

Towards Rigorously Designed Preference Visualizations for Group Decision Making

Emily Hindalong*
Dialpad, Inc.

Jordon Johnson
University of British Columbia

Giuseppe Carenini†
University of British Columbia

Tamara Munzner
University of British Columbia

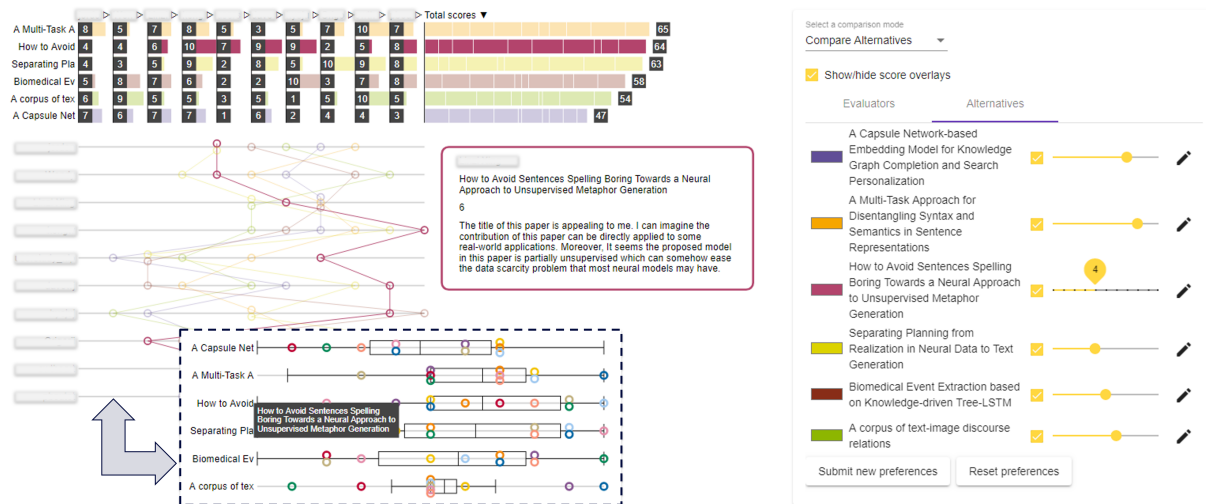


Figure 1: Our prototype, with idioms on the left and controls on the right. In Compare Alternatives mode, the parallel coordinates plot is rendered; switching to Compare Evaluators mode displays the box plot instead. Controls enable users to switch comparison mode, filter evaluators and alternatives, toggle overlays, and update preferences. Identifying information has been redacted.

ABSTRACT

Group decision making is when two or more individuals must collectively choose among a competing set of alternatives based on their individual preferences. In these situations, it can be helpful for decision makers to model and visually compare their preferences in order to better understand each others' points of view. Although a number of tools for preference modelling and inspection exist, none are based on detailed data and task models that capture the demands of group decision making in particular. This paper is a first step in addressing this gap. By going through the four stages of the nested model of visualization design, we have developed and tested a prototype to support group decision making when decision makers express their preferences directly on the alternatives.

Index Terms: Group Decision Making—Preference Models—Preference Inspection—Multiform Visualizations; Nested Model—Task Abstractions—Idiom Evaluation—User Studies

1 INTRODUCTION

Group decision making (GDM) situations arise when two or more individuals must collectively choose among a competing set of alternatives. These situations are ubiquitous in organizational and public planning [9]; examples include selecting an office location, hiring a candidate, or choosing a wastewater management system for a city.

Progress toward a satisfactory resolution is often impeded by differences in preferences [21] [29]. In some cases, group members

*Contributed while at UBC as an M.Sc. student.

†e-mail: carenini@cs.ubc.ca

may not even have a complete understanding of their own preferences. Organizations often elect to remain with the status quo because they do not have the resources to tackle these complexities [9] [14]. This inertia can stifle positive change.

One viable approach to GDM that is grounded in the literature is to have decision makers explicitly model their preferences over the alternatives. Then, the group members can compare how alternatives perform in order to reach a better understanding of other points of view. There is evidence that this approach can encourage reflection, promote transparency and inclusiveness, and ultimately lead to greater satisfaction with the outcome [2] [26].

The success of this endeavour depends on how quickly and effectively decision makers can glean insights from their own and others' preferences. It can be difficult to spot differences and trends when the data is represented in text-based formats such as spreadsheets. Visualizations show more promise because they leverage the pattern recognition and pre-attentive capabilities of the human visual system [24]. Note that such visualizations need not be intended to replace group discussion, but to facilitate it; understanding the rationale behind user preferences is essential to achieving consensus.

However, not all graphical methods are equally effective, and a poorly-chosen graphic can diminish the efficacy of the decision making process [1]. Unfortunately, there are few tools that support inspection of multiple people's preferences simultaneously, and none seem to be grounded in a comprehensive understanding of the needs (and corresponding tasks) of decision makers. Furthermore, there are currently no guidelines for designers of such tools. This paper is a first step in addressing this gap by focusing on relatively simple preference models in which decision makers express their preferences directly on the alternatives.

The stages of our work roughly correspond to the levels of the Nested Model of Visualization Design [23]. However, while the Nested Model is typically applied to design problems in a single

domain, we have elected to consider GDM (when alternatives are directly scored) as a whole in order to take advantage of the structural similarity of GDM problems across domains. At the *domain situation* level, from detailed summaries of seven GDM scenarios, we have identified high-level analysis goals, along with a set of situational features from which we have identified classes of decision makers. At the *data/task abstraction* level, we developed a data model based on multi-dimensional tables, as this abstraction is well-understood in the context of InfoVis [24]. To obtain the task model, we coalesced scenario-specific goals into a hierarchy of scenario-independent goals and subgoals. Then, for each subgoal, we enumerated high-level visualization tasks to support them. At the *encodings* level, we conduct a deep assessment of the pros and cons of core visual encodings for our target preference model with respect to the identified tasks. Then, we discuss options for combining encodings into a composite view and offer specific recommendations for different user classes. To obtain preliminary evidence of utility for our recommendations, we have implemented a prototype (*algorithm* level) and performed two user studies. The results suggest that the recommended idioms and interactions are effective, resulting in more participatory discussions, increased information exchange, and more easily identified areas of (dis)agreement.

2 BACKGROUND AND RELATED WORK

We provide some background on GDM, describe existing tools for visualizing preferences and related data, and discuss relevant visualization task analyses.

2.1 Group Decision Making (GDM)

GDM occurs when multiple stakeholders with possibly conflicting preferences need to make a choice together. One consistent finding in decision analysis is that formal methods can improve decision making [10]. When these are applied to GDM they typically involve at least the following steps (e.g. [34]): (1) context clarification and identification of group members; (2) generation of alternatives; (3) elicitation of preferences from stakeholders; and (4) synthesis and communication of decision recommendations.

Our work is primarily concerned with Step 4, when decision makers' preferences are combined to make a final decision. For simplicity, decision makers might sum the individual scores and select the alternative with the highest total score, but this would rob them of the opportunity to reconcile their differences and learn from other points of view. Ideally, they would examine and discuss the totals in the context of individual scores and preference models, an activity we call **preference inspection**. Note that the decision making process is often iterative: decision makers may revisit earlier steps to refine the decision model as they develop a deeper understanding of the problem, so preference inspection activities such as those supported by our prototype are critical.

