knowledge: familiarity with the problem domain, pre-

vious experience using visual-analytics interfaces, and previous experience in solving multi-objective problems.

The set of all items being produced for the measurement of the main effect (i.e., H1), together with all other items populated for the measurement of the resulting user experience (i.e., H2), were jointly considered as the input pool for the first card sorting round. We fragmented this preliminary set as illustrated in Table 2.

Table 2. - Preliminary construct-wise item pool for card-sorting input

Conflict understanding	7	Usability	8	Ease of use	6
Mental effort	1	Task difficulty	3	Decision confidence	1

To eliminate ambiguous items, we conducted three separate card sorting rounds with 21 independent judges. As noted by Moore et al. [24], this was aimed at increasing construct validity and our confidence in the developed scale. The procedure was an 'open card sort' style, implying that the judges in the first round were not told what the underlying constructs (i.e., item categories) were, rather they were asked to provide their own construct labels and definitions. We used three indicators to identify construct validity and the convergence of the sorting procedure: (1) Inter-rater agreement between the judges for each construct in each round, (2) Hit ratio ("convergent" validity) – item-wise portion of true item classifications within its "target" construct, and (3) Missed ratio ("discriminant" validity) – construct-wise portion of wrong item classification within each construct. In each round, a "doesn't fit" category was also permitted for items that did not match any of the other clusters. A summary of results is illustrated in Table 3.

After the first sorting round, we observed a significant cross loading between ease-of-use and usability. As noted above, this can be traced back to [20] having the SUS scale being composed of two sub-factors to measure usability and learn-ability, the former having conceptual overlap with ease-of-use. In addition, two items had poor loading on either ease-of-use or usability. We dropped these items and split the remaining items as being measures of either ease-of-use or learn-ability, considering usability as the combination of both (i.e., a second degree construct). **Learn-ability** is defined as the degree to which the tool enables the user to learn how to use it.

Table 3. - Card sorting results

	Round 1 (8 judges)				Round 2 (6 judges)				Round 3 (7 judges)			
	# items	Inter-rater agree	Hit Ratio	Missed Ratio	# items	Inter-rater agree	Hit Ratio	Missed Ratio	# items	Inter-rater agree	Hit Ratio	Missed Ratio
Conflict Understanding	7	0.75	1.00	0.05	7	0.76	0.76	0	7	0.71	0.8	0.03
Usability	8	0.61	0.28	0.50	-	-	-	-	-	-	-	-
Ease of use	6	0.55	0.50	0.59	8	0.57	0.75	0.25	7	0.51	0.88	0.14
Learnability	-	-	-	-	4	0.46	0.75	0.14	3	0.67	0.86	0.1
Mental effort	1	0.67	0.25	0.50	-	-	-	-	-	-	-	-
Task difficulty	3	0.52	0.71	0.41	-	-	-	-	-	-	-	-
Helfulness	-	-	-	-	3	0.38	0.55	0.57	3	0.5	0.86	0.14
Decision Confidence	1	1.00	0.25	0.00	-	-	-	-	-	-	-	-
Choice confidence	-	-	-	-	1	0.33	0.67	0.67	1	0.24	0.86	0.45
Doesn't fit		0.42	-	-		1	-	-		None	-	-
Average		0.68	0.50	0.34		0.50	0.70	0.33		0.53	0.85	0.17

The single item measuring mental-effort had a significantly stronger load on ease-of-use than on itself. This was also explained by the definition of ease-of-use being:

"the degree to which an individual believes that using a particular system would be free of physical and mental effort". Hence, we decided to drop the entire construct, since it was sufficiently captured in the measurements for ease-of-use.

Task-difficulty had very poor inter-rater agreement and inconsistent definitions across the judges. Consequently, we concluded that the actual construct being measured should be labeled as **decision-making helpfulness.** This captured the extent to which the visualization aid helped in the decision making process.

Decision-confidence also had very low hit ratio. Hence, we reworded the item to better reflect the certainty of choice made based on the perception of one's own knowledge and experience. We labeled this construct as **choice confidence**. This concluded all modifications made according to the results of the first sorting round.

