Objective Evaluation of Subjective Metrics for Interactive Decision-Making Tasks by Non-experts

Yann Laurillau, Joëlle Coutaz, van Bao Nguyen, Gaëlle Calvary, Daniel Llerena

Objective Evaluation of Subjective Metrics for Interactive Decision-Making Tasks by Non-experts

Yann Laurillau[1], Joëlle Coutaz[1], Van Bao Nguyen[1], Gaëlle Calvary[1], and Daniel Llerena[2]

Univ. Grenoble-Alpes, CNRS, Grenoble INP, [1]LIG, [2]GAEL, 38000 Grenoble, France
<first name>.<last name>@univ-grenoble-alpes.fr

Abstract. This work addresses the evaluation of interaction techniques for decision-making tasks performed by non-experts in the context of multi-objective optimization problems. Such tasks require making trade-offs between antagonistic criteria, according to individual subjective preferences. Evaluating such techniques is made difficult by the subjective nature of such tasks, as well as by a lack of rigorous methods for assessment. Our primary contributions to this problem are two-fold: (1) a set of subjective metrics including decision accuracy, choice satisfaction, and incentives to explore; (2) the use of a pragmatic approach to map these subjective metrics onto objective quantitative measures. To illustrate how this subjective-objective mapping can be performed in a pragmatic manner, we have conducted an experiment involving 177 participants to objectively measure and compare two multi-slider interaction techniques for decision-making tasks performed by non-experts in the context of domestic energy management. The results of this evaluation constitute a secondary contribution.

Keywords: multi-criteria decision-making, evaluation, decision-making task, Pareto front, optimization problem, tightly coupled sliders, energy management.

1 Introduction

In Psychology, decision-making is a cognitive process that results in the selection of an alternative from multiple possibilities. To help this process, the field of Multi-Criteria-Decision Making (MCDM) has developed mathematical models, methods and algorithms that generate solutions for problems that involve multiple, possibly conflicting, criteria [30]. For such problems, there is no unique solution but a set of alternatives from which decision makers must choose the solution that best fits their objectives or preferences. To make informed decisions, decision makers should be supported by appropriate tools such as visualization techniques.

Most work on visualization for decision-making has focused on Multi-Attribute Decision Analysis, a subclass of MCDM for which the solution space is discrete, finite, and predetermined (such as finding a hotel room for a vacation). Dimara et al. refer to a "*multi-attribute choice task (MACT) as a task that consists of choosing the best alternative among a fixed set of alternatives where alternatives are defined across several*

attributes" [9]. In this article, we are concerned with Multi-Objective Optimization problems, another subclass of MCDM problems, for which the set of alternatives is continuous, possibly infinite, not known explicitly in advance, and where the criteria are strongly interdependent. To complement Dimara et al., we define a *"multi-objective choice task (MOCT) as a task that consists of choosing the best alternative from a continuous set of alternatives where criteria are strongly interdependent"*.

For Multi-Objective Optimization problems, it is impossible to find a solution that simultaneously gives the optimal value for all the criteria. Rather, there exist many solutions, called Pareto-optimal [1], that satisfy the problem mathematically. Because all Pareto-optimal solutions are equally good from the mathematical point of view, decision makers have to select the preferred "best" solution. This requires making trade-offs between the criteria, where making trade-offs means giving up on at least one criterion to allow the improvement of others.

As a typical example of this class of problems, consider Alice who would like to be warm with good air quality at the lowest possible cost. Suppose that her home is equipped with an e-coach energy management system capable of generating Pareto-optimal solutions for her problem [3]. The Pareto front, which corresponds to the set of optimal solutions, delimits the frontier between the set of feasible but non-optimal solutions from the set of unfeasible solutions. To be optimal, Alice must select her preferred solution from the Pareto front by deciding how much she is ready to give up on thermal comfort and air quality to reduce financial cost or vice versa. To make this final decision, Alice must draw on her subjective preferences.

In this article, we are concerned with the problem of objectively evaluating and comparing interaction techniques designed for tasks that are inherently subjective. There is a growing research interest in addressing this issue in the field of interactive visualization [2, 4, 5]. In this area, evaluation generally consists of assessing the usability of the technique as in [31], or evaluating the capacity of the technique to support data exploration for analytical tasks such as retrieving a particular value [10]. Based on qualitative studies, Boukhelifa et al. [6] show how experts resolve conflicts between competing objectives but do not address the comparative evaluation of multiple techniques. In [9], Dimara et al. investigate metrics, such as decision accuracy, to objectively compare interactive visualizations for MACTs. In this article, we address this issue for MOCTs – which, by definition, are more complex than MACTs, when performed by non-experts – who, like Alice, are not trained in this type of tasks.

Our primary contributions to the objective evaluation of interaction techniques for MOCTs are two-fold: (1) a set of subjective metrics including decision accuracy, choice satisfaction, and incentives to explore, that can be objectively and quantitatively measured; (2) the use of a pragmatic approach to map these subjective metrics onto objective quantitative measures. To illustrate how this subjective-objective mapping can be performed in a pragmatic manner, we have conducted an experiment involving 177 participants to objectively measure and compare Sliders4DM and P4DM, our own re-implementations of two existing multi-slider interaction techniques for exploring Pareto fronts, respectively, TOP-Slider [19] aimed at non-experts performing multi-objective choice tasks, and 'Pareto sliders' developed for expert surgeons [25].

In the following, we provide an overview of related work and justify our choice for multi-slider interaction techniques as representative case studies for MOCTs. We then propose a set of objective and subjective metrics for assessing these techniques and show how a pragmatic approach can be used to measure the proposed subjective metrics, both quantitatively and objectively, from logged data. The experiment conducted to show the feasibility of our approach and the results are then presented in detail. We finally discuss our findings and conclude with implications for future research.

2 Related work

In this section, we provide an overview of previous work on visualization techniques for optimization problems, along with the requirements for supporting non-experts. We then cover related work on the evaluation of tools for MOCTs.

2.1 Visualization Techniques for Optimization Problems

Visualization techniques for optimization problems have been proposed for experts in specific areas such as engineering design, business intelligence, and surgery. A number of methods including HSDC [1], 3D-RadVis [14], and ParetoBrowser [28] have been developed to visualize Pareto fronts for complex optimization problems. In particular, 3D-RadVis maps large dimensional objective spaces to 3D representations while preserving the shape of the Pareto front. However, 3D representations require specific training for interpretation. To alleviate this problem, ParetoBrowser combines 2D and 3D graphs with parallel coordinates representations. However, ParetoBrowser is intended for domain experts.