Several preference models have been proposed in the decision analysis literature, with increasing complexity: from decision makers simply ranking alternatives, to giving a score to each alternative, to scoring alternatives on a set of attributes that could be weighted differently [17]. In this paper, we focus on preference models in which alternatives are scored directly, which represents a solid foundation for performing similar work for more complex models.

In terms of user tasks, the decision analysis literature does not provide the level of detail needed for designing visualizations; so we have performed an in-depth survey of seven GDM scenarios (in the spirit of a recent examination of an undergraduate admissions process [37]), identified a hierarchy of common analysis goals, and decomposed them into visualization tasks that guided our design.

2.2 Visual and Interactive Tools

Although tools to support decision making are plentiful, not all of them are able to integrate and display multiple people's preferences

simultaneously [27] [34]; some applications to GDM involve joint construction of only a single preference model by all members of the group [34]. There are tools that allow multiple users to input their preferences, such as M-MACBETH [3], D-Sight [11], and 1000Minds [15]; but they typically only show aggregated preferences using non-interactive charts and tables. Another tool, Web-HIPRE [25], uses only stacked bar charts, which are not effective for many of the GDM tasks identified in this work. Web ValueCharts [22] is derived from a task model for individual decision making [4]; however, its generalization to GDM was not investigated. Finally, ConsensUs [20] only allows users to compare their own preferences to the group average or one other user at a time; it does not support informative comparisons among any desired subset of evaluators.

In addition, there are tools for exploring data similar to preferences. For instance, LineUp supports comparison of ranked entities across attributes [13]; SurveyVisualizer [8] supports exploration of hierarchical multi-attribute survey data; and QStack ranks search results from multi-tag datasets based on tag frequency [30]. Further, since GDM is an inherently collaborative process, the field of collaborative visualizations [18] is closely related to our work; however, in these tools GDM is often supported only to achieve a particular shared task, such as selecting antibiotic prescriptions [19], rather than to support GDM in general as we do here.

Tools for visualizing preferences or similarly structured information tend to use a restricted set of encodings, as summarized in Table 1 for the most popular systems. However, unfortunately, the choice and integration of such encodings in each tool has not been guided by detailed data and task models, informed by a systematic study of the needs of decision makers; this is what we aim to do in this paper.

2.3 Task Analyses

To our knowledge, there have been no attempts to formally model the tasks for preference inspection in a GDM context.

Bautista and Carenini developed a task model for individual decision making [4]. This model divides the decision process into three phases: model construction, inspection, and sensitivity analysis. Each phase is decomposed into a task hierarchy using terms and concepts from information visualization and decision theory.

In a similar vein, Rempel and Young conducted a comprehensive task analysis for portfolio construction, a decision problem that involves iterative refinement of investment portfolios [32]. They open-coded relevant literature and field notes, coalesced the codes into themes, and described their codes in terms of visualization tasks. Their analysis is specific to portfolio construction, lacking the generality we aim for in this paper. Furthermore, they deal with decision scenarios in which the alternatives are potentially infinite, while we consider the more common finite setting.

3 FROM DOMAINS TO DATA AND TASK ABSTRACTIONS

In this section we describe the process by which we synthesize data and task abstractions from our selected GDM scenarios. Goals and tasks identified during this process that apply to more complex preference models will be similarly treated in future work.

3.1 Scenario Characterization

In order to identify a hierarchy of analysis goals for GDM, we surveyed seven scenarios covering as much breadth as possible within our time constraints. For four of the scenarios, we interviewed decision makers after the decision process, using a script with specific questions about problem formulation in decision analysis terms and open-ended questions about the process and analysis goals. Some interviews were conducted by the first author; others were conducted by colleagues. The script was adapted to each scenario and refined as data was collected. The remaining scenarios were described in published sources (journal papers or webinars). Table 2 provides a brief description of each scenario and the elicitation method used.

Table 1: Some well-known techniques for visualizing preferences or similar data.

	Context	Main Encodings
ConsensUs [20]	Group Decision Making	Dot plots in small multiples
Group ValueCharts [2]	Group Decision Making	Stacked bar chart; tabular bar chart
LineUp [13]	Multi-attribute rankings	Slope graph; stacked or tabular bar chart
QStack [30]	Multi-attribute rankings	Stacked bar chart
Survey Visualizer [8]	Multi-attribute survey results	Parallel coordinates tree
Web-HIPRE [25]	Group Decision Making	Stacked bar chart

Table 2: Descriptions of the seven group decision making scenarios.

Name	Description	Elicitation Method	Alternatives	Evaluators
BP: Best Paper	Researchers choose Best Paper Award recipient	Interview	4 - 15	5
FH: Faculty Hiring	Faculty members choose which candidate to hire	Interview	1 - 4	50 - 100
CR: Campbell River	Stakeholders from diverse interest groups choose a watershed operation strategy	Webinar observation	6	15
VY: Voyager [12]	Scientists choose trajectories for Voyager 1 & 2	Journal Paper	32	10
NC: Nuclear Crisis [27]	Emergency planners choose a strategy in response to a mock nuclear crisis	Journal Paper	6	6
SW: Software	Software company employees choose a technology stack	Interview + In-person observation	2	5
GI: Gift	Lab members choose a gift for a colleague	Interview	3	10

The first author produced summaries identifying the following:

1. the decision problem and decision-making process
2. the data, along with how preferences are modeled
3. the preference inspection goals and how they are achieved
4. other interesting features of the situation

For the first point, we summarized the steps taken by the decision makers. For the second point, we identified aspects of the problem (such as alternatives and evaluators) and described how preferences were modeled (if at all). For the third point, we open coded the sources, using ‘goal’ for any desired result (e.g. consensus) or qualitative query (e.g. identify best alternative) and ‘task’ for any process (e.g. discussion) or quantitative query (e.g. identify tallest bar) used to achieve a goal. For the final point, we identified interesting situational variables by looking for explicitly specified features (e.g. How much time is devoted to preference inspection?) as well as implicit features (e.g. How high are the stakes?). Five situational variables were identified, and discrete levels were defined based on the ranges covered by the scenarios. The levels for implicit features were decided based on the authors’ judgment. The results for the *Best Paper* scenario are provided in the supplemental material.

3.2 Situational Analysis

By examining the situational features identified in the previous step (see rows in Table 3), we were able to sort the scenarios into groups, which we call user classes (see columns in Table 3). The **Specialized** class covers very high-stakes, one-time decision problems where the decision makers devote one or more days to analyzing with the help of experts (i.e., consultant with knowledge of formal decision making methods). The **Professional** class covers high and medium-stakes decision problems that recur annually or monthly where the decision makers devote only a few hours to analysis. The **Casual** class covers low-stakes decisions made in a more casual setting.

Each class is likely to have different needs for a support system. Decision makers in the first class may benefit most from advanced analytic features and visual encodings, since they have the time, incentive, and expertise to take advantage of them. Decision makers in the second class are more likely to benefit from systems that are easy to learn and deliver insights quickly; if a system is visually too complex or cumbersome, they may not be willing to put in the effort to learn and use it. The third class may have an even greater

Table 3: Summary of situational features of seven scenarios.

