Understanding the Role of User Interface for Multi-Criteria Decision-Making in Supporting Exploratory Usage of Information Systems

Sungsoo Ray Hong[∞], Rafal Kocielnik[∀], Cecilia Aragon[∀], Sarah Battersby[∇], Juho Kim[△]
[∞]George Mason University, [∀]University of Washington, [∇]Tableau Research, [△]KAIST
shong31@gmu.edu, {rkoc,aragon}@uw.edu, sbattersby@tableau.com, juhokim@kaist.ac.kr

Abstract

Multi-Criteria Decision-Making (MCDM) is one of the essential tasks people encounter in their everyday use of information systems. In MCDM, people weigh the relative importance of multiple decision criteria to assess decision candidates. Such an MCDM task is ubiquitous yet can be cognitively taxing without the explicit support of user interfaces (UIs). However, there has been a lack of approaches aiming at systemically understanding how the design of UIs can affect users' attitudes and behavior in performing their exploratory use of information systems under MCDM scenarios. To better understand the role of UIs in MCDM, we determine two factors in characterizing UI for MCDM; (1) the internal representation, the way that UI frames end-users in determining preferences of decision criteria (i.e., individual, proportional, and pairwise quantifies), and (2) the external representation, the way that UI externalizes user preferences while the users interact with systems (1D and 2D layouts). We conducted two studies to understand how different design choices affect users' MCDM processes. We found 2D layout improves a set of attitudinal aspects in MCDM scenarios while using different quantifiers introduces a set of trade-offs.

1. Introduction

People frequently encounter the situation of Multi-Criteria Decision-Making (MCDM) in their everyday use of information systems [1]. Tasks associated with MCDM is prevalent, ranging from casual usage (e.g., searching places and items [2]) to high-stake decision-making situations in a professional context (e.g., deciding on the treatment of patients in ICU or the location of a company branch [3]). To better support MCDM, researchers have put significant effort into constructing models that explicate users' decision-making process [4] and information-seeking behavior associated with MCDM [5]. Starting from users' identification of their "current situation"

(i.e. a user's awareness of important criteria to consider), studies have framed MCDM as a process of determining their ideal "structure" of relative preferences regarding the criteria that returns the most desired outcome, through exploratory system usage [6]. Because user interface (UI) serves as a major vehicle in helping people to realize their structure, researchers in Human-Computer Interaction (HCI), Computer-Supported Cooperative Work (CSCW), and Information Visualization (InfoVis) have devised novel UIs to support users in identifying, weighing, and aggregating multiple decision-criteria related to the decision-context [7], aiming at leading the users to a "better" outcome [1].

Grounded in theories of information-seeking [8, 9], studies have found how the design of MCDM user interfaces can specifically make differences in supporting users in exploring decision candidates [10, 11, 12, 13, 14]. For instance, without having a way of indicating their preference to a system, studies found users often make a decision without understanding the consequence of their choice [15], not being able to fully explore alternative candidates [4], and perceiving a lack of control in their search [16]. Explicit support of UI can help users to be more transparent [17] by helping them to more efficiently and effectively communicate with the system in terms of which criterion they feel relevant to their decision-context and how much more important any given criterion is than the others [2]. Despite the large body of research aimed at devising different types of UIs in supporting MCDM scenarios, there has been a lack of approaches exploring the design space of UIs and characterizing the factors that can affect users' attitudinal and behavioral usage patterns. To date, the designers have no concrete guidelines that inform their design-decisions in building front-end side of exploratory decision-support tools [13].