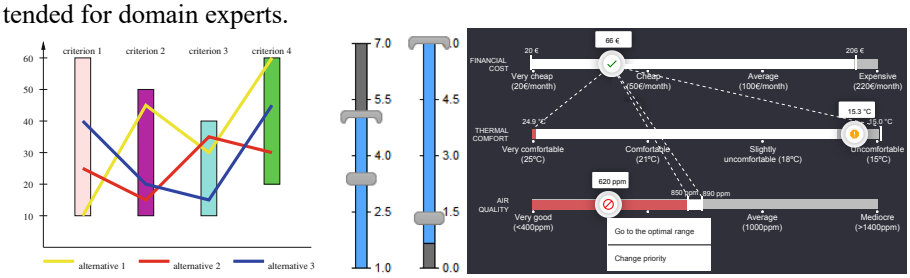

Fig. 1. Alternatives represented as value paths (extract from [21]) (left); The Pareto Slider for surgery (PSS) (extract from [25]) for two criteria (middle); TOP-Slider with three criteria (right): financial cost, thermal comfort, and air quality (extract from [19]).

For an untrained user such as Alice who has little or no knowledge in thermal models, we hypothesize the following requirements. Visualization techniques targeted at non-experts performing multi-objective choice tasks should: (R1) Hide the complexity of the optimization problem while facilitating the understanding of the mutual influence between the criteria; (R2) Favor the exploration of the Pareto front to find the preferred "best" solution in an informed manner; (R3) Notify users when moving away from the optimal solution space; (R4) Make the Pareto front observable in order to limit the attraction effect bias and to incite users to select optimal solutions [10].

Sliders are commonly used interactive tools for exploring data spaces and selecting values. Thus, we hypothesize that multi-slider based techniques can provide an appropriate foundation for supporting non-experts performing multi-objective choice tasks. As illustrated by Value Paths [21], Pareto Slider [25], and TOP-Slider [19] discussed below, multiple sliders can be combined to implicitly represent Pareto fronts in a 2D-space (cf. R4), and thereby, hiding the complexity of the underlying optimization problem (cf. R1) while favoring the exploration of the solution space (cf. R2).

Value Paths visualize Pareto optimal solutions as a set of parallel vertical bars in a 2D-coordinate space (See Fig. 1-left). Each criterion is represented by a bar whose size and location on the y-axis express the range (provided that it is known) of the criterion in the Pareto optimal set. Alternatives are represented by polygonal lines called value paths. Similarly to the parallel coordinates technique, the number of criteria can be increased to a certain degree, but having too many alternatives makes interpretation and comparison difficult [21].

Many multi-slider interactive techniques have been developed for multi-attribute choice tasks [24, 26, 27, 32], but very few have targeted multi-objective choice tasks. Pareto Slider designed for Surgeons (PSS in short) [25] and TOP-Slider [19] are notable exceptions. Both of them are composed of parallel sliders, using one slider per criterion. The Pareto front is represented implicitly as ranges of optimal values distributed across the sliders. Whereas Value Paths represent Pareto optimal ranges only, PSS shows both the optimal and unfeasible ranges using color-coding (see Fig. 1-middle). TOP-Slider goes one step further by representing the optimal, unfeasible, as well as feasible but non-optimal Pareto ranges using white, red and grey color-coding respectively (see Fig. 1-right as an illustration). Color-coding is one way to indicate users when moving away from the Pareto front (cf. R3).

Both PSS and TOP-Slider express the interdependence between criteria in a tightly-coupled manner (cf. R1). However, these techniques differ in the way they provide feedback when a cursor is moved. In PSS, moving the cursor of one slider to modify the value of its corresponding criterion, moves the cursor of the other sliders automatically so that the new position of the cursors corresponds to a Pareto optimal solution. The strategy used for choosing the new Pareto optimal solution among the possible ones is computed algorithmically, not decided by users. By contrast, with TOP-Slider, moving one cursor does not move the other cursors. Instead, as shown in Fig. 1-right, two pairs of lines pop up to show the impact of the current position of the cursor on the Pareto ranges of the other criteria. As a consequence and contrary to PSS, TOP-Slider allows users to choose their preferred optimal solution as well as to explore trade-offs that are not necessarily Pareto optimal (cf. R2).

2.2 Evaluation of Tools for Multi-Objective Choice Tasks

In their analysis of evaluation methods of tools for multi-attribute choice tasks (MACT) [9], Dimara et al. observe that "*there is a lack of methodological guidance in the information visualization literature on how to do so*." The problem is two-fold: (1) Objective metrics are not enough to capture the quality of a decision, given that "*finding a good trade-off"* is subjective. Subjective metrics such as self-reported satisfaction are useful,

but unreliable as they may be subject to cognitive biases [10]. (2) There is a lack of clear references for identifying an appropriate baseline for comparative assessment.

As a first step towards a more rigorous approach to the evaluation of tools for multi-attribute choice tasks, Dimara et al. [4] propose a combination of objective and subjective metrics for comparing parallel coordinates, scatterplot, and tabular visualizations, three commonly used elementary visualization techniques: accuracy and time-on-task as objective metrics; technique preference, satisfaction, confidence, easiness, and attachment as subjective metrics. Dimara et al. report that, for decision-making, the three techniques are comparable across the metrics with "*a slight speed advantage for the tabular visualization*". Therefore "*time-on-task can be a useful differentiating factor*." Another interesting conclusion is that "*testing real decision tasks can provide more insights*."

Although table-based visualization techniques seem more effective for decision-making than scatter plots and parallel coordinates [15], they are not applicable to optimization problems where the set of alternatives is continuous and possibly unknown in advance. As we are concerned with multi-objective choice tasks, we have elected the slider, another commonly used elementary interactive tool that supports choosing a value in a range of continuous numeric values. Although Dimara et al's work is an important contribution to the problem of evaluating tools for decision making by non-experts, sliders have not been covered by their study.

Sharing similarities with Multi-Objective Optimization problems, geospatial multi-criteria decision-making (GIS-MCDM) problems deal with an infinite and continuous set of alternatives (e.g. geographical distance). Milutinovic et al. [22] developed GISwaps, a novel method based on the concept of Even Swaps, "*a trade-off-based method for multiple criteria decision-making under certainty*". This method consists of adjusting alternatives depending on a reference criterion, and a response criterion depending on a set of "*virtual alternatives*": the key idea is to make the reference criterion "irrelevant" thanks to a compensation value applied to the response criterion. Based on this method, they conducted a quantitative comparative study to objectively evaluate the impact of interactive visualization on trade-off-based geospatial decision-making. To do so, they compute "*the average trade-off value for each virtual alternative in each swap turn (ranking results)*" and "*variation in compensation values in trade-offs*". The key result shows that interactive visualization leads to more consistent trade-offs.

Boukhelifa et al. [6] have conducted an observational study in order to understand how experts, in a collaborative setup, develop strategies to deal with "multiple competing objectives" for exploring complex solutions spaces in the context of Multi-objective Optimization. In this study, the underlying optimization model is a multi-dimensional Pareto front. One key observed strategy for trade-off is prioritization: experts often start with the most important criterion according to their expertise, then refine their exploration based on a secondary criterion.

PSS targets surgeons for planning medical radiofrequency ablation, and has been evaluated with only 2 surgeons. As discussed above, PSS enforces the choice for an optimal solution as the result of the value change of one criterion. The experiment does not address the adequacy of this strategy. To the best of our knowledge, TOP-Slider is the only example of a multi-slider interaction technique targeted at non-experts for

multi-objective choice tasks. TOP-Slider, however, has only been evaluated qualitatively with 16 participants [19].