Type	Specialized	Professional	Casual
Scenarios	CR, VY, NC	BP, FH, SW	GI
Stakes	Very High	Medium – High	Low
Work Context	Expert Assistance	Professional	Casual
Timeframe	Days	Hours	Hours
Decision Frequency	Once	Monthly – Annually	Once

preference for usability over sophistication, and in some cases may not derive sufficient benefit from support systems to use them at all.

3.3 Data Abstraction

Since we focus on preference models in which the evaluators express preferences directly on the alternatives, our data model may include:

- A set of Alternatives $A : |A| \geq 1$
- A set of Evaluators $E : |E| \geq 2$
- A function $score(e, a)$ that returns the score given by evaluator e to alternative a (e.g. an integer in $[1, 10]$)
- A function $rank(e, a)$ that returns the rank given by evaluator e to alternative a (with ties assigned the average of the spanned ranks). Note this can be derived from $score(e, a)$.

To more clearly combine our data model with the visualization tasks identified in the next section, we mapped it to a model based on multi-dimensional tables, as this abstraction is well-understood in the context of information visualization [24]; sets became dimensions, and functions became measures. We obtained the dimensions *Alternative* and *Evaluator* and the measures *AltScore* and *AltRank*; we also derived the measures *TotalScore* and *TotalRank* to represent aggregated scores/ranks over all evaluators.

3.4 Task Abstraction

The development of task abstractions began with a goals analysis step, in which we coalesced scenario-specific goals into a hierarchy

of scenario-independent goals. Next, we enumerated high-level visualization tasks to support each subgoal. The task names are drawn from Brehmer and Munzner’s typology [7], and the task inputs and outputs are entities defined by the data model. We did so by casting scenario-specific goals and supporting activities as tasks (if more specific than the goal in question) and conducting additional analysis to discover further tasks that could support the goal in question. Finally, we decomposed the high-level tasks into sequences of low-level tasks, again from Brehmer and Munzner’s typology. Because many of these tasks consist of the same operations on different inputs, we defined them in terms of **auxiliary task (AT)** functions, whose inputs are entities defined by the data model.

3.4.1 Goal Analysis

Table 4 lists the goals and subgoals resulting from our analysis along with the scenarios in which each of them appeared. In all scenarios, the overarching goal was to achieve consensus or make a well-informed decision that most stakeholders could accept. This goal was primarily achieved through discussion, with quantitative summaries serving as a guide. A crucial point is that in no scenario did the quantitative summaries completely supplant verbal exchange; rather, their role was to focus analysis on points of interest, which can greatly enhance the efficiency of the process. In particular, decision makers used the quantitative summaries for the first three goals, which pertain to understanding the model. An additional goal is to *validate* the model (G4). It is not uncommon for evaluators to adjust their preferences (Section 2.1, step 3) after preference inspection [34]. In some cases, it may also be necessary for the decision makers to revise the alternatives (Section 2.1, step 2). Quantitative models are seldom sufficient to fully capture individual preferences, so a final goal is to discover nuances that the models do not capture (G5).

3.4.2 Task Analysis

After identifying the goals in Table 4, we related them to abstract visualization tasks from Brehmer and Munzner’s typology: the high-level family of *Discover* tasks and the lower-level families of *Query* and *Search* tasks [7]. We provide the full analysis for goal *G1: Discover Viable Alternatives* here; the supplemental material contains similar analyses for goals G2 - G4. We exclude G5 since it does not involve visual inspection of preference models.

High-level tasks to support G1 are shown in Table 5. All of these tasks and the goals they support are instances of the high-level task *Consume: Discover*, which covers many facets of inquiry [7].

Supporting tasks for G1a include finding alternatives with high overall scores or low variation in scores across evaluators, as these may constitute viable ‘compromise’ alternatives. Another way to focus the analysis is to identify non-dominated¹ alternatives. Excluding dominated alternatives can help minimize distraction. To narrow the list further, it is essential to be able to consider trade-offs between competitive alternatives. Finally, it may be necessary to look at pros and cons of one alternative (where the ‘pros’ are evaluators with high scores and the ‘cons’ are evaluators with low scores), especially if selecting no alternatives is an option. G1b is concerned with identifying high-performing alternatives for a particular evaluator.

We decomposed the high-level *Discover* tasks into low-level *Search* and *Query* auxiliary tasks on various inputs. This fine-grained analysis allows potential designs to be evaluated more efficiently. All the high-level tasks can be accomplished using a combination of just ten AT functions; Table 6 shows the complete list.

The AT functions and their inputs allows us to precisely describe how to accomplish the high-level tasks. For example, Task *T4: Discover trade-offs in AltRanks/AltScores between alternatives a and b* can be accomplished for AltScores as follows:

1. $AT6(AltScores(a)) \rightarrow X$ (Locate AltScores for a)
2. $AT6(AltScores(b)) \rightarrow Y$ (Locate AltScores for b)
3. $AT3(X,Y) \rightarrow Z$ (Get differences between distributions)

Similar sequences can be used to achieve all the high-level tasks, as shown in the supplemental material.

4 ENCODINGS ANALYSIS

In the context of our data and task abstractions, we conduct an in-depth assessment of the pros and cons of core visual encodings. Then, we discuss options for combining encodings into composite views and offer specific recommendations for different user classes. Our analysis focuses on the case in which alternatives are scored directly. The ranking case is simpler (ranks can be derived from scores, but not vice versa), and similar encodings would be suitable.

With respect to scalability, when the number of alternatives/evaluators is less than a dozen, they can be differentiated using colour alone [24]. Otherwise, colour should only be used for highlighting subsets of the data.

4.1 Single-view Analysis - Core Idioms

This section discusses the major competitive idioms for presenting small-scale tabular data with categorical keys and quantitative values. To reiterate, the data abstraction when evaluators are scoring alternatives directly is a two-dimensional table with Evaluators and Alternatives as keys and AltScores as values. TotalScores are obtained by summing AltScores over Evaluators. All but two of the idioms we discuss encode values using position on a common scale, which is the most effective channel for encoding magnitude [24]. Although non-radial designs can be oriented horizontally or vertically, our figures show only the horizontal orientation for succinctness.

4.1.1 Bar-based Idioms

Three styles of bar charts are suitable for two-dimensional tabular data: stacked bar charts, multi-bar charts, and tabular bar charts [13].

Stacked Bar Charts are appropriate when a one-dimensional measure is the sum of a two-dimensional measure, as is the case with TotalScores [13] [24]. To improve discriminability, segments within stacked bars can be assigned different color hues [24]. Stacked bar charts are not effective for tasks that require comparison of AltScores [24] [36]; however, they are effective at supporting TotalScore comparisons while providing information about the relative contribution of each AltScore.

Multi-bar Charts map spatial regions to dimensions in a nested fashion such that all bars are aligned to a common baseline. Additionally, color hue may be mapped to the secondary grouping to facilitate comparison across regions. Figure 2 includes two possible designs given the available mappings from spatial region and color hue to Evaluators and Alternatives.