After the second card sorting round, we dropped two additional items due to loading ambiguities. One was originally intended to measure ease-of-use but appeared to load equally on choice-confidence. The second item was originally intended to measure learn-ability and appeared to load equally on ease-of-use. To further increase item loading, we reworded two additional items (7 and 20) to better reflect their intended constructs. This concluded the modifications for the second sorting round.

Finally, after the third sorting round, the only apparent problem was a relatively low discriminant validity for choice confidence. Due to its high convergent validity, and no consistent classification to any of the items wrongly classified as its potential measures, no further actions seemed necessary. All remaining constructs demonstrated reasonable inter-rater agreements, alongside good convergent and discriminant validities. Hence, we decided to keep all items unchanged and concluded sorting rounds with the final loadings as illustrated in Table 4. A summary of all scale items is depicted in Table 5. The developed scale is our first contribution in this paper and can be used to assess any multivariate visualization aimed to promote MCDM.

Table 4. - Item loading matrix after the sorting procedure

	Choice confidence	Conflict understanding	Decision making helpfulness	Ease of use	Learn-ability
EOU 1				100.00%	
EOU 2			14.29%	71.43%	14.29%
EOU 3	14.29%			85.71%	
EOU 4				85.71%	
EOU 5				100.00%	
EOU 6				100.00%	
EOU 7				71.43%	14.29%
LBT 8				14.29%	85.71%
LBT 9				14.29%	85.71%
LBT 10				14.29%	85.71%
CON 11		85.71%		14.29%	
CON 12		85.71%		14.29%	
CON 13		85.71%	14.29%		
CON 14	14.29%	71.43%		14.29%	
CON 15	14.29%	71.43%		14.29%	
CON 16	14.29%	85.71%			
CON 17	14.29%	71.43%	14.29%		
HLP 18			85.71%		
HLP 19	14.29%		85.71%		
HLP 20		14.29%	85.71%		
CNF 21	85.71%				

Table 5. - Final items after the sorting procedure (EOU – ease-of-use, LBT – learnability, CON[3] – conflict understanding, HLP – decision-making helpfulness, CNF – choice confidence)

Item	Item Description
EOU1	The visualization aid was cumbersone to use
EOU2	Using the visualization aid required a lot of mental effort
EOU3	using the visualization aid was often frustrating
EOU4	using the visualization aid was clear and understandable
EOU5	Overall, I believe that the visualization aid is easy to use
EOU6	I found the usage of the visualization aid unnecessarily complex
EOU7	I think there was too much inconsistency in the usage of the visualization aid
LBT1	Learning how to use the visualization aid was easy for me
LBT2	I would imagine that most people would learn to use the visualiztion aid very quickly
LBT3	I needed to learn a lot of things before I could start using the visualization aid
CON1	Based on all presented alternatives, an increase in the value of <objective1> requires a decrease in the value of <objective2>
CON2	Based on most presented alternatives, an increase in the value of <objective1> requires a decrease in the value of <objective2>
CON3	In most presented alternatives, the two objectives <objective1> and <objective2> seem to be conflicting
CON4	With regard to the conflict between <objective1> and <objective2>, there are solutions that denote a good compromise between them
CON5	With regard to the conflict between <objective1> and <objective2>, there are solutions that denote a good compromise between them
CON6	Solutions <1> and <2> are strong indicators to having a conflict between <objective1> and <objective2>
CON7	Solution <1> denotes a good compromise between <objective1> and <objective2>
HLP1	The visualization aid did not help me find the selection solution
HLP2	Using the visualization aid helped me focus the selection to the finding of a preferred solution
HLP3	The visualization aid eased the decision making task be helping me identify groups of solutions that have similar characteristics
CNF1	I have strong confidence that my choice fits my personal preferences

4 Experimental Design

We pursued a 4X2 (visualization type, problem domain) mixed experimental design to evaluate our hypotheses. In the laboratory setting, each participant was randomly assigned to one of four visualization interface groups: Tabular, PC, SOMMOS, PC and SOMMOS. The participants were then asked to make two consecutive choices, each corresponding to a different problem domain.