The primary goal of this work is to better understand the role of UI design in MCDM. Our directions for design exploration are two-folds. In the first direction, we explore the ways the UI frames structuring users'

HÏCSS
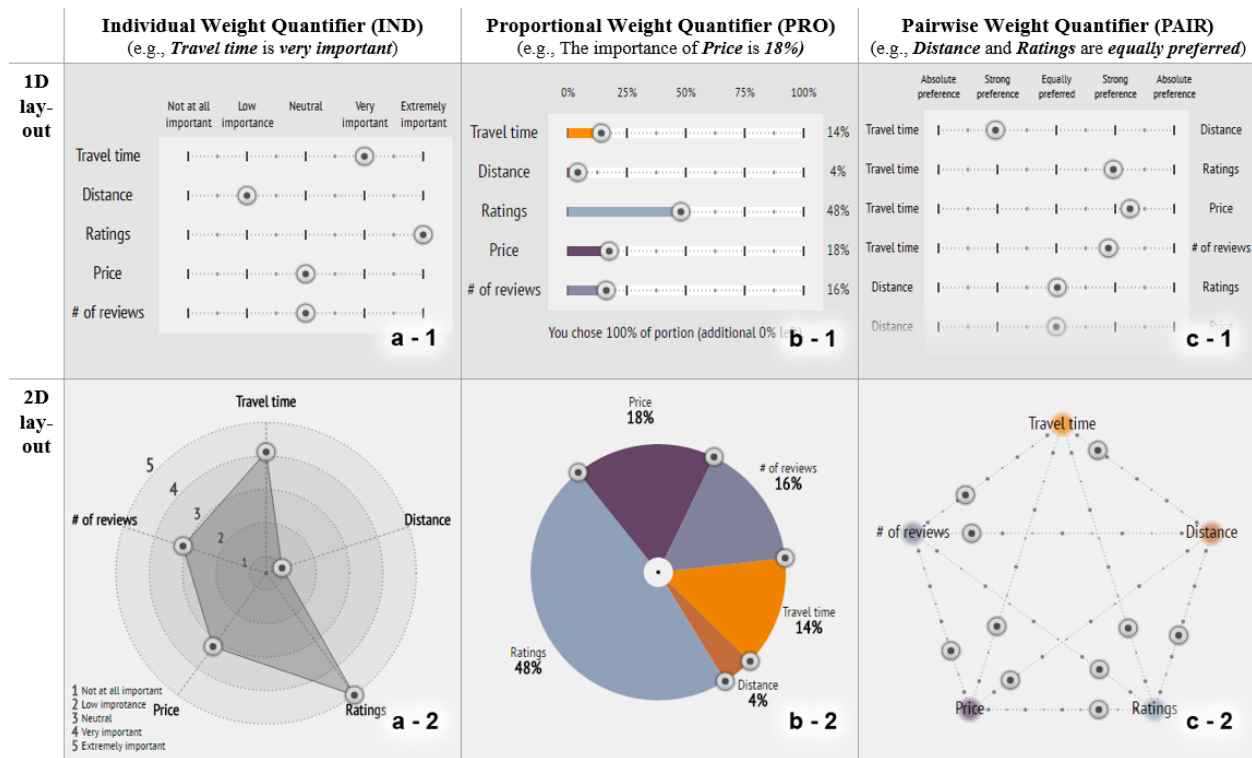
Figure 1. Each column shows different internal representation (or weight quantifiers). From the left, individual, proportional, and pairwise weight quantifiers. Each row shows different external representation (or layouts) From the top, baseline (conventional) and two-dimensional layout.

mind (or their *internal representation*) in determining relative preferences of decision-criteria, based on studies in MCDM investigated in economics, HCI, and organizational research [1]. In particular, we identified three different ways of eliciting users' importance of decision-criteria; individual, proportional, and pairwise *weight quantifiers* (see the three columns in Fig. 1). The second direction is informed by the theory of affordances developed by McGrenere & Ho [18], Norman [19, 20], and Gibson [21]. Affordances refer to the properties of user interfaces that help users to reason which actions they can take using some external cues and/or metaphor [18]. While a wide variety of designs have been suggested in MCDM [11, 22, 23], there has been little research focusing on understanding how the UIs can augment their affordances using visual cues. Consequently, we derive alternative *external representations* that use visual cues in visualizing user preferences for the three quantifiers (see Fig. 1, the second row with *2D layouts*).