Tabular Bar Charts map dimensions to spatial regions in a grid. There are four possible designs given the available mappings from spatial region and color hue to evaluators and alternatives. Figure 2 shows two variations that map colour to the column dimension. Note that tabular bar charts pair nicely with stacked bar charts, facilitating transitions between comparing TotalScores, inspecting the breakdown of TotalScores into AltScores, and comparing AltScores for a particular evaluator. Tabular bar charts are more compact than multi-bar charts, but they are also less precise because the same axis range is compressed and repeated across columns. Additionally, making comparisons across tabular columns is less accurate than comparisons across regions in multi-bar charts [36].

4.1.2 Point-based Idioms

Strip Plots use position along a common axis to encode values (Figure 2). Two-dimensional tabular data can be represented as a series of strip plots with one dimension separated by region (each

¹Alternative a **dominates** alternative b if at least one evaluator prefers a to b and all others either prefer a or are indifferent between the two.

Table 4: Goals for preference inspection and the scenarios in which they appear. BP=Best Paper, FH=Faculty Hiring, CR=Campbell River, VY=Voyager, NC=Nuclear Crisis, SW=Software, GI=Gift

GENERIC GOAL		SCENARIOS
G1	Discover Viable Alternatives	
	a Discover high-performing alternatives across evaluators/evaluator groups	<i>all but</i> FH
	b Discover high-performing alternatives for a single evaluator/evaluator group	SW, GI
G2	Discover Sources of Disagreement (discrepancies across evaluators)	
	a Discover and explain disagreement about an alternative (across evaluators/evaluator groups)	<i>all but</i> GI
G3	Explain Individual Scores	
	a Analyze contribution of different evaluators to an alternative’s total score	NC
G4	Validate Model	
	a Understand sensitivity of preferences to change	NC
	b Understand sensitivity of total scores to aggregation method	VY
	c Discover discrepancies between the preferences of one evaluator and the rest	BP
G5	Discover Nuances (not captured by preference models)	<i>all</i>

Table 5: Tasks to Support G1: Discover Viable Alternatives.

TASK	
G1a. Discover high-performing alternatives across evaluators	
T1	Discover alternative(s) with best TotalRank/TotalScore
T2	Discover alternative(s) with low variance in AltRanks/AltScores across evaluators
T3	Discover non-dominated alternatives across evaluators
T4	Discover trade-offs in AltRanks/AltScores between alternatives <i>a</i> and <i>b</i>
T5	Discover pros and cons in AltRanks/AltScores for alternative <i>a</i>
G1b. Discover high-performing alternatives for a single evaluator	
T6	Discover alternative(s) with best AltRank/AltScore for evaluator <i>e</i>

with its own strip plot) and the other distinguished using another channel, typically colour hue (though another option is mark shape, sometimes used in concert with hue [33]).

The key strength of strip plots relative to bar charts is that they place an entire dimension on a single axis, uniting the precision of multi-bar charts with the compactness of tabular bar charts. This also makes them superior to bar charts for tasks related to spread, such as identifying clusters and outliers, since the user can scan a single spatial dimension to obtain all relevant information. However, strip plots are less effective than bar charts at supporting look-up tasks because the secondary dimension is often differentiated using colour alone. The use of colour also limits scalability, since people can only differentiate up to around a dozen hues [24]. Their efficacy is contingent on the quality of the palette, which should be highly discriminable and accessible to people with colour-blindness [24].

Another challenge associated with strip plots is that occlusion may occur if two or more points have the same (or nearly the same) value. Perhaps the most scalable way to address this challenge is a combination of stacking and fill removal, which means plotting multiple unfilled points in a vertical ‘stack’ at the same x-coordinate.

Strip plots can be augmented to further support comparison of distributions. First, each axis can be overlaid with distribution information to create range, box, or violin plots. For succinctness, we only consider box plots. Alternatively, points corresponding to items in the secondary dimension can be connected with straight lines of the same colour; the resulting *parallel coordinates* plots can be effective for inspection and comparison across axes [24].

A variation on parallel coordinates is the *radar chart*, which is

less effective for comparison of values across axes since the axes are not aligned. Furthermore, the layout may be misleading if the data is not cyclic [24]. Other problems with radar charts arise in cases where one or more values approach zero (see Bob in Radar Chart 2 of Figure 2). One benefit of radar charts is that the area of a polygon is roughly proportional to the squared sum of the axis scores. This means that Radar Chart 2 in Figure 2 roughly encodes TotalScores.

4.1.3 Heatmaps

Heatmaps are arranged as a grid like tabular bar charts, except it does not matter which dimension is assigned to which axis because the color mark used to encode values has no orientation.

The main strength of heatmaps is information density, which makes them suitable for providing high-level summaries of large datasets at a glance [24]. They can also reveal interesting trends if the rows and columns are arranged in a meaningful way [24].

There are two popular types of color scales for heatmaps: rainbows, which progress from one hue to another, and sequential scales, which vary the luminance or saturation of a single hue. Most experts recommend the latter [16] [5], so we adopt this choice.

4.2 Single-view Analysis - Task-based Evaluation

We performed an analytic evaluation of the suitability of each encoding for each auxiliary task function identified in Section 3.4.2. The evaluation was based on a combination of established principles and the experience of the authors. Tasks that apply to more than one type of input were split into cases. The evaluation details are provided in the supplemental material. Figure 4 summarizes the results.

A key takeaway is that the two most complementary encodings that are in the top half of the table (i.e., with high total score) are Parallel Coordinates 2 and Tabular Bar Chart 1. This pair of independently strong encodings achieves a combined score of 33 out of a possible 45, which is the score achieved if the best encoding for each task were selected. This finding suggests that these two encodings can be used in conjunction to support most tasks.

Another observation is that parallel coordinates dominate radar charts. In light of this finding and the problems with radar charts discussed earlier, we eliminate them from further consideration.

4.3 Multiple-view Analysis

In this section, we offer recommendations on how different views can be integrated into a single window to support most tasks.

4.3.1 Number and Arrangement of Views

It is clear from Figure 4 that no single encoding is sufficient to support all tasks. For this reason, our recommendations employ