The first problem domain was in the area of property rental, and the task was to select an apartment for rent (namely, RENT). We selected this domain because it was an area with which most participants are familiar with. The corresponding visualization of the solution space was composed of 20 apartments to be examined according to 3 objectives: number of roommates, price, and distance from university; all these objectives were intended to be minimized.

The second problem domain was in the area of transportation asset management (namely, TAM). We chose this domain because it was an area of less familiarity. The

---

[3] Note that in all CON items, the phrases appearing in brackets were populated by concrete objectives/solutions in the context of the concrete problem domain.

corresponding visualization of the solution space was composed of 100 possible transportation plans (e.g., a portfolio of transportation projects), to be examined according to 5 objectives: air quality, safety, foreseen economic growth, traffic scarcity, number of pedestrian/cyclist trails; all these objectives were intended to be maximized. The order in which the two problems were presented to each participant was counter balanced across participants.

5   Method

Procedure. The entire experimental procedure took an hour on average. In the beginning of the experiment, each participant was given a 15 minute computer-based training session, which included: a primer to multi-criteria decision making and a video demonstration of the visualization type corresponding to the participant's group. The students were then given a quiz with 10 questions to determine whether they had understood the concepts and the usage of the tool presented in the training. A passing score of 70% was targeted to ensure all participants understood the visualization presented to them and its operation. Following the quiz, each participant was given a questionnaire booklet designed to facilitate the experimental procedure. The booklet had the following sections: participation consent, a set of questions about the participant's background knowledge and experience, and two question sets, corresponding to the two decision-making problems. In each such set, participants were first instructed to use the visualization aid for the purpose of choosing a single alternative out of the presented set, and then record both their selection, and a short explanation justifying their choice. Next, participants were presented with a randomly shuffled set of questions about usage experience and objective trade-off understanding.

Pilot. We carried out a pilot with 30 undergraduate students to further test the procedure and all corresponding experimental materials. Results indicated a significant effect of the visualizations on the perceptions of ease-of-use ($F_{(3,11)}=3.94$, $p=.039$) and usability ($F_{(3,11)}=3.76$, $p=.044$) for the RENT problem domain. Although an impact on conflict understanding was not yet apparent, it was decided to go ahead with the main experiment as the sample size in the pilot was fairly limited. We incorporated some minor re-wording after the research team reviewed the pilot responses.

Participants. A group of 93 undergraduate students volunteered to participate in the main experiment, which followed a similar scheme to the pilot. As an incentive for participation, students were offered 2 bonus points in a related academic course. After eliminating students who failed the quiz (i.e., 3 in total who scored below the 70% score threshold), and questionnaires that had been improperly filled out not adhering to the instructions, 172 useful responses remained in total: 85 for the RENT problem

domain and 87 for the TAM problem domain. These responses were analyzed and all results are detailed in the following section.

6   Results Analysis

Data Preparation. Prior to running the statistical analysis, we explored the data using BOXPLOT to determine extreme responses. We concluded that 5 responses were outliers, scoring more than 1.5 times the inter-quartile range[4] (i.e., 4 in RENT, and 1 in TAM). Correspondingly, we executed all subsequent analyses twice, with and without the outliers. We found that outlier elimination does not affect any of the significant findings. Hence, final results reporting utilized 80 responses for the rental problem domain, and 85 for the transportation problem domain.

Scale Reliability. Our analysis of the results checked the scale reliability for the measurement of usage experience constructs. Item-wise reliability scores for each construct are illustrated in Table 6, showing that the reliability level indicated by Cronbach's alpha met the suggested tolerance (> 0.7, [23]).

We skipped the reliability check for the scale developed for conflict understanding. Unlike the other constructs that are traditionally considered reflective [8], the items used to measure conflict understanding were not expected to be correlated to one another (i.e., a formative construct). Indeed, this was corroborated by exploratory factor analysis that has been attempted for each problem-domain question set, yielding 3 different principle components associated with each question set in each problem domain. Furthermore, no conceptual grounds could have been attributed to the underlying components, supporting the irrelevance of this test [8]. That is, the inherent content validity of the scale developed for conflict understanding is shaped by how the questions complement each other.