3 Mapping Subjective Metrics with Objective Measures Using Pragmatism: Comparing TOP-Slider (S4DM) with PSS (P4DM) as a Case Study

In this section, we address the following research question: how to objectively and experimentally evaluate and compare interaction techniques designed for MOCTs? More specifically, in addition to the usual objective metrics such as time-on-task, what objective measures should be used in practice to evaluate metrics that are inherently subjective? In the following, we first present our approach to address these questions using a pragmatic approach. We then detail each step of this approach.

3.1 A 3-step Pragmatic Approach

As discussed in [9], decision-accuracy and decision-satisfaction are inherent to decision-making tasks. However, defining a measure for these metrics is challenging because of the subjective nature of decision tasks and because of the difficulty to find "good" solutions without objective methods such as Pareto-based models. To address this difficulty, we have adopted a pragmatic approach, drawing on the experimental context to map aggregated logged data as objective measures for subjective metrics.

Pragmatism is "*thinking about solving problems in a practical and sensible way rather than by having fixed ideas and theories*" (Cambdridge Dictionnary). In research, a pragmatic approach focuses on finding useful/practical solutions in a realistic context through experiential inquiry [13], "*rather than becoming mired in discussions regarding generalizability*" [16]. For this study, we propose a 3-step pragmatic approach that consists for the experimenters (1) to elicit the characteristics shared by the target users, (2) to consider the key differentiating features of the interaction techniques under evaluation, and (3) to draw on the context of use.

As a concrete example of interaction techniques and realistic context of use, we considered TOP-Sliders [19] and PSS respectively [25], in the context of residential homes equipped with a smart energy management system. Typically, users are not experts in thermal modeling, but they are familiar with thermal comfort, air quality and financial cost. Using either one of these techniques, inhabitants would express their preferences as a trade-off between comfort, air quality and financial cost. As users modify their preferences, the system would update the Pareto-based optimal solution space from which users could iteratively pick the most appropriate solution for them.

The choice for TOP-Sliders and PSS is motivated by the following: (1) They both address MOCTs using sliders as the elementary interactive technique; and (2) both of them use color-coding but differ in the way the interdependence of the criteria is reflected as well as how they suggest or enforce the choice for optimal solutions. For the sake of conformity and comparison with the qualitative study performed in [19] for

TOP-Sliders, we have elicited the same three criteria for expressing users' preferences: financial cost, thermal comfort, and air quality.

The choice for energy management as context of use is motivated by the following: (1) energy is a major world societal grand challenge for the upcoming decades (e.g., United Nation's sustainable development goals [29]); and (2) energy management is a typical example addressed by research on optimization models such as finding trade-offs between energy consumption and thermal comfort (e.g., [33]).

3.2 Step 1: Profiling with a Preliminary Study

As a first step, we propose to *identify a primary criterion that reflects the participants' profile*. This contrasts with Dimara et al. who, to measure decision-accuracy [9], relied on self-reported preferences, which are not necessarily reliable [10]

We conducted a preliminary users study [19] to evaluate TOP-Sliders qualitatively involving a limited number of participants in order (1) to improve the design of TOP-Sliders until the requirements were met satisfactorily, (2) to identify the strategies that users developed to find their preferred solution. Participants ranged between 17 and 71, of which 6 were over 40, with an average of ~38. The participants included 1 computer scientist, 9 students, and 6 family members (of which 4 retired healthy persons). 7 participants had concerns for energy consumption and financial cost and 3 of the retired participants used a technical solution to manage their own consumption at home (e.g., programming heating periods). In particular, the semi-structured interviews uncovered that all the students asserted that financial cost was more important than thermal comfort. Therefore, in the context of energy saving by participants with low income, we hypothesized that financial cost is the primary criterion.

In the following, we will refer to Sliders4DM as the reimplementation and improvement of TOP-Slider [19] and to P4DM as our own implementation of PSS [25].

3.3 Step2: Metrics for Comparison

We considered a combination of objective and subjective metrics to compare Sliders4DM with P4DM: time-on-task and interaction-workload as objective metrics; decision-accuracy, decision-satisfaction, and incentive-to-explore as subjective metrics. Time-on-task is frequently used in HCI in comparative studies. Interaction-workload is relevant, as decision-making tasks are cognitively demanding. Incentive-to-explore makes it possible to assess Requirement R2 (cf. Section 2).

C1 – Decision-Accuracy as a subjective metrics. As in [9], decision-accuracy is our first class metrics. In the experiment presented here, we considered students with limited financial resource. Thus, as observed in step 1, the choice made by the participants for financial cost can serve as an objective measure for decision-accuracy. This is motivated by the following: (1) the solution space is already Pareto optimal. As a result, the difficulty to find optimality is alleviated and (2) the participants share the same profile.

C2 – Decision-Satisfaction as a subjective metrics. Instead of post-questionnaire for subjective choice assessment [9], we propose the final position of the sliders as an

objective measure. This is motivated by the following: (1) in the instructions, the participants were asked to click the '*Validate*' button when they "were satisfied" with their choice; (2) according to Cialdini's influence principles [8], a person always tries to seek for consistency while taking decisions, especially when a decision is recorded – which was the case, as the participants were made aware that their choice was logged. In addition, referring to step 1, all interviewees involved in the preliminary users study, but one, indicated that they were satisfied with their choice.

C3 – Incentive-to-Explore as a subjective metrics. For this criterion, we used the order in which the participants used the sliders. In particular, we focused on the order of the first three sliders used to analyze the exploration and possibly detect corrective actions. Data exploration is a basic analytic task, a necessary component for multi-attribute choice tasks [8]. Furthermore, according to Dimara et al. [8], "analytic tasks are informative when evaluating visualization tools for decision support, because good decisions require a good understanding of the relevant data." Consequently, we consider that a "good understanding" implies understanding the impact of each criterion through the manipulation of each slider.

C4 – Time-on-Task as an objective metrics. As in [9], we considered the time to achieve the task as well at the fine grain action level using the time spent to drag cursors or to reach and click buttons. This duration includes the durations of idle moments (e.g., no interaction) and the total activity duration that is the sum of the duration of the atomic actions such as moving a cursor.

C5 – Interaction-workload as an objective metrics. For the purpose of comparing P4DM with Sliders4DM at the interaction level, we have considered the number of atomic actions (e.g., dragging a cursor or a restrictor knob, clicking a button), the number of mouse movements to drag a cursor or a restrictor knob, as well as trajectory lengths (in pixels) of the cursors when moved with the mouse. The goal is to ensure that usability does not impact the decision-making task.