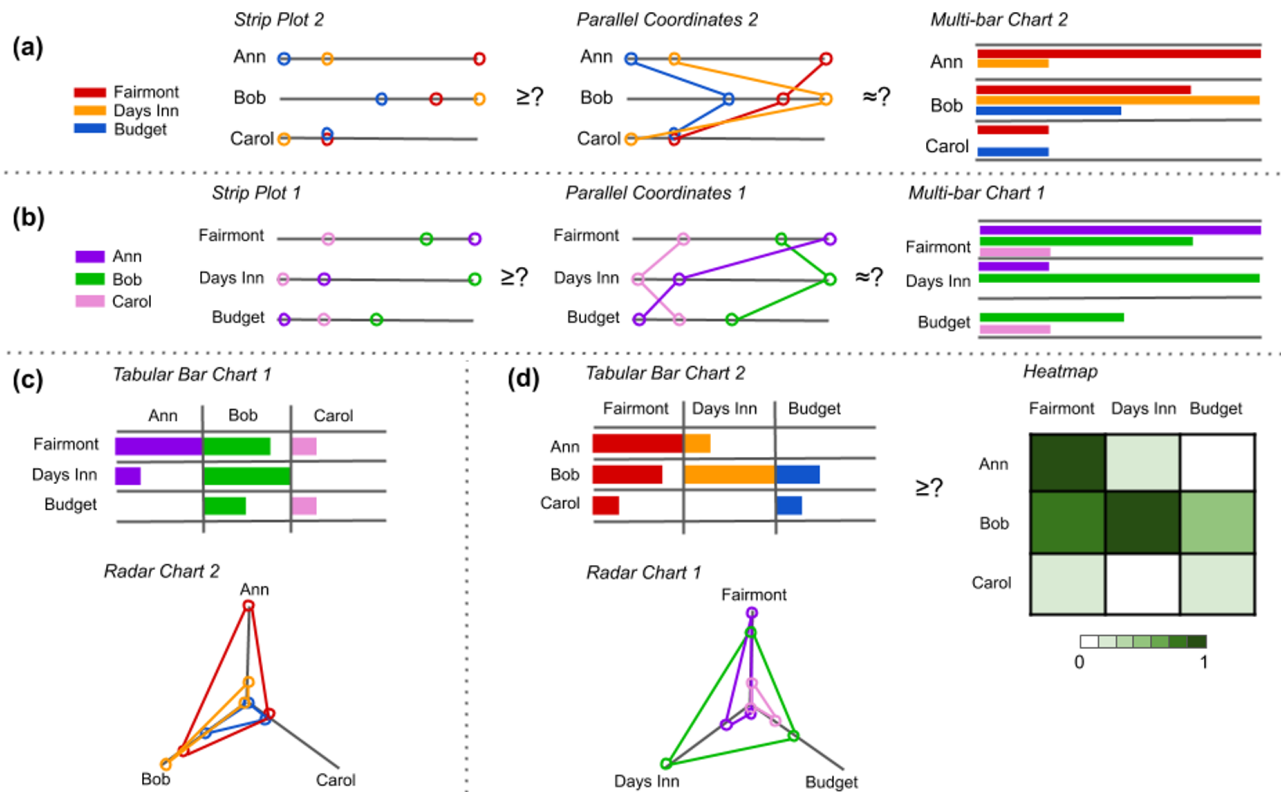


Figure 2: What is the best encoding for comparing Fairmont and Budget for Bob? The encodings are divided into four efficacy groups according to key principles. (a) Highly effective - comparisons are performed along a single axis or within single region; (b) Less effective - comparisons are made across axes or regions; (c) Less effective - axes are condensed and offer less precision; (d) Least effective - requires comparison of unaligned widths, unaligned positions, or colors. The rankings within groups are nuanced, as discussed in the supplemental material.

the *multiform* design choice, in which the same data is faceted into two views that use different encodings [24]. If the intended platform is a desktop or laptop computer, we recommend splitting the window horizontally and populating each half with encodings in the horizontal orientation to maximize precision.

There is a cost associated with multiple views, both in terms of cognitive load and screen real estate [38], so we do not advise supporting more than two views. As the next few sections will demonstrate, it is possible to strongly support all tasks using combinations of just two encodings and a few basic interactions.

We present recommendations for both **Casual** and **Advanced** users, with the latter including the Specialized and Professional classes from Section 3.2, since their needs are likely to be similar for cases where alternatives are scored directly.

4.3.2 Recommendations for Casual Users

Our recommendations for casual users strike a balance between sophistication and accessibility.

Tabular Bar Chart 1 + Stacked Bar Chart + Box Plot 1. In order to identify potentially strong combinations for a dual-view design, we computed a score for each pair of encodings by taking the sum of the maximum score on each task. The parallel coordinate designs were excluded from consideration due to the learning curve and lack of familiarity for most people [24] [28]; including unfamiliar idioms may confuse casual users, making them less likely to use the tool.

Of the pairs that were included, the top scoring combinations were: Tabular Bar Chart 1 and Box Plot 1; Tabular Bar Chart 1 and Strip Plot 2; Box Plot 1 and Strip Plot 2; and Box Plot 1 and Multi-bar Chart 2. Of these, only the first uses the same colour map-

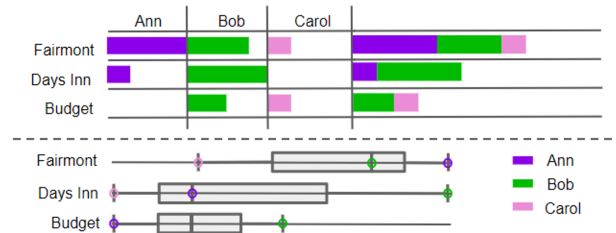


Figure 3: Recommendation for Casual Users with Tabular Bar Chart 1 + Stacked Bar Chart (top view) and Box Plot 1 (bottom view).

ping in both encodings, which is desirable because it preserves the semantics of colour across views [31]. The remaining pairs would require two distinct colour palettes; otherwise, they risk implying connections between unrelated marks [31]. Introducing a second colour palette limits scalability to about a dozen entities (alternatives and evaluators) *in total*. Thus, we recommend the first pairing (i.e. Tabular Bar Chart 1 and Box Plot 1) above all others.

Figure 3 shows a dual view based on this pairing, augmented with a Stacked Bar Chart. This design strongly supports all tasks except AT9:A (*Search: Browse for min/max values for one evaluator across alternatives*), which is weakly supported by the Tabular Bar Chart; this weakness can be mitigated by including sort functionality and text labels for bar values. One drawback of multiform designs is that users might become confused shifting attention between views, since they use different idioms and the axes do not correspond.

Table 6: Auxiliary Task Functions. The Input column also indicates when there are different cases of the task.

	Action	Input	Output	Supported by
AT1	Query: Identify	single value or distribution	its key-set	
AT2	Query: Compare	pair of values A: one evaluator, two alternatives B: one alternative, two evaluators	difference	
AT3	Query: Compare	pair of same-type distributions A: all evaluators, two alternatives B: all alternatives, two evaluators	tuple of differences	AT2
AT4	Query: Compare	pair of same-type distributions	dominance relation	AT3
AT5	Query: Summarize	single distribution	summary of variance	
AT6	Search: Locate	key-set A: one alternative, one evaluator B: one alternative C: one evaluator	single value or distribution	
AT7	Search: Lookup (in context)	key-set + single value or distribution	single value or distribution	AT6
AT8	Search: Browse	single distribution	outliers	AT2
AT9	Search: Browse	single distribution A: one evaluator B: all data	min/max values	AT2
AT10	Search: Browse	set of distributions	non-dominated distributions	AT4

	AT1	AT2:A	AT2:B	AT3:A	AT3:B	AT4	AT5	AT6:A	AT6:B	AT6:C	AT7	AT8	AT9:A	AT9:B	AT10	Total
Parallel Coords 2	1	2	1	3	2	3	1	1	1	2	n/a	1	3	3	3	27
Parallel Coords 1	1	1	2	2	3	2	2	1	2	0	n/a	2	1	1	2	22
Radar Chart 2	1	1	0	2	0	3	1	1	1	2	1	1	2	2	3	21
Multi-bar 2	2	2	1	1	1	1	1	3	1	2	n/a	1	2	2	1	21
Multi-bar 1	2	1	2	1	1	1	1	3	2	1	n/a	2	1	1	1	20
Tabular Bar 1	3	1	0	2	0	2	0	3	2	3	n/a	0	1	1	2	20
Box Plot 1	1	1	2	1	1	2	3	1	2	0	n/a	3	0	0	2	19
Strip Plot 2	1	3	1	1	1	1	1	1	0	2	n/a	0	3	3	1	19
Box Plot 2	1	2	1	1	1	1	1	1	0	2	n/a	0	3	3	1	18
Heatmap	3	0	0	0	0	1	0	3	2	2	n/a	2	2	2	1	18
Strip Plot 1	1	1	3	1	1	2	2	1	2	0	n/a	2	0	0	2	18
Tabular Bar 2	3	0	1	0	2	0	1	3	3	2	n/a	1	0	0	0	16
Radar Chart 1	1	0	1	0	2	0	2	1	2	2	n/a	2	1	1	0	15
Stacked Bar Chart	1	0	0	0	0	0	0	1	1	2	3	0	0	3	0	11

Figure 4: Support for each auxiliary task function (see Table 6 above for full descriptions) by encoding. 3=best, 2=strongly effective, 1=weakly effective, 0=ineffective. *n/a* indicates the encoding is not applicable. The *Total* column contains the row totals. The rows are sorted by Total.