Table 6. - Reliability scores (Cronbach's alpha)

Variable	Ease of use	Helpfulness	Learnability	Usability	Confidence
Alpha	0.87	0.71	0.70	0.88	Single-item

6.1   Subjective Decision Quality

We conducted a separate one-way ANOVA for each problem domain to examine the effect of the visualization interface (between groups independent variable) on how well participants understood the relationship between objectives (dependent variable). The results depicted in Table 7 indicate a significant effect of the visualization type only for the case of RENT ($F_{(3,77)}=2.86$, $p=.042$). We performed the analysis for each problem domain separately since the actual measurements for correspondence understanding (i.e., items in questionnaire) were different between the two problem domains.

---

[4] The inter-quartile range (IQR) contains the middle 50 percent of the distribution. If the data is normally distributed, a range that is 1.5 times the IQR covers ~99.3% of the distribution.

Table 7. - Inter-objective correspondence understanding (scores ranged from 1 = least understanding to 9 = highest level of understanding)

Case	PC			SOMMOS			PC+SOMMOS			Table			ANOVA		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n	df	F	Sig.
Rent	7.51	0.72	22	7.06	0.85	22	7.13	0.56	16	6.85	0.82	21	(3,77)	2.86	0.042*
TAM	6.51	0.76	23	6.65	0.97	22	6.41	0.85	20	6.30	0.88	21	(3,82)	0.636	0.594

Post-hocs. We conducted pair-wise comparisons between the visualization interfaces once the significant main effect was apparent for the case of RENT. Consistent with the ANOVA analysis, LSD post-hoc comparisons indicated a significant difference between the PC visualization (*M*=7.51, *SD*=.72) and the plain representation (*M*=6.85, *SD*=.82), p=.005. No further significant pair-wise differences were evident.

6.2 Subjective Usage Experience

We conducted a two-way (visualization type, problem domain) MANOVA to test the possible effects of the visualization type and problem domain on usage experience. In this case, it was possible to integrate problem domains in the analysis as the questions pertaining to all usage experience factors were the same in both problem domains. The two-way interaction (visualization type*problem domain) was not found to be significant for any of the usage experience factors, allowing us to interpret each of the main effects separately.

Descriptive statistics for the effect of visualization type on the usage experience factors are illustrated in Table 9. The main effect of the visualization type on the perception of ease-of-use was significant ($F_{(3,157)}$=2.81, *p*=.04). Similarly, the effect of the visualization type on the perception of usability was also significant ($F_{(3,157)}$=2.81,*p*=.04). The effect of the visualization type on the remaining usage experience factors was not found to be significant. However, it might be worth noting a somewhat moderately significant effect of the visualization type on learn-ability ($F_{(3,157)}$, *p*=.06).

Descriptive statistics for the effect of problem domain on the usage experience factors are illustrated in Table 10. As in the case of visualization type, the main effect of the problem domain was also significant with regard to its impact on the perception of ease-of-use ($F_{(1,157)}$=9.51, *p*=.002), and on the perception of usability ($F_{(1,157)}$=7.66, *p*=.006). In addition, choice confidence was also significantly affected by the problem domain ($F_{(1,157)}$=4.64, *p*=.03). All remaining usage experience factors were not found to be significantly affected by the problem domain.

Post hocs. We conducted pair-wise comparisons between the visualization interfaces following the conclusion of the significant main effects for ease-of-use and usability. With respect to the effect of the visualization type on ease-of-use, LSD post-hoc comparisons indicated two significant differences: between the PC visualization (*M*=6.13, *SD*=1.63) and the plain representation (*M*=7.06, *SD*=1.36), p=.005, and also between the latter and the combined visualization of PC & SOMMOS (*M*=6.37, *SD*=1.47), p=.048. No further significant pair-wise differences were evident.