3.4 Step 3: Comparing Sliders4DM with P4DM

Fig. 2 shows Sliders4DM and P4DM. For both of them, we have reused the color-coding scheme and layout of TOP-Sliders [19]. Similarly, Sliders4DM and P4DM share the same Pareto front modeled by Equation (1) where each criterion is represented by a normalized value between 0 and 1. This model was developed with experts in energy consumption to satisfy the "real decision task" condition put forward by Dimara et al [9]. The goal is to focus on the intrinsic differences between the two interaction designs.

$$4((x-1)^2+\tfrac{1}{5})(y-1)^2-\tfrac{8}{15}\,x^2-2\,z+\tfrac{3}{10}=0 \text{ with (x,y,z)} \in [0,\,1] \quad (1)$$

In the following, we detail the main characteristics of Sliders4DM and P4DM as well as their key differences. From this link [20] the interested reader can play with the two techniques.

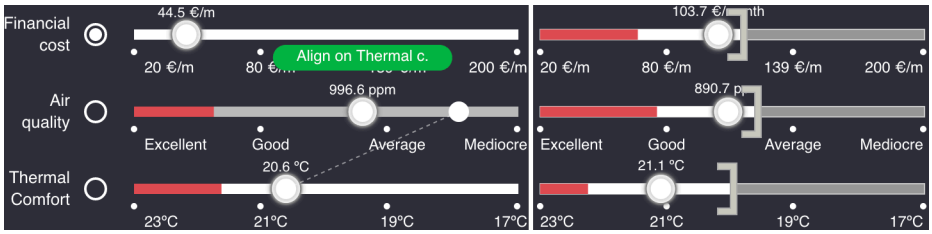

Fig. 2. (left) Screenshot of Sliders4DM, the adapted version of TOP-Slider used for the comparative experiment. Here, financial cost has been selected as the primary criterion. The user is now moving the cursor for Thermal comfort. As a result, a dashed-white line pops up and links this cursor to a small white circle that suggests an optimal solution for the third criterion. As the cursor is moved, the white-filled circle moves accordingly in a tightly-coupled manner. (right) Screenshot of P4DM, our own implementation of PSS.

Sliders4DM: TOP-Slider adapted. As shown in Fig. 2-left, Sliders4DM integrates the suggestions from [19], such as introducing radio buttons as an explicit means to support the priority-based strategy developed by the participants during the qualitative experiment. This is backed up by Milutinovic et al.'s observational study [22] (see section 2.2). In addition, we have improved the interactive behavior of TOP-Slider as the result of a number of expert evaluations conducted with colleagues.

The notable differences between Sliders4DM and TOP-Slider are the following:

Primary criterion. When a primary criterion is selected using one of the radio buttons, two pairs of dashed-white lines pop up to show the interdependence with the other two criteria. This is a change from TOP-Slider where the interdependence lines were visible only during the displacement of a cursor. Like for TOP-Slider, the Pareto ranges are updated in a tightly-coupled manner with cursor displacement, but only for the cursor of a primary criterion.

Secondary criterion. Differing from TOP-Slider, when moving the cursor of a secondary criterion, the two pairs of dashed lines are now re-placed with a single dashed-white line resulting in a simplification of the visual cues. This line links the cursor of the secondary criterion to a white-filled circle that moves within the slider bar of the third criterion synchronously with the displacement of this cursor. The circle suggests an optimal choice for the third criterion, given the current choice for the primary criterion and the position of the secondary criterion.

Tight-coupling visualization. As with TOP-Slider, the Pareto ranges are updated in a tightly-coupled manner with cursor displacement. Differing from TOP-Slider with the introduction of these radio buttons, the Pareto ranges of the sliders are updated only for the cursor of a primary criterion and are kept unchanged for the other criteria. By doing so, we improve screen visual stability.

'Align on …' button. A contextual *'Align on …'*, green button replaces the contextual pop-menu of TOP-Slider that was used rarely in the qualitative experiment [15]. Clicking the green button would then move the linked cursor to the current position of the white circle. The green button, which appears only when the cursor of a secondary criterion is moved, has been introduced to facilitate the alignment of the cursors on the Pareto front while leaving the user free to explore non-optimal solutions.

P4DM: Pareto Slider for Surgery adapted. Like PSS, one distinctive feature of P4DM is the presence of a square bracket shape cursor, named “restrictor knob” whose position on a slider delimits the optimal from the undesired values for the corresponding criterion (cf. Fig. 2-right). By moving the restrictor of a slider, users can exclude values for the corresponding criterion. In addition, a cursor cannot be moved outside its white range: moving the round cursor of one slider moves the cursor of the other sliders automatically so that the new position of the cursors corresponds to a Pareto optimal solution.

In PSS, the strategy used for choosing the new Pareto optimal solution is decided by the designer of the algorithm, not by the user. In our re-implementation, we have reproduced the strategy described in [25]: a point on the Pareto front is selected so that the movement of the two untouched cursors is kept minimal.

4 Experimental User Study

This section presents the details of the experimental study conducted with 177 students to compare Sliders4DM with P4DM. In this experiment, objective quantitative data was logged automatically then processed to measure the objective and subjective metrics presented in section 3.

4.1 Apparatus

Using standard web browsers, both Sliders4DM and P4DM were available as web applications developed with JavaScript (client and server), SVG (visual rendering), node.js (storage of interaction traces, and participant authentication). Both user interfaces were designed with a minimal 900x560 pixels footprint. Therefore, the participants were asked to use a standard desktop computer with mouse for input, and connected to the Internet with regular communication speed (i.e. no tablet or smartphone device). Logs show an average resolution width of 1419 px (σ=168 px).

The code of the two interaction techniques was instrumented to collect mouse events where a log entry includes: a timestamp, an event type (motion, press, release) and the widget concerned (slider cursor, priority button, alignment button, restrictor knob), the slider index, and the cursor position (value normalized between 0 and 1). The log files, one per participant, were stored on a server in JSON format. Logged data was analyzed with Python scripts using the SciPy library.

4.2 Participants

The experiment was the third and last session of a larger experiment that involved 201 students over a two-month period. The subjects were told that they could earn up to 20€ for participating in the first two sessions and that they could earn a 5€ bonus if they achieved the task of the third session, the scope of this article. Students were told that payment would occur at the end of the third session. In addition, they did not know how much they had already earned in participating in the first two sessions before the end

of the third session. Therefore, the participants share the same profile, that is, students with limited financial resources: only 24 over 201 students chose not to participate in our experiment. It is thus reasonable to consider that (1) money was the motivation for the remaining 177 students; and (2) that financial cost is effectively the primary criterion for this experiment.

Table 1. Groups of participants: mean age and studies.