4.3.3 Recommendations for Advanced Users

The space of viable options here is somewhat larger than for casual users, since designers may want to provide more or less flexibility depending on the work context and expertise of potential users. Even so, our recommended design strongly supports all tasks while maintaining consistency and discriminability between views.

Comparison Mode Switch. A careful analysis of Figure 4 indicates that all tasks are strongly supported by at least one of the following: Stacked Bar Chart, Tabular Bar Chart 1, Box Plot 1, and Parallel Coordinates 2. Overall, these encodings achieve a combined score of 41 out of a possible 45, which is the score achieved if the best encoding for each were selected. A design that can effectively integrate these four encodings is presented in (Figure 5). When the user selects *Compare Evaluators* mode, color is mapped to evaluators and the design is identical to the one for causal users. However, when the *Compare Alternatives* mode is selected, colour is mapped to alternatives, and the strengths of Tabular bar charts, Stacked Bar charts and Parallel Coordinates are synergistically combined.

4.3.4 Interactivity

This section describes how various options for transforming the view support users in performing auxiliary tasks identified in this work.

Users should be able to *sort* alternatives by TotalScore or by AltScore for a particular evaluator, as this ordering is essential for inspecting top values (AT9) and identifying dominance relationships (AT4). The ability to sort columns by AltScore for a particular alternative is not essential but could be useful. Ideally, users should also be able to *reorder* point plots, rows, and columns manually, as giving them control over which items are adjacent can improve their performance on comparison tasks (AT2, AT3, and AT4). Whenever the columns in the tabular bar chart are reordered, the segments in the corresponding stacked bar chart should be reordered as well.

Linked highlighting of marks for a particular evaluator or alternative could improve users' ability to locate distributions (AT6) and compare them (AT3), especially in cases where the distributions are spread across regions or axes. Supporting selection of multiple distributions for highlighting at once could also facilitate comparison tasks. Further, linked highlighting across all views can help users stay oriented when shifting attention between views.

There are two types of *filtering* a designer might want to support: filtering on entities and filtering on values. Filtering on entities is the ability to select a subset of alternatives or evaluators to inspect at any time, which can facilitate many tasks by removing distracting elements. Filtering is especially important when working with parallel coordinates, since the distributions occupy the same space.

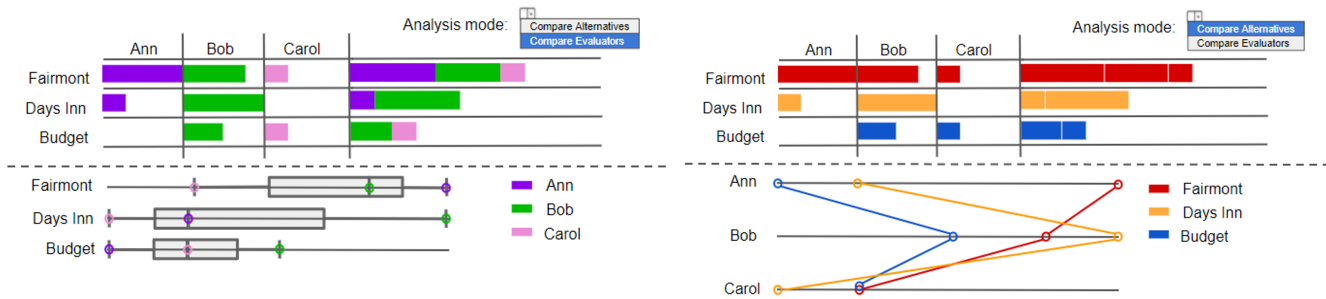


Figure 5: Design with Comparison Mode Switch.

Filtering on values is the ability to exclude alternatives based on TotalScore or AltScore for a particular evaluator, which allows users to set *qualifying thresholds* (i.e. constraints) that must be met for an alternative to be considered. This feature is not required to support any of the tasks we identified, but it could be useful in scenarios where qualifying thresholds are important.

A final type of transformation involves augmenting the display with *details on-demand*. For example, users might want to query the value encoded by a mark, as it may be difficult to glean from the graphical representation alone. Possible implementations include toggled value overlays or tooltips that appear when users hover over a mark. Tooltips could also include the label for the mark in order to expedite identification (AT1). Details of particular importance include text explanations supplied by evaluators; they could be displayed whenever a user interacts with the corresponding mark. Other forms of textual information designers might consider making available on demand include averages, variances, and axis details.

5 EVALUATION

We have implemented an online tool (see Figure 1) based on the recommendations for Option 2 in Sections 4.3.3–4.3.4 and as depicted in Figure 5. It allows preference elicitation and analysis to be performed remotely and/or asynchronously; it also allows evaluators to update their preferences while they view those of the group. We have performed a preliminary evaluation with two user studies.

5.1 Preference Elicitation

Prospective evaluators are presented with a list of alternatives, each of which may have a description and/or reference hyperlink. Then, each evaluator can assign AltScore values using sliders on a simple one-to-ten scale and provide the rationale behind their score for each alternative. Scores and comments are saved to an online database.

5.2 Visualization Design and Implementation

As discussed in Section 4.3.3, our prototype displays tabular and stacked bar charts as well as a box or parallel coordinates plot (depending on comparison mode). It was implemented in TypeScript using a MEAN (MongoDB, Express, Angular, NodeJS) stack; and D3 [6] is used to generate the visualizations.

Our prototype incorporates many of the interactions discussed in Section 4.3.4. Alternatives may be *sorted* by TotalScore or AltScore by clicking on text labels above the tabular and stacked bar charts. *Linked highlighting* occurs when hovering over a mark; marks associated with that evaluator or alternative (depending on comparison mode) are emphasized by reducing the saturation of all other marks. Toggles for each evaluator and alternative enable *filtering*. *Details* are made available in a number of ways. Clicking the name of an alternative in the control area displays a dialog containing the alternative name, description, and reference hyperlink. Tooltips in the box and parallel coordinate plots display AltScore values for marks. A numerical overlay for the bars and stacked bars is available via

toggle. Since names may require truncation, tooltips for text labels display full names. Finally, hovering over any mark reveals a detail pane containing the evaluator and alternative names, the AltScore value, and any comment about the alternative by the evaluator.

Finally, evaluators can change their AltScore values with sliders or open dialogs to edit their comments. These changes are immediate locally and may be discarded or shared with the group. This seamlessly supports the iterative nature of GDM by allowing users to move between inspection and elicitation (see Section 2.1).