We conducted two rounds of studies to understand how choosing different internal and external representations in designing MCDM UIs can affect the ways in which users interact with information systems in performing their MCDM tasks. In Study 1 (S1), we asked 117 Amazon Mechanical Turk users (Turkers) to use a 1D and 2D pair of one of the three weight quantifiers and examined how using different interfaces changes their interaction with a system in place finding tasks. In Study 2 (S2), we asked 8 participants to freely use the 6 conditions and captured their perceived differences across conditions through semi-structured interviews to more deeply understand our findings in S1. Through the studies, we found applying 2D improves a set of attitudinal aspects in MCDM scenarios, while using different quantifiers can introduce a set of perceived trade-offs.

This work offers the following contributions. First, we characterize the design space of user interface for supporting MCDM using an internal and external representation of eliciting and presenting user preferences of decision-criteria and instantiate interactive MCDM UIs based on our characterization. Second, we provide insights we gained through the two studies, helping researchers and practitioners to better understand the expected consequence of their design choice in building UI for MCDM. Finally, we present a web-based open-source library, MC^2.

2. Related Work

We review research in MCDM especially focusing on the UI design to characterize the design space of UIs for MCDM. First, we look at the MCDM work in HCI, InfoVis, and other fields focusing on the design of UI for better supporting exploratory information-seeking. In this direction, we aim at understanding how UI can support structuring people's internal representation of an MCDM problem. Then we dive deeper into the theory of affordances as a framework for improving the external representation of UIs for improving communication between a user and a system.

2.1. Information-Seeking Models and MCDM

In decision-making scenarios, the fundamental goal of information systems is to guide users to choose an optimal candidate while reducing information overload [22]. Within the information retrieval community, there has been a rich body of research aimed to improve the quality of recommendations through better algorithm designs [24] such as collaborative filtering and item-based filtering [25, 26]). However, decision-making scenarios are known to be iterative and exploratory, and the necessity of facilitating active human involvement using better design has grown in HCI, InfoVis, and CSCW community [27, 13, 28].

Models in Information-seeking behavior provide a useful framework for understanding users' behavior to satisfy their information needs using information systems [29]. For example, a sense-making framework sees the process of searching the information as an iterative process which aims to fill a gap between the "current situation" and the outcome desired by a user [6]. Ingwersen argues that the user has an initial model of their information needs, which is usually implicit and imperfect [30]. This imperfect understanding of their information needs [1, 31] leads users to go through multiple interactions until they can fully clarify the structure of their needs [32]. Aligning with the findings in this line of research, studies in MCDM found that users' perceived cognitive load increases as they consider multiple aspects of the candidates [2] and especially coping with trade-offs [15, 33]. Presenting suitable UIs to elicit the structure of user needs at an early stage of communicating has the potential to reduce the uncertainty of decision-making [34].

With the progress of understanding information-seeking models for explaining how humans behave, communities in HCI, InfoVis, and Information Retrieval devised and studied a wide range of application areas, including e-commerce [35, 15], media contents browsing [11, 22], travel and tourism planning [16, 23, 36, 37], conference presentations attending plan [38], and group decision-making scenarios [12, 39]. Studies in MCDM found positive aspects of how UI can better support users in seeking information. For instance, facilitating users' exploration of a target information space allows users to have a better sense of why a particular set of recommendations is made [1, 36], improving recommendation quality using users' multifaceted preferences regarding criteria [22], engaging in providing more feedback about their preferences [40].

Despite the rich body of design research on MCDM UI that covers broad application areas, there has been no dedicated research focusing on systemically understanding how different design choice of UIs impact users' building of the relative importance of criteria, or helping them to form their *internal representation* of the "current situation". A more dedicated effort may be required for characterizing UI for MCDM [40].

2.2. Theories of Affordances and MCDM

To explore design opportunities of UIs in MCDM, we turned to the theory of affordances [18]. The theory focuses on visual feedback and clear communication of permissible interactions and the constraints directly in the UI, thus offering a valuable framework for effective redesign of UIs for MCDM for facilitating users' iterative communication in identifying their multi-faceted needs.