Group	# Part.	Mean age	Studies			
			Eco.	Lit.	Law	Sci.
S4DM	91 (50 m./42 f.)	21.3 (σ=2.1)	45 (49.4%)	16 (17.6%)	16 (17.6%)	14 (15.4%)
P4DM	86 (49 m/36 f.)	21.4 (σ=1.6)	36 (41.9%)	20 (23.2%)	17 (19.8%)	13 (15.1%)

Table 1 shows the distribution of the participants: 99 males and 78 females (average age: 21.35) studying economy and/or management (81), literature (36), law and/or politics (33), and sciences (27). We used a between-subjects approach with the interaction technique as the independent variable [12]. As in experimental economics [7, 17], we adopt a between-subjects (or between-groups) experimental design, so that each person is exposed to a single interaction technique. The main reasons for this choice are the following: first, this experimental design minimizes learning and transfer across experimental conditions. In a within-subjects design, the subjects are more knowledgeable about the domain after the first user interface's use, and that knowledge will likely help subjects to become more efficient on the second tested user interface. In our case, the learning effect is precisely in the course of our study. Secondly, between-subjects studies have shorter sessions than within-subject ones, which allows it to be less tiring or boring, and also more appropriate for remote non-moderated testing. Between-subjects experimental design requires care in the constitution of the two subject groups. The groups must meet homogeneity conditions to ensure that the assignment of subjects does not affect the results of the study. For this reason, in our experiment, subjects were randomly distributed into two groups, one for each interaction technique. We chose to recruit students as subjects because Step 1 indicated their sensitivity to financial cost for decision making. In addition, socio-economic diversity is less pronounced for this class of subjects making it easier to satisfy the requirements for similarity in characteristics between the two groups. In the following, we denote S4DM the group of participants that used Sliders4DM, and P4DM the group of participants that used P4DM.

4.3 Decision Task

Participants were asked to perform the following decision task: "*As a student with limited financial resources, you are asked to select the values for financial cost, air quality and thermal comfort that best suit you for your home. When you have found a combination that satisfies your objectives, please click the 'Validate' button*".

As shown in Fig. 2, financial cost ranged between 20 €/month and 200 €/month, air quality between excellent (400 ppm) and mediocre (1400 ppm), and thermal comfort between 17 ºC and 23ºC. The maximum and minimal values for air quality and thermal comfort were chosen in accordance with the outdoor conditions at the time of the experiment (i.e., early April in France).

4.4 Procedure

The first step of our experiment consisted of providing the participants with the necessary information displayed on their screen, including a detailed description of the interaction technique to be used (either S4DM or P4DM), color-coding schemes, tight-coupling of the sliders, and the task to achieve. For both P4DM and Sliders4DM, the sliders were displayed in the same order as follows: financial cost (top), air quality (middle), thermal comfort (bottom).

The participants were informed that: (1) the goal was to set the cursors on a position suitable for them; (2) the initial position of the sliders cursors corresponded to an arbitrary choice (i.e. minimal cost, bad air quality, and cold temperature); (3) there was no time limit to achieve the decision task but one trial only was taken into account; (4) they had to click the '*Validate*' button when satisfied with their choice; (5) validating was mandatory to record their choice and to earn the financial bonus; (6) all their actions were recorded automatically; and (7) the session would start in two days and would be available online for only 24 hours.

In the second step of the experiment, participants had to authenticate themselves using an identification number and a password in order to be able to interact with one of the two interaction techniques.

5 Results

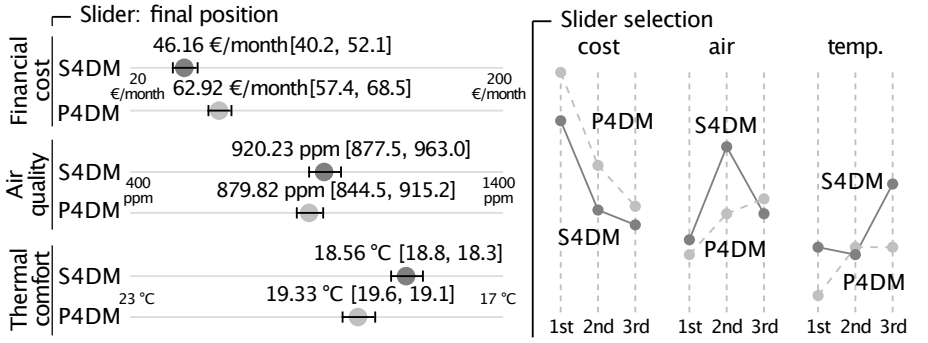

Fig. 3. Final cursor position for each slider denoting the choice of the decision task (left); The first three sliders used by the P4DM and S4DM groups (right).

This section reports and analyzes the data logs from the experiment. Interval estimation is used to interpret the inferential statistics [11]: we adopted the approach recommended in [11] based on the overlapping of the confidence intervals (disjoint, less than 25%, more than 25%) to assess practical evidence (respectively: strong, some, none). In the following, the graphs that report a mean value also display a 95% BCa bootstrap confidence interval (CI) [18], graphically and numerically (within square brackets). As well, as recommended by [11], effect size for mean difference (diff. = P4DM-S4DM) is reported as a 95% BCa confidence interval. In addition, we used the following color-coding scheme: dark grey for S4DM, and light grey for P4DM.

Multi-objective choice task. The results are shown in Fig. 3-left. Each of the three horizontal panels corresponds to a criterion: financial cost, air quality, and thermal comfort. An horizontal panel reports the mean final position of the cursor for both interaction techniques, representing the result of the decision task.

For financial cost, there is strong evidence that the P4DM group is willing to spend more money (62.92 €/month) than the S4DM group (46.16 €/month), by 36% (diff.=16.76 €/month, CI=[9, 24.4]). For thermal comfort, there is strong evidence that the P4DM group chose a more comfortable level for thermal comfort (19.33 ºC) than the S4DM group (18.56 ºC), by 8.4% (diff.=0.77 ºC, CI=[0.38, 1.13]). For air quality, with strong evidence, both groups chose a similar level of air quality between good (733 ppm) and average (1066 ppm), respectively ~920 ppm for the S4DM group, and ~879 ppm for the P4DM group (diff.=-40.41 ppm, CI=[-97.81, 13.59])

In short, we observe a strong correlation between financial cost and thermal comfort: a lower financial cost for the S4DM group, and a higher level of thermal comfort for the P4DM group.

Table 2. Statistics for the first three sliders used (including standardized residuals); residuals (dof = 2) are in bold if the value is greater than 1.96 (or less than -1.96).

	First use		Second use		Third use	
Group	S4DM	P4DM	S4DM	P4DM	S4DM	P4DM
Financial cost	52 (**-2.50**)	65 (**2.50**)	28 (**-2.10**)	40 (**2.10**)	24 (-1.26)	29 (1.26)
Air quality	20 (0.59)	16 (-0.59)	45 (**2.51**)	27 (**-2.51**)	27 (-1.11)	31 (1.11)
Thermal Com.	18 (**2.79**)	5 (**-2.79**)	16 (-0.53)	18 (0.53)	35 (**2.41**)	18 (**-2.41**)
χ^2 (p-value)	9.15 (0.01)		6.64 (0.036)		5.82 (0.054)	

The first three sliders used. The results are reported in Table 2 as well as in Fig. 3-right. Table 2 shows the numerical values used to generate the three graphs of Fig.3-right, one per slider, respectively from left to right: financial cost, air quality, and thermal comfort. Each graph represents the number of participants (vertical axis) using the related slider for their first three uses (horizontal axis).

In order to identify differences between the two groups, we applied a multivariate statistical test using 3x2 contingency tables based on the χ^2 probability law (dof=2). The bottom row of Table 2 reports the computed χ^2 value and p-value of the statistical test. For each count, Table 2 also reports the standardized residuals (dof=2). For the first use, with strong evidence, both groups use the slider related to financial cost. In addition, we observe that: (1) a higher number of users (u.) of P4DM (65 u.) used the financial cost slider first compared to S4DM (52 u.); (2) a very few number of P4DM users (5 u.) used the thermal comfort slider.