5.3 User Studies

The decisions considered in our two user studies varied in scope and impact, but both could be modeled by scoring alternatives directly.

Study Format: Before the group meeting, participants were sent an email with instructions for registering with the tool and submitting their preferences. During the meeting, the participants were directed to a questionnaire and filled out the portion concerning their prior visualization experience. Their collected preferences were then used to give a demonstration of the prototype. Afterwards, participants were given up to five minutes to individually analyze the group’s preferences, with a focus on finding areas of (dis)agreement. From this point, they were also permitted to update their preferences.

The meeting was then opened up for group discussion, with the goal of reaching a consensus. Note that, while the preferences of the author who conducted the study were displayed along with those of the participants, the author did not participate in the discussions other than to answer direct questions about their preferences.

The participants then filled out the remainder of the questionnaire (adapted from the one used in Bajracharya et al. [2]) to provide feedback on the tool. Most of the questions used a 5-point Likert scale, but there were also places for free-form feedback.

User Study 1: MERIDIAN² A consortium working to provide resources and infrastructure for marine environmental research. Two of the authors are currently members of MERIDIAN, though only the author conducting the study was involved in it. This study was held via video conference with eight MERIDIAN members in four locations; the conducting author shared their screen as a common reference display, while the participants used laptops to access the tool for individual analysis. The task was to rank current MERIDIAN projects in terms of priority for future work. This task was significant because the participants had a personal stake in the outcome, even though the discussion was only preliminary.

User Study 2: NLP The Natural Language Processing (NLP) reading group attended by two of the authors; only the author conducting this study was involved. It was held in-person with nine participants tasked with ranking papers to be presented at future meetings. Compared to the MERIDIAN group, this task was less impactful for the participants, since it would not have long term effects on their research projects; even so, the alternatives were selected from different areas of NLP, so some disagreement was expected.

²<https://meridian.cs.dal.ca/>

Table 7: Mean responses to questions on a 5-point Likert scale (1 = “strongly disagree”, 5 = “strongly agree” unless otherwise specified).

	MER	NLP
Prior experience: (<i>1 = none, 5 = extensive</i>)		
Bar chart	4.83	4.12
Stacked bar chart	4.17	3.00
Box plot	4.00	3.25
Parallel coordinates plot	1.83	1.75
Microsoft Excel	4.50	3.62
Tableau	1.83	2.00
Visualizations in general	3.67	3.12
Usefulness: (<i>1 = not useful, 5 = very useful</i>)		
Bar chart	4.33	4.25
Stacked bar chart	4.67	4.62
Box plot	4.50	3.12
Parallel coordinates plot	2.33	3.12
Legend	4.67	4.38
Mean of Visualization Elements	4.10	3.90
Showing/hiding alternatives	4.00	4.00
Showing/hiding evaluators	4.00	4.12
Sorting by evaluator/total score	4.50	4.88
Switching between comparison modes	3.00	3.62
Showing/hiding the score overlay	3.33	4.12
Linked highlighting	4.50	3.88
Mean of Interactive Techniques	3.89	4.10

	MER	NLP
I modified my own preferences during the session (<i>1 = not at all, 5 = significantly</i>)	4.33	2.12
This tool helps make our discussions more participatory	4.33	4.12
The tool helps identify agreements and disagreements	4.67	4.88
The tool helps make informed decisions based on everyone’s preferences	3.67	4.00
I would be happy if the alternative with the highest aggregate score were chosen	3.67	3.62
Please rate the tool’s potential to affect group interaction (<i>1 = worse, 5 = better</i>)	3.83	4.00
Please rate the tool’s potential to affect information exchange among participants (<i>1 = less exchange, 5 = more exchange</i>)	4.33	3.75
This tool was suitable for the complexity of the decision being made	3.67	3.88
I would like to use this tool for making similar collaborative decisions in the future	4.00	3.62
I would like to use this tool for making more complex collaborative decisions in the future	3.83	3.75

5.4 User Study Results

Six of the MERIDIAN and eight of the NLP participants filled out the questionnaire. Table 7 shows aggregated responses to the Likert scale questions; see the supplemental material for full results.

Participants easily identified areas of disagreement, particularly in cases where there was some apparent agreement, but also a small number of outliers. Both groups sought to find high-performing alternatives with as few low evaluator scores as possible.

The participants found most of the visualization elements effective. The primary exception seems to be parallel coordinates; this may be related to the participants’ lack of prior experience (< 2) with that idiom, rather than an assessment of its general effectiveness. Also noteworthy is that all but one participant scored at least one idiom as “very useful”; further, that idiom was not the same for all participants, suggesting that the multiform approach may have utility across a broad user base. Even so, the respondents seemed to find the stacked bar chart the most useful overall. While it is not very effective in most auxiliary tasks (see Figure 4 and Table 6), it excels in auxiliary task AT9:B (*Search:Browse to obtain min/max values*), highlighting its importance in the decision process. Interactions were also found to be very effective. The one deemed least effective (but still with micro-average 3.35) was switching comparison mode; this is likely related to the scores for parallel coordinates.

The use of comments was mixed between and within the study groups. In MERIDIAN, only three participants annotated their scores with comments, and two of those only commented on their own projects. Conversely, while only four NLP participants commented on their scores, two of those annotated all of their scores, while a third annotated all but one score. In both studies, however, participants found evaluator comments to be useful during preference inspection and used them as reference material during the group discussions. This suggests that such annotation can be of benefit to the GDM process, even though decision makers’ motivations may vary significantly, and thus may require additional support.

All of the MERIDIAN respondents modified at least one of their preferences, as did half of the NLP respondents. The majority of the preference changes occurred during group discussions, as noted by the conducting author. For MERIDIAN, one project was given low scores by nearly all participants because it seemed unrelated to the goals of the consortium; when its applicability was explained,

the participants increased their scores considerably for that project. Another project had its scores reduced when it was noted that its intended purpose could be achieved relatively inexpensively by existing systems. For NLP, the participant scores were largely driven by relevance to their own research projects; the majority of the changes occurred when it was revealed that the techniques used in one of the papers for a particular domain could be applied more broadly.

The mean scores also suggest the tool enables participatory discussions (4.33 and 4.12), helps identify areas of (dis)agreement (4.67 and 4.88), and can increase information exchange (4.33 and 3.75). Even so, it is noteworthy that the NLP group seemed less interested in using the tool than the MERIDIAN group. This may be due in part to the difference in impact that the tasks given to each group would have. Reading group meetings are a small part of a research group’s time and effort, and so there would be less inclination to spend time coming to a consensus about them; whereas the prioritization or removal of a project in MERIDIAN would have a significant impact on the day-to-day operations of its members. We note that the NLP group, on average, would be more likely to use our tool for more complex decisions than for ones similar to their task; this may suggest some conflation between decision complexity and decision impact, which can be explored in future user and usability studies along with the other insights gained from this work.

The most frequent feature request was for the ability to anonymize the evaluators; people whose preferences are outliers may feel pressured to conform to the rest of the group, and a layer of anonymity could help offset that effect. Another request was for additional sorting options for alternatives, such as sorting by variance. This matches the observed tendency to look for alternatives that maximize aggregate score, while minimizing disagreement. Participants also noted that the visualization would become harder to read for much larger scenarios; addressing scalability concerns without sacrificing effectiveness is important for future work.