With respect to the effect of the visualization type on usability, LSD post-hoc comparisons indicated three significant differences: between the PC visualization (*M*=6.55, *SD*=1.28) and the plain representation (*M*=7.32, *SD*=1.18), *p*=.008; between the SOMMOS visualization (*M*=6.74, *SD*=1.53) and the plain representation, *p*=0.046; and between the combined visualization of PC & SOMMOS (*M*=6.67, *SD*=1.26) and the plain representation, *p*=0.032. There were no further significant pair-wise differences evident in the effect on usability.

Table 8. - Usage experience - between problem domains and between visualization types

Effect source	Dependent Variable	(df, 157)	F	Sig.
VisualType	Ease of use	3	2.81	0.04*
	Learnability	3	2.49	0.06
	Helpfulness	3	0.73	0.54
	Choice confidence	3	0.29	0.83
	Usability	3	2.81	0.04*
Problem domain	Ease of use	1	9.51	0.002*
	Learnability	1	0.97	0.33
	Helpfulness	1	2.09	0.15
	Choice confidence	1	4.64	0.03*
	Usability	1	7.66	0.006*

Table 9. – Descriptive statistics for usage experience factor per visualization type (1-9 scale)

Variable	PC			SOMMOS			PC+SOMMOS			Table		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n
Ease of use	6.13	1.63	45	6.53	1.65	43	6.37	1.47	36	7.06	1.36	41
Helpfulness	7.17	1.34	45	7.35	1.30	43	7.57	1.11	36	7.26	1.21	41
Learnability	7.53	.99	45	7.24	1.56	43	7.37	1.15	36	7.93	1.14	41
Usability	6.55	1.28	45	6.74	1.53	43	6.67	1.26	36	7.32	1.18	41
Confidence	7.56	1.25	45	7.40	1.76	43	7.33	1.47	36	7.61	1.39	41

Table 10. – Descriptive statistics for usage experience factor per problem domain (1-9 scale)

Variable	Rent			TAM		
	M	SD	n	M	SD	n
Ease of use	6.90	1.48	80	6.16	1.56	85
Helpfulness	7.48	1.10	80	7.19	1.47	85
Learnability	7.63	1.17	80	7.42	1.30	85
Usability	7.12	1.29	80	6.54	1.34	85
Confidence	7.74	1.25	80	7.24	1.62	85

6.3 Moderator Analysis

With respect to each of the factor groups proposed as possible moderators to the influence of visualization type on inter-objective correspondence understanding (i.e., H1), we hereby note the following results:

Problem Characteristics. We incorporated both problem domain and problem complexity factors into the experimental design, along with the manipulation of the four visualization types. As implied from Table 7, there was an interaction between problem characteristics and the main effect of the visualization interface. This was apparent in having a significant impact only for the case of lower complexity (i.e., RENT). However, since the manipulation of problem-characteristics was simultaneously facilitated by changing the number of alternatives in the solution

space (20 vs. 100), the number of objective (3 vs. 5), and the problem domain (RENT vs. TAM), it is impossible to conclude at this point which one (or more) of the three individual factors were responsible for the moderation.

Participant Factors. Background knowledge in any of the two problem domains, and prior experience in using visual interfaces and in solving multi-objective problems were all incorporated as pre-test questions. We statistically tested the potential effect of these factors as possible covariates. Specifically, we incorporated the following factors as potential covariates: *knowledge in problem domain*, *experience in using visual interfaces*, and *experience in multi-objective decision making*. None of the participant factors were found to significantly moderate the main affect of the visualization type on conflict understanding.

Subjective Usage Experience. It is apparent in the results that the manipulation of problem domain also had a significant impact on *ease-of-use*, *choice confidence*, and *usability* as depicted in Table 6. Similar to the moderation of problem characteristics, this impact may have influenced the potential main effect of the visualization interfaces in the case of the more complicated problem domain. Hence, we statistically tested the potential effect of these factors as possible covariates. None of the usage experience factors were found to significantly moderate the main effect of the visualization type on conflict understanding. Correspondingly, we conducted all post-hoc analyses illustrated in the previous section without integrating the covariates.