For the second use, with strong evidence, a majority of the S4DM group (45 u.) used the air quality slider while the majority of the P4DM group (40 u.) still used the financial cost slider.

For the third use, there is a small difference between the two groups ($p \sim 0.05$). Analyzing the use of the thermal comfort slider in details by computing a 2x2 contingency table where the data related to financial cost and air quality criteria are aggregated (35 vs. 51 for S4DM, 18 vs. 60 for P4DM), we observe with strong evidence (χ^2

= 5.0285, dof = 1, p-value = 0.025) that, as a third use, the S4DM group has used the thermal comfort slider more than the P4DM participants.

In summary, the "first three sliders used" patterns are the following:

- For S4DM: financial cost / air quality / thermal comfort.
- For P4DM: financial cost / financial cost / (air quality or financial cost).

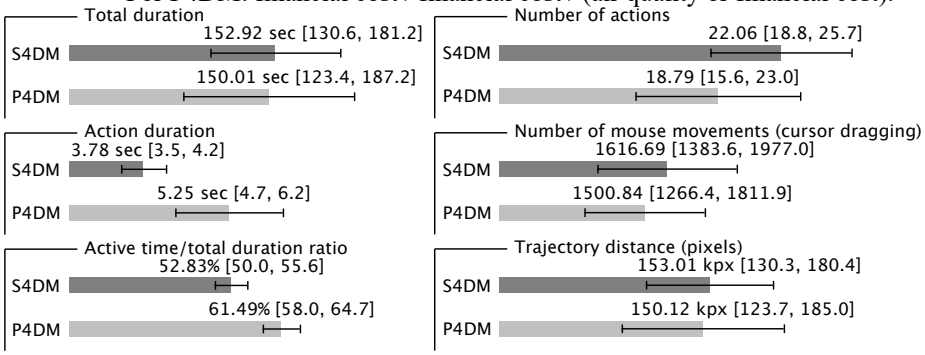

Fig. 4. Durations of the decision task and of actions (left); Interaction statistics (right).

Completion Time and Duration. The results are presented in Fig. 4-left. With strong evidence, both groups achieved the decision task by the same amount of time (Fig. 4, left-top panel): 2 minutes and 32.92 seconds for the SADM group, and 2 minutes and 30.01 seconds for the P4DM group (diff.=-2.91 sec, CI=[-40.0, 38.2]).

At a finer grain though, with strong evidence (Fig.6, left-middle panel), the S4DM group achieved atomic actions faster (3.78 seconds) than the P4DM group (5.25 seconds, diff.=1.473 sec, CI=[0.77, 2.35]). This result is based on measuring the time spent to achieve actions including the time used to drag cursors between two positions and the time to click a *'Priority'* or *'Align on ...*' button.

We calculated the active time ratio as the total of each action duration divided by the total duration of the task. With strong evidence (Fig. 4, left-bottom panel), the P4DM group (61.49%) spent more time to interact than the S4DM group (52.83%), by 8.7% (CI=[4.3,12.9]).

Interaction-workload. As shown in Fig. 4 (right-top panel), the S4DM group achieved the decision task within a mean number of ~22 actions while the P4DM group used a mean number of 18.78 actions, without significant difference (diff.=-3.26 actions, CI=[-8.23, 2.12]). Using raw mouse events, we considered the number of mouse movements for dragging cursors or restrictor knobs during the decision task (Fig. 4, right-middle panel). The S4DM group moved the mouse ~1616 times while the P4DM group moved the mouse ~1500 times, without significant difference (diff.=-116 times, CI=[-512, 275]).

We measured the length in pixels (1 Kpx=1000 pixels) of the trajectory followed by the mouse cursor. For this purpose, we considered a mean display width (1419 pixels) computed from the logs (see 4.1, Apparatus). For S4DM, a trajectory included the mouse movements to reach and click buttons (radio buttons to select a criteria priority and the alignment buttons). For both groups (Fig. 4, right-bottom), the distance is about 150 Kpx, without significant difference (diff.=-2.89 Kpx, CI=[-43.7, 35.7).

Use of Buttons and Restrictor Knobs. We computed how many times the S4DM buttons and the P4DM restrictor knobs were used, as well as the percentage of participants that used these widgets. We observe that some participants have not used one of the following widgets: 44% for the '*Priority*' radio buttons, 68.13% for the '*Align on…*' button, and 27,91% for the restrictor knobs.

A majority of the S4DM participants used the priority buttons (3 times in average), once (20.88%), or twice (13.19%) while the remaining 23% used the buttons from three to twelve times. The selected criteria for the priority buttons are: (1) air quality for the first use (37/51 participants); (2) financial cost for the second use (18/32 participants), over thermal comfort (10/32 participants); (3) air quality for the third use (12/21 participants and between 4-5 participants for the two other criteria). Similarly, most participants used the '*Align on…*' button once (10.99%) or twice (6.59%). The remaining 14% used this button from three to eleven times.

As for P4DM, the restrictor knobs were used from once to seven times by 47.67% of the participants. The remaining 24.42% used them from eight to thirty-eight times.

6 DISCUSSION

Based on the quantitative details presented above, we now analyze and interpret the results according to the 5 metrics specified in Section 3.1. We then summarize and generalize the main findings and point out the limits of this work.

6.1 Analysis

Decision-Accuracy. Using financial cost as an objective measure for decision-accuracy, we observe a significant difference between the two interaction techniques: Sliders4DM leads to a more optimal decision than P4DM as the S4DM group save 16.76 €/month. The P4DM group chose a more expensive option by ~36%. For Sliders4DM, these results are consistent with the preferences provided by the students involved in the qualitative experiment for TOP-Slider (i.e. low financial cost priming over thermal comfort). Moreover, air quality might be more important over thermal comfort as participants selected air quality twice as the primary criterion. This might explain why they chose a better level of air quality over thermal comfort.

Decision-Satisfaction. In terms of satisfaction, we may consider that Sliders4DM helps users to reach a suitable compromise faster than P4DM. As reported in Section 5 for the "three first sliders used", the P4DM group changed the value of the financial cost twice while the S4DM group did so only once. Furthermore, 2/3 of the P4DM group needed to manipulate the restrictor knob. We interpret this as an initial unsatisfactory choice for financial cost.

More specifically, Sliders4DM may help users to reach a satisfactory choice in an efficient manner as (1) each slider is used once at the beginning (cf. the 'financial cost-air quality-thermal comfort' pattern reported in section 5) and (2) only 1/3 of the S4DM participants used the '*Align on…*' button, meaning that often the cursors were already set at a suitable position.