According to Norman, affordances are perceived properties of a designed object that provide clues to the operation of the object and suggest a range of possibilities [20]. Furthermore, Norman advocates the "make it visible" guideline to make the affordances directly perceived [19]. The interface for setting user preferences can visually communicate the relative importance of the criteria and clearly indicate to the user how to make a particular criterion more or less important than other criteria. Taking full advantage of the concept of affordances can involve enhancing the visual salience of the multi-criteria structure. Aside from affordances themselves, Norman defines the concepts of constraints, with physical constraints closely related to affordances [19]. As exemplified by Norman, a physical constraint could be "locking the mouse button when clicking is not desired". In the context of our design space, the constraint could be used to limit the ability of the user to assign more importance to one criterion, unless at an explicit expense of the other criterion. Hence, by using constraints, the GUI could

directly communicate to the user the inter-related nature of multiple criteria. The final concept from the theory of affordances, relevant to our work, is the notion of feedback. Norman describes it as a "mechanism" that advertises the affordances and he further argues that the presence of feedback in the interface can dramatically improve the usability and the transparency of the system independent of the affordances themselves [18].

We see using the concepts of affordance, constraints, and feedback from the theory of affordances as useful for enhancing *external representation* of UIs in MCDM. While "good" UI design practice involves the proper use of affordances (e.g. a slider should communicate it can be moved), affordances present in generic UI elements are not specifically designed with the purpose of helping users understand the capabilities and constraints specific to the MCDM. In our redesign, we aim to design affordances, constraints, and feedback specifically tailored to the MCDM context. We further note that while Norman makes a distinction between real affordances (that allow for certain interaction but may not be visually communicated) and perceived affordances (visually communicated but not necessarily there), for our purposes this distinction is not as relevant and thus we use the term visual affordance to indicate a real affordance that is also perceived by the user.

3. Characterization of MCDM UI

Through this review, we determined the two dimensions that can characterize the UI designed for MCDM scenarios; (1) internal representation, the way the design frames uses' understanding of one's own preference structure in MCDM, and (2) external representation, the way that the design externalizes the elicited preference structure.

3.1. Internal representation: Weight Quantifiers

A user's basic unit operation in MCDM is indicating one's perceived importance regarding single or multiple criteria through a UI. While going through such operations, a user gradually structures one's internal representation regarding information space-of-interest and determine the most suitable decision candidate [6]. Through the analysis of literature in MCDM, we determined the three different ways of eliciting the relative importance, which we call weight quantifier (WQ) hereinafter. Those are (1) *individual* WQ; a WQ that allows a user to set the importance of criteria separately without considering the impact of changes related to other criteria, (2) *proportional* WQ, a WQ that asks a user to set proportion for setting the importance

of a target criterion and simultaneously sums every importance to 100%, and (3) *pairwise* WQ, a WQ that asks a user to consider relative importance between a pair two criteria. Core attributes of each WQ are highlighted in Table 1.

3.1.1. Individual WQ (IND) IND allows users to set the degree of importance of each criterion separately, and is widely used in research and practice (e.g., [16, 11, 22, 15, 38], see Fig. 1 (a-1)). While interacting with IND, users consider only one criterion at a time. Typically, across every criterion, the same scale is used to represent their weight (e.g., the Likert scale that indicates the degree of importance, or numeric scale from 0 to 1.0). A user decides relative ranking between criteria while setting each criterion's degree of importance one by one. However, users may not be able to have an explicit sense of the precise weights used for yielding the rank. In general, this additive model uses the overall value of the item as follows:

$$V(a) = \sum_{i=1}^{n} w_i v_i(a)$$

where $V(a)$ is the overall value of alternative a, w_i is the weight assigned to reflect the importance of ith criterion (out of n criteria) of a, and $v_i(a)$ is the standardized score of a's ith criterion. To yield $V(a)$, w_i is divided by $\sum_{k=1}^{n} w_k$ for normalization. Specifically, we assign the weight of used scale based on [1] (see p.137). $v_i(a)$ is derived as follows:

$$v_i(a) = \frac{a_i - x \in A_i min f(x)}{x \in A_i max f(x) - y \in A_i min f(y)}$$

where A_i is a set of values in ith criterion of every item, and a_i is the actual value of a's ith criterion. Note that a denominator of the above equation can be set to $x \in A_i min f(x) - a_i$, where $x \in A_i$, in case less quantity is desirable for ith factor.

3.1.2. Proportional WQ (PRO) Users can set the exact amount of proportion for each criterion with PRO (e.g., [23, 36], see Fig. 1 (b-1)). A numeric scale, such as 26%, is used to indicate the importance of a criterion. Users can identify not only the ranks between criteria but also how much one criterion is more influential than other criteria. In using PRO, to add more proportion to one criterion, it is required that the user must redistribute the same amount and remove from the rest of the criteria. w_i can be set directly from a user's setting

	Individual	Proportional	Pairwise
User cognitive scope while interaction	One criterion at once	Up to all criteria	Two criteria at once
User realization of ranks between criteria	Explicit	Explicit	Not Explicit
User realization of exact weights of criteria	Not explicit	Explicit	Not explicit
# of handles upon *n* criteria	*n*	*n*	*n(n-1)/2*

Table 1. Notable characteristics of three WQ types

without normalization (e.g., 26% will converted to 0.26 as w_i).

3.1.3. Pairwise WQ (PAIR) Pairwise WQ has been used in [35, 37], and helps users dismantling a complex problem to more easily compare two criteria at one time (see Fig. 1 (c-1)). A Likert scale is commonly used to aid users to set the relative importance between two paired factors. A known difficulty of PW is the exponentially increasing number of WQs as users consider more criteria (i.e., n criteria will generate *n(n-1)/2* UI widgets, see 10 handles in Fig. 1 (c-2) for covering 5 criteria). Having many handles also results in spreading out each criterion when positioning UI widgets (In the worst case, a user has to scan through every UI widget although one is interested in only a specific criterion). To derive the weights, the Analytic Hierarchy Process (AHP) is dominantly used [41]. AHP constructs an n by n pairwise comparison matrix M, where rows and columns indicate n factors being used, and each element M_{pq} in the matrix specifies how much one prefers a factor in the corresponding row p over the corresponding column q (see [1], page 151). The aim of AHP is to find a set of values $w_1,\ldots,w_n$, such that M_{pq} are approximated as closely as the corresponding ratios w_p/w_q. This can be solved by yielding eigenvector w of the equation below giving the maximum eigenvalue λ. Once all w_i are obtained, each is normalized to the sum of the w (i.e., $\sum_{k=1}^{n} w_k$).

$$\sum_{q=1}^{n} M_{pq} W_q = \lambda W_p$$

3.2. External representation: 2D Layouts

In creating an alternative design of the 3 WQs for better affordances, we adopt 2D layouts that visualizes a user's preferences while they indicate their multifaceted preference regarding criteria. The 2D layout is known to be descriptive and explanatory (Zhang Norman, 1994), presenting an improved external representation of the users' internal representation about their preference structure. In terms of the design of UI for MCDM, however, the majority of approaches apply a 1D layout such as stacked radio buttons [15, 35], sliders [11, 36], or drop-down lists [16, 37]. The key to reinforce clear visual affordances is choosing 2D visualization layouts that align with a user's internal representation of preferences structure [42]. In choosing the layout for MCUI, we consider (1) *familiarity*, how widely the layout is used, and (2) *intelligibility*, how easy the layout can be interpreted, from the general user. Consequently, we chose a radar chart for IND, a pie chart for PRO, and a radial chord diagram for PAIR. We assume such a design choice can help a user to understand how the system aggregates their inputs in a more transparent manner, helping them to understand which actions they would take further.