7 Conclusions

Based on the results and the research model in Figure 2, we concluded that:

— The visualization of PC is more effective in terms of inter-objective conflict understanding than a tabular visualization when used in the context of a multi-objective decision problem that is not overly complicated and in a problem domain with which the decision maker has higher familiarity. Beyond a certain complexity threshold and\or when the problem domain is less familiar to the user, the gain from using a visual interface diminishes. This finding is in line with prior research [15] according to which domain familiarity should be taken into account as a significant background factor that can interact with the effectiveness of treatments in the context of problem solving tasks. Interestingly, this impact is not affected by the decision maker's background knowledge and experience with using multivariate visualizations.
— The type of the visualization aid has a significant impact on subjective usage experience, supporting H2. Specifically, the perceptions of ease-of-use and usability show that a plain representation typically scores higher than a graphical visualization aid. This may be explained by the widespread use of tabular visualizations.
— The decision making problem at hand may also facilitate different perceptions of ease-of-use and usability. Further, decision making confidence is also affected by

the problem characteristics. This effect should be carefully taken into consideration as it may also indirectly mask the effectiveness of the visualization aid in promoting better inter-objective correspondence understanding. Our experimental design employed a confounded manipulation, simultaneously tweaking problem domain (RENT vs. TAM), and problem complexity in two dimensions: size of solution space (20 vs. 100), and number of objectives (3 vs. 5). Thus, further work is required to determine which of the concrete dimensions may actually be a source for effect moderation.

Aside from the above conclusions inferred directly from the statistical analysis, our research team deduced some qualitative insights while administrating the experimental procedure. The most prominent observations included the following:

— Despite our expectations regarding the superiority of the combined visualization (PC+SOMMOS) over the individual constituents, it did not appear to have fulfilled this expectation. Observing user actions revealed that most participants who were assigned to this group used each of the two visualization interfaces independent of the other, practicing a fairly limited number of transitions between them. The root cause we observed for this behavior was the mere use of visual 'tabbing' which did not allow participants to simultaneously watch the two visualizations on a single screen. This may also be attributed to the content of the training session presented to this group.
— It was observed that users do not immediately rush into making a choice. Rather, the decision making process progressed through several phases. This included: high-level visual skimming of the entire solution space, prioritizing between the presented objectives, identification of choice groups that seem to possess similar characteristics, flipping back and forth between the upper and lower bounds for each objective range and observing its effect on the visible solution space while doing so, and gradually iterating through all previous phases while eventually converging on a single choice. Based on this observation our conclusion is twofold: (a) the suitability of concrete visualization means with regard to each of the above phases may vary, and (b) the multi-objective decision making process itself should be further investigated to more diligently learn about its methodological constituents.

8   Future Work

Different levels of complexity revealed the possible existence of a certain cognitive threshold beyond which the effectiveness of the visualization means is impaired. Our immediate intention is to pursue an experiment aimed at testing the exact complexity parameters that may play a significant role in this regard (e.g., number of objectives, solution space size).

As concluded from the observed methodological aspects of the decision making process, and in congruence with [31], a longitudinal investigation is necessary to fur-

ther discover the actual cognitive stages associated with more effective decision making in multi-objective settings.

Our preliminary experiment was focused on the understanding of trade-off as the main outcome variable. Understanding the potential impact of visual interfaces on other outcome variables (e.g., cluster specific trade-off understanding) requires further investigation.

Finally, it is clear that certain features in the visualization interface can promote better decision making performance. For example, the 'filtering' feature allowing users to restrict the accepted range per objective was used extensively, regardless of the actual multi-dimensional visualization type. There are still a handful of additional features that need to be tested for similar impacts, including visual clustering, objective layout shuffling, glyph shapes, visualization layouts, and more.