Incentive-to-Explore. In terms of exploration, the "three first sliders used" patterns show differences between the two interaction techniques. Correlated to our previous observation, whereas the P4DM group achieved a corrective pattern 'financial cost-financial cost-air quality', the S4DM group used each slider once in a 'financial cost-air quality-thermal comfort' sequence. Moreover, 2/3 of the S4DM group adopted the 'air quality-financial cost-air quality' pattern for the priority criteria. We suspect that the tight coupling between cursor positions enforced by P4DM results in corrective patterns.

Time-on-Task. Although both groups spent a similar amount of time to achieve the decision task (~150 seconds), we observe differences at the action level: with Sliders4DM, participants' actions are clearly shorter than with P4DM (~30% less). This is correlated with a significant smaller ratio *(total active time) / (total duration)* for Sliders4DM. We hypothesize that Sliders4DM allows users to find a suitable compromise more quickly.

Interaction-Workload. We expected that interaction workload would be higher for Sliders4DM given that Sliders4DM includes three radio buttons and one contextual '*Align on…*' button displayed next to the sliders. In fact, both interaction techniques show very similar results. The presence of the radio buttons and the '*Align on…*' button did not increase trajectory lengths significantly. The total number of mouse movements to drag a slider cursor or to drag a restrictor knob is also similar for the two techniques. Consequently, both interaction techniques seem to impose a similar interaction workload.

6.2 Summary of the findings

Our comparative study indicates that Sliders4DM is more effective than P4DM in terms of decision-accuracy, decision-satisfaction, and incentive-to-explore. This may be due to the difference in the way the interdependence between the criteria is expressed in the two techniques. P4DM automatically positions cursors at optimal solutions, necessarily guiding users to safe choices at the risk of imposing non-preferred best solutions[1], and thereby restricting "what if" thinking. Sliders4DM encourages users to explore by decoupling cursor placements. In this way, users are free to position cursors in any range, whether it be optimal, non-optimal or even unfeasible. This freedom is necessary to accommodate situations in which users may not be looking for optimality, but for a satisfactory solution.

However, if cursor placements rely on human decision, then two conditions must be satisfied: (1) The implementation must provide *tightly-coupled visual feedback for cursor displacements* to express the interdependence between the criteria; and (2) The system must warn users when moving away from the Pareto front, for example, using color coding. These features can be complemented with *a system-suggested optimal solution* that users may choose to accept, for example, through a contextual speed-up button.

---

[1] Although they have not explicitly investigated this lack of freedom, Schuman et al. observed that the two surgeons recruited for their experiment preferred manual selection to using PSS [25].

The experiment has provided significant results with the aggregated logged data used to objectively measure key subjective metrics: decision-accuracy, decision-satisfaction, and incentive-to-explore. In particular, for decision-accuracy, we were able to observe a clear difference between the two techniques using a metrics based on financial cost and a metric based on the final position of the cursors. In addition, the "three first sliders used" pattern allowed us to assess decision-satisfaction as well as incentive-to-explore. For the latter, the observation of the use of the '*Priority*' radio buttons was considered in conjunction with the "three first sliders used" pattern to assess decision-satisfaction. This allowed us to observe clear differences between the two groups.

Compared to Dimara et al's methodology [9], our approach goes one step further as we were able to objectively measure decision-accuracy and decision-satisfaction without asking participants to self-report their confidence about the decision made. Besides, Dimaral et al. concede that self-reported confidence "can be subject to biases" [10]. Instead, we have adopted a pragmatic approach, drawing on the experimental context to map aggregated logged data as objective measures for subjective criteria, that is: (1) a controlled profile for participants (students with limited financial resources), (2) a clear primary criterion (financial resources), and (3) a limited number of interdependent criteria (financial cost, air quality, thermal comfort). Our approach, nevertheless, requires a preliminary user study to identify the primary criterion to be used for the participants' profile. For this, we relied on a preliminary study for documenting that financial cost is primary for students whereas thermal comfort is primary for elderly people.

The design choices made for a particular user interface dedicated to decision-making may introduce biases that influence the decision process (this has been demonstrated in the context of information visualization [10]). First, in order to minimize this side effect, we choose to reimplement TOP-Slider and PSS to share the same visual and feedback design, while respecting their interactional differences. Consequently, even if the final design might introduce biases, the results of our study show significant differences between the two interaction techniques. Second, to avoid framing, we chose to set the cursors on an ultimate but impossible best solution (best comfort at lowest cost).

6.3 Limitations and caveats

All the participants involved in the comparative experiment were students. This may have affected the results. In addition, the mapping we used between the objective logged data and the subjective metrics may have also influenced the results. In particular, we have assumed that financial cost was a primary criterion and thus used the choices made by the participants for financial cost to measure decision-accuracy. While this assumption is confirmed with strong evidence, it may not be valid when applied to participants with very different cultural backgrounds. Despite these limitations, although a pragmatic approach does not seek for generalizability, our approach opens perspectives to investigate these metrics further as well as to consider new subjective metrics to cover heterogeneous profiles and/or multiple primary criteria. At the methodological levels, it also requires to investigate how to design preliminary studies to build relevant profiles in relation to criteria.

7 Conclusion and Take-away Message

In this research, we have explored the problem of evaluating interaction techniques for multi-objective decision-making. These techniques are difficult to rigorously evaluate due to the subjective nature of decision-making as well as a lack of methodological guidance. Qualitative methods such as self-reporting, are commonly used for evaluating subjective metrics. However, self-reporting is known to be sensitive to cognitive bias. We believe that objective measures can bring additional rigor to the evaluation process.

In this article, we have shown how objective quantitative measures, aggregated from logged data, can be used to evaluate subjective metrics including decision accuracy, choice satisfaction, and incentive to explore. We have shown how these metrics can serve for conducting objective comparative experiments of interaction techniques for multi-objective decision-making tasks with a significant number of participants. We have selected two existing multi-slider interaction techniques as a case study and involved 177 participants.

We have proposed the following pragmatic considerations for defining the mapping of subjective metrics from objective aggregated data: (1) the characteristics shared by the target users, such as the "primary criterion for choice", (2) the key differentiating features of the interaction techniques under evaluation such as "suggestion vs enforcement of optimal solution", (3) the context of use such as energy management. We hope our approach will inspire researchers to extend these heuristics as methodological guidelines and principles for objectively evaluating and comparing interaction techniques designed for tasks that are inherently subjective.

Acknowledgements

This work has been partially sponsored by the French ANR, project INVOLVED reference ANR-14-CE22-0020, as well as by PIA project reference ANR-11-EQPX-0002, Amiqual4Home, and by Eco-SESA, a "Cross Disciplinary Program" project of the IDEX of Université Grenoble-Alpes, France, ANR project ANR-15-IDEX-02. We also thank Marie Cronfalt-Godet for her help for the organization of the experimental study.

References

1. Agrawal, G., Bloebaum, C., Lewis, K., Chugh, K., Huang, C.-H., Parashar, S.: Intuitive Visualization of Pareto Frontier for Multiobjective Optimization in n-Dimensional Performance Space. In: 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, 2004-4434. American Institute of Aeronautics and Astronautics (2004).
2. Akram Hassan K, Liu Y, Besançon L, Johansson J, Rönnberg N: A Study on Visual Representations for Active Plant Wall Data Analysis. Data 4(2), 74 (2019).
3. Alzhouri Alyafi, A., Nguyen, V.B., Laurillau, Y., Reignier, P, Ploix, S., Calvary, G., Coutaz, J., Pal, M., Guilbaud, J.-P.: From Usable to Incentive-Building Energy Management Systems. Modeling and Using Context, 2(1) (2018).