6 CONCLUSIONS AND FUTURE WORK

Effective GDM is critical for the advancement of our society; however, it can be challenging due to its multi-variate and interpersonal nature. There is considerable evidence that structured decision processes and individual preference modeling in particular can promote more fruitful analysis and discussion, ultimately leading to greater

satisfaction with outcomes. InfoVis solutions have the potential to enrich this process, but only a handful have been attempted, and none seem to be grounded in a comprehensive understanding of user needs and design options. This paper is a first step in addressing this gap. By going through the nested model of visualization design, we have developed and tested a prototype to support GDM when decision makers express their preferences directly on the alternatives.

One limitation of our analysis of real-world scenarios is that they were not examined *in situ*, and so it is possible that some tangible realities were missed. Thus, a promising area for future work would be to conduct design studies following the methodology proposed in Sedlmair et al. [35], with our data/task models and current prototype providing an initial framework that we could refine to satisfy more specific requirements. This would complement the breadth of our current work with additional depth and practical grounding.

As a first step in developing soundly designed preference visualizations for GDM, our encodings analysis has for now been focused on relatively small-scale decision problems where each evaluator scores each alternative. The next step will be to extend the design space to decision scenarios requiring more complex preference models (e.g. where alternatives are represented by a set of attributes) as well as larger sets of alternatives and evaluators. Although this may require considerable additional effort, we are optimistic on its feasibility, since the core data abstractions and tasks will be similar, and so the analytic evaluation of tabular encodings presented in this paper will inform the rest of its development. In addition, the insights and new questions gained from our user studies will inform new user and usability studies to evaluate future GDM support tools.

REFERENCES

- [1] K. Abraham, F. Flager, J. Macedo, D. Gerber, and M. Lepech. Multi-attribute decision-making and data visualization for multi-disciplinary group building project decisions. In *Working Paper Series, Proc. of the Engineering Project Organization Conference*. Winter Park, CO, 2014.
- [2] S. Bajracharya et al. Interactive visualization for group decision-analysis. *Submitted to: Decision Support Systems*, 2017.
- [3] C. A. Bana e Costa and J.-C. Vansnick. The MACBETH approach: Basic ideas, software, and an application. In *Advances in Decision Analysis*, pp. 131–157. Springer, 1999.
- [4] J. Bautista and G. Carenini. An integrated task-based framework for the design and evaluation of visualizations to support preferential choice. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp. 217–224. ACM, 2006.
- [5] D. Borland and R. M. Taylor II. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Appl.*, 27(2):14–17, 2007.
- [6] M. Bostock, V. Ogievetsky, and J. Heer. D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.
- [7] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
- [8] D. Brodbeck and L. Girardin. Visualization of large-scale customer satisfaction surveys using a parallel coordinate tree. In *IEEE Symposium on Information Visualization*, pp. 197–201. IEEE, 2003.
- [9] V. A. Brown, J. A. Harris, and J. Y. Russell. *Tackling Wicked Problems: Through the Transdisciplinary Imagination*. Earthscan, 2010.
- [10] R. T. Clemen. Making hard decisions: an introduction to decision analysis. 1996. *Duxbury Press*, 1996.
- [11] D-Sight, 4 Rue des Pères Blancs 1040, Brussels, Belgium. D-Sight. <http://www.d-sight.com/>, 2015.
- [12] J. S. Dyer and R. F. Miles Jr. An actual application of collective choice theory to the selection of trajectories for the Mariner Jupiter/Saturn 1977 project. *Operations Research*, 24(2):220–244, 1976.
- [13] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013.
- [14] J. S. Guest et al. A new planning and design paradigm to achieve sustainable resource recovery from wastewater. *Environmental Science & Technology*, 43(16):6126–6130, 2009.
- [15] P. Hansen and F. Ombler. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis*, 15(3-4):87–107, 2008.
- [16] M. Harrower and C. A. Brewer. Colorbrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.
- [17] C. L. Hwang and K. Yoon. *Multiple Attribute Decision Making: Methods and Applications State-of-the-Art Survey*, vol. 186. Springer Science & Business Media, 2012.
- [18] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen. Collaborative visualization: Definition, challenges, and research agenda. *Information Visualization*, 10(4):310–326, 2011.
- [19] W. Jin, D. Gromala, C. Neustaedter, and X. Tong. A collaborative visualization tool to support doctors’ shared decision-making on antibiotic prescription. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 211–214. ACM, 2017.
- [20] N. Mahyar et al. ConsensusUs: Visualizing points of disagreement for multi-criteria collaborative decision making. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 17–20. ACM, 2017.
- [21] R. D. Margerum. Collaborative planning: Building consensus and building a distinct model for practice. *Journal of planning education and research*, 21(3):237–253, 2002.
- [22] A. P. Mishkin and E. A. Hindalong. Web ValueCharts. <http://www.cs.ubc.ca/group/iui/VALUECHARTS/>, 2018.
- [23] T. Munzner. A nested model for visualization design and validation. *IEEE Trans. on Visualization and Computer Graphics*, 15(6), 2009.
- [24] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.
- [25] J. Mustajoki and R. P. Hämäläinen. Web-HIPRE: Global decision support by value tree and AHP analysis. *INFOR: Information Systems and Operational Research*, 38(3):208–220, 2000.
- [26] J. Mustajoki, R. P. Hämäläinen, and M. Marttunen. Participatory multicriteria decision analysis with web-hipre: a case of lake regulation policy. *Environmental Modelling & Software*, 19(6):537–547, 2004.
- [27] J. Mustajoki, R. P. Hämäläinen, and K. Sinkko. Interactive computer support in decision conferencing: Two cases on off-site nuclear emergency management. *Decision Support Sys.*, 42(4):2247–2260, 2007.
- [28] S. Pajer et al. WeightLifter: Visual weight space exploration for multi-criteria decision making. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):611–620, 2017.
- [29] C. Pavitt. *Small group communication: a theoretical approach*. Gorsuch Scarisbrick, 1994.
- [30] P. G. Pham and M. L. Huang. Qstack: Multi-tag visual rankings. *Journal of Software*, 11(7):695–703, 2016.
- [31] Z. Qu and J. Hullman. Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):468–477, 2018.
- [32] M. Rempel and C. Young. Vipor: A visual analytics decision support tool for capital investment planning. Defence Research and Development Canada, 2017.
- [33] N. B. Robbins. Dot plots: A useful alternative to bar charts. *Business Intelligence Network Newsletter*, 2006.
- [34] A. Salo and R. P. Hämäläinen. Multicriteria decision analysis in group decision processes. In *Handbook of Group Decision and Negotiation*, pp. 269–283. Springer, 2010.
- [35] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.
- [36] M. Streit and N. Gehlenborg. Points of view: Bar charts and box plots. *Nature Methods*, 11(2):117–117, 2014.
- [37] P. Talkad Sukumar, R. Metoyer, and S. He. Making a pecan pie: Understanding and supporting the holistic review process in admissions. *Proc. of the ACM on HCI, 2(CSCW)*:169, 2018.
- [38] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proc. of the Working Conf. on Adv. Visual Interfaces*, pp. 110–119. ACM, 2000.