References

1. Bagajewicz, M., Cabrera, E.: Pareto optimal solutions visualization techniques for multiobjective design and upgrade of instrumentation networks. Industrial & engineering chemistry research. 42, ACS Publications, 5195–5203 (2003).
2. Brooke, J.: SUS-A quick and dirty usability scale. Usability evaluation in industry. (1996).
3. Davis, F.: A technology acceptance model for empirically testing new end-user information systems: Theory and results. (1985).
4. Davis, F.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly. (1989).
5. DeLeeuw, K., Mayer, R.: A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. Journal of Educational Psychology. (2008).
6. Ehrgott, M.: Multicriteria optimization. (2005).
7. Fleming, P. et al.: Many-objective optimization: An engineering design perspective. Evolutionary Multi-Criterion Optimization. pp. 14–32 Springer (2005).
8. Freeze, R., Raschke, R.L.: An Assessment of Formative and Reflective Constructs in IS Research. ECIS 2007 Proceedings. (2007).
9. Gemino, A., Wand, Y.: A framework for empirical evaluation of conceptual modeling techniques. Requirements Engineering. 9, 4, 248–260 (2004).
10. Green, P., Srinivasan, V.: Conjoint analysis in consumer research: issues and outlook. Journal of consumer research. (1978).
11. Green, P.E., Wind, Y.: New Way to Measure Consumers' Judgments. Harvard Business Review. 53, pp. 107–117 (1975).
12. Horn, J.: Multicriteria decision making and evolutionary computation. (1996).
13. Inselberg, A., Dimsdale, B.: Parallel coordinates: a tool for visualizing multi-dimensional geometry. Proceedings of the 1st conference on Visualization'90. 361–378 (1990).
14. Keim, D. et al.: Visual Analytics: Definition, Process, and Challenges. Neurosciences. 154–175 (2008).

15. Khatri, V. et al.: Understanding conceptual schemas: Exploring the role of application and IS domain knowledge. Information Systems Research. 17, 1, 81–99 (2006).
16. Kidwell, B. et al.: Emotional calibration effects on consumer choice. Journal of Consumer Research. (2008).
17. Kohonen, T. et al.: Self-Organizing Maps. Springer, Berlin (2001).
18. Köksalan, M. et al.: Multiple Criteria Decision Making: From Early History to the 21st Century. (2011).
19. Korhonen, P., Wallenius, J.: Visualization in the multiple objective decision-making framework. Multiobjective Optimization. (2008).
20. Lewis, J., Sauro, J.: The factor structure of the system usability scale. Human Centered Design. (2009).
21. Lotov, A. et al.: Interactive decision maps: Approximation and visualization of Pareto frontier. (2004).
22. Lotov, A., Miettinen, K.: Visualizing the Pareto frontier. Multiobjective Optimization. (2008).
23. M.K. Trochim, W.: Research methods knowledge base. (2007).
24. Moore, G.C., Benbasat, I.: Development of an instrument to measure the perceptions of adopting an information technology innovation. Information systems research. 2, 3, 192–222 (1991).
25. Obayashi, S. et al.: Multi-objective design exploration for aerodynamic configurations. AIAA Paper. (2005).
26. Orme, B.: Getting started with conjoint analysis. (2006).
27. Paas, F. et al.: Cognitive load measurement as a means to advance cognitive load theory. Educational psychologist. 38, 63–71 (2003).
28. Plaisant, C.: The challenge of information visualization evaluation. Proceedings of the working conference on Advanced visual interfaces AVI 04. 11, 4, 109–116 (2004).
29. Purshouse, R., Fleming, P.: Conflict, harmony, and independence: Relationships in evolutionary multi-criterion optimisation. Evolutionary Multi-Criterion Optimization. (2003).
30. Shneiderman, B. et al.: Creativity support tools: Report from a US National Science Foundation sponsored workshop. International Journal of Human-Computer Interaction. 20, 2, 61–77 (2006).
31. Shneiderman, B., Plaisant, C.: Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization. pp. 1–7 ACM (2006).
32. Witowski, K. et al.: Decision making in Multi-Objective Optimization for Industrial Applications–Data mining and visualization of Pareto data. Proceedings of the 7th European LS-DYNA Conference. (2009).
33. Yang, J. et al.: Interactive hierarchical displays: a general framework for visualization and exploration of large multivariate data sets. Computers & Graphics. (2003).