4. Bajracharya, S., Carenini, G., Chamberlain, B., Chen, K., Klein, D., Poole, D., Taheri, H., Öberg, G.: Interactive Visualization for Group Decision Analysis. International Journal of Information Technology & Decision Making, 17(6), pp. 1839–1864 (2018).
5. Booshehrian, M., Möller, T., Peterman, M., Munzner, T.: Vismon: Facilitating Analysis of Trade-Offs, Uncertainty, and Sensitivity In Fisheries Management Decision Making. Computer Graphics Forum, 31(3pt3), pp. 1235–1244 (2012).
6. Boukhelifa, N., Bezerianos, A., Trelea, C., Perrot, N., Lutton, E.: An Exploratory Study on Visual Exploration of Model Simulations by Multiple Types of Experts. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), pp. 1–14. ACM Press, Glasgow, Scotland UK (2019).
7. Charness, G., Gneezy, U., Kuhn, M.: Experimental methods: Between-subject and within-subject design. Journal of Economic Behavior & Organization, 81 (1), pp. 1-8 (2012).
8. Cialdini. R.: Influence. William Morrow and Company, New York (1984).
9. Dimara, E., Bezerianos, A., Dragicevic, P.: Conceptual and Methodological Issues in Evaluating Multidimensional Visualizations for Decision Support. IEEE Transactions on Visualization and Computer Graphics, 24(1), pp. 749–759 (2018).
10. Dimara, E., Franconeri, S., Plaisant, C., Bezerianos, A., Dragicevic, P.: A Task-based Taxonomy of Cognitive Biases for Information Visualization. IEEE Transactions on Visualization and Computer Graphics, 26(2), pp. 1413-1432 (2018).
11. Dragicevic, P.: Fair Statistical Communication in HCI. In: Robertson, J., Kaptein, M. (eds.) Modern Statistical Methods for HCI, pp. 291–330. Springer, Switzerland (2016).
12. Fréchette, G.R., Schotter, A.: Handbook of Experimental Economic Methodology. Oxford University Press, Oxford (2015).
13. Goldkuhl, G.: Pragmatism vs interpretivism in qualitative information systems research. European Journal of Information Systems 21(2), pp. 135–146 (2012).
14. Ibrahim, A., Rahnamayan, S., Martin, M.V., Deb, K.: 3D-RadVis: Visualization of Pareto front in many-objective optimization. In: IEEE Congress on Evolutionary Computation (CEC'16), pp 736-745. IEEE (2016).
15. Inselberg, A., Dimsdale, B.: Parallel Coordinates for Visualizing Multi-Dimensional Geometry. In: Kunii T.L. (eds) Computer Graphics 1987. pp. 25–44. Springer, Tokyo (1987).
16. Kelly L.M., Cordeiro M.: Three principles of pragmatism for research on organizational processes. Methodological Innovations, 13(2), 2059799120937242 (2020).
17. Keren G.B., Raaijmakers Jeroen G. W.: On between-subjects versus within-subjects comparisons in testing utility theory. Organizational Behavior and Human Decision Process, 41(2), pp. 233-241 (1988).
18. Kirby, K.N., Gerlanc, D.: BootES: An R package for bootstrap confidence intervals on effect sizes. Behavior Research Methods, 45(4), pp. 905–927 (2013).
19. Laurillau, Y., Nguyen, V.-B., Coutaz, J., Calvary, G., Mandran, N., Camara, F., Balzarini, R.: The TOP-slider for Multi-criteria Decision Making by Non-specialists. In: Proceedings of the 10th Nordic Conference on Human-Computer Interaction (NordiCHI'18), pp. 642–653. ACM, New York, NY, USA (2018).
20. Laurillau, Y., Nguyen, V.-B., Coutaz, J., Calvary, G. : Slider4DM and P4DM widgets. http://iihm.imag.fr/laurillau/S4DM/, last accessed 2021/06/01.
21. Miettinen, K.: Survey of methods to visualize alternatives in multiple criteria decision making problems. OR Spectrum, 36(1), pp. 3–37 (2014).
22. Milutinović, G., Ahonen-Jonnarth, U., Seipel, S., Brandt, S.A.: The impact of interactive visualization on trade-off-based geospatial decision-making. International Journal of Geographical Information Science, 33(10), pp. 2094–2123 (2019).

23. Monz, M., Küfer, K.H., Bortfeld, T.R., Thieke, C.: Pareto navigation—algorithmic foundation of interactive multi-criteria IMRT planning. Physics in Medicine and Biology, 53 (4), pp. 985–998 (2008).
24. Pajer, S., Streit, M., Torsney-Weir, T., Spechtenhauser, F., Möller, T., Piringer, H.: WeightLifter: Visual Weight Space Exploration for Multi-Criteria Decision Making. IEEE Transactions on Visualization and Computer Graphics, 23(1), pp. 611–620 (2017).
25. Schumann, C., Rieder, C., Haase, S., Teichert, K., Süss, P., Isfort, P., Bruners, P., Preusser, T.: Interactive multi-criteria planning for radiofrequency ablation. International Journal of Computer Assisted Radiology and Surgery, 10(6), pp. 879–889 (2015).
26. Sifer, M.: Filter co-ordinations for exploring multi-dimensional data. Journal of Visual Languages & Computing, 17(2), pp. 107–125 (2006).
27. Tweedie, L., Spence, R., Dawkes, H., Su, H.: Externalising Abstract Mathematical Models. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'96), p. 406–412. ACM, New York, NY, USA (1996).
28. Vallerio, M., Hufkens, J., Van Impe, J., Logist, F.: An interactive decision-support system for multi-objective optimization of nonlinear dynamic processes with uncertainty. Expert Systems with Applications, 42(21), pp. 7710–7731 (2015).
29. United Nations, www.un.org/sustainabledevelopment/energy/, last accessed 2021/05/11.
30. Velasquez, M., Hester, P.T.: An Analysis of Multi-Criteria Decision Making Methods. International Journal of Operations Research, 10(2), pp. 56–66 (2013).
31. Weng, D., Zhu, H., Bao, J., Zheng, Y., Zu, Y.: HomeFinder Revisited: Finding Ideal Homes with Reachability-Centric Multi-Criteria Decision Making. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), pp. 1–12. ACM Press, Montreal QC, Canada (2018).
32. Wittenburg, K., Lanning, T., Heinrichs, M., Stanton, M.: Parallel Bargrams for Consumer-based Information Exploration and Choice. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST'01), pp. 51–60. ACM, New York, NY, USA (2001).
33. Wu, B., Cai, W., Chen, H.: A model-based multi-objective optimization of energy consumption and thermal comfort for active chilled beam systems. Applied Energy, 287, 116531 (2021).