3.2.1. IND-2D Radar chartis widely used to present multivariate information in the form of a two-dimensional space [43]. The layout is especially well known to the public in terms of presenting multifaceted aspects of one entity (e.g., an NBA player's performance, a multi-faceted beer flavor, or medical record of a patient). In our design, each axis of a radar serves as a slider for each corresponding factor. Every slider has a handle that users can drag to set their preference for each factor. We use a 5-level Likert scale to present the degree of importance. Putting the full-scale description (e.g., "Very important") onto the radar can cause visual clutter and overlaps. We instead display numbers from 1 to 5 on each circle, along with a legend that maps these numbers to the Likert scale. 2D-IND is shown in the left on Fig. 1 (a-2).

3.2.2. PRO-2D A pie chart has been widely adopted "in the wild"; random sampling from Visual.ly revealed that 36% of infographics used the pie chart [44]. Due to its popularity, we choose the pie chart to improve PRO-1D. In PRO-2D, one slice of a pie indicates one of the factors, and the degree of the angle of the slice indicates the weight of the corresponding factor. To control the weight of each factor, we present handles between every two adjacent slices. A user can drag the handles to arrange the weights between two adjacent factors. One benefit of adopting this design is to present clear visual affordance regarding the constraints that

PRO has; the sum of a user's weight indication cannot exceed 100% and one needs to decrease a certain portion of the weight to increase the weight of another. To provide smooth continuity when adding or removing a criterion, we use an animated transition by linearly interpolating the two stages in our implementation. Fig. 1 (b-2). presents PRO-2D.

3.2.3. PAIR-2D Our design of WQ with 2D layout uses a pairwise comparison matrix to indicate the strength of their preference for one factor over another. A chord diagram is widely used for visualizing such a matrix. To design PAIR-2D, we chose one type of chord diagram called a radial network diagram, which positions factors in radial form and connects every possible edge between them. In our implementation, we put sliders over every edge in a graph (see Fig. 1 (c-2)). Because the layout is a graph, color can be assigned to each node (or a criterion). Users who aim to set the weights relevant to a specific criterion can take advantage of the visual affordance of this layout because every UI handle relevant to the criterion is connected to one node in a chord diagram. For example, users can set the UI handles relevant to "Travel Time" by looking through the lines connected to the top node in Fig. 1 (c-2).

4. Studies

We conducted two studies to understand how different ways of (1) supporting a user's internal representation forming process (using IND, PRO, and PAIR weight quantifiers), and (2) applying different external representation (using 1D and 2D layouts) influence user behavior and perception in MCDM scenarios. In Study 1 (S1), we presented a web interface to AMT participants (Turkers) where they were asked to use a pair of WQs using different layouts (1D, 2D) to decide on their travel destination in a hypothetical scenario. Then we measured Turkers' behavior and attitudinal data. In Study 2 (S2), we conducted interviews to draw deeper insights into explaining the reasons behind the patterns we identified in S1.

To conduct the studies, we built a web-based destination recommendation systems. The apparatus presents three panels. The first panel presents one of the two scenarios we prepared for the studies, a person finds a French restaurant for celebrating the fifth anniversary with a partner or a Japanese restaurant for an important business meeting (See Fig. 2 (a)). We prepared two scenarios to remove the learning effect about datasets in S1 where each participant makes decisions twice, one

using 1D and another using 2D. We also note that we chose high-stakes scenarios to help Turkers put serious effort into seeking the best destination suitable for the contexts. The second panel presents one of the six UIs (3, IND, PRO, and PAIR, by 2, 1D, and 2D) that a participant can use to explore decision candidates (See Fig. 2 (b)). We asked participants to consider five criteria relevant for destination finding tasks suggested in previous work [4, 45]. The criteria are review ratings, price range, and the number of reviews of restaurants, travel time, and travel distance to a restaurant. To eliminate the effect of color, we kept color gray in all six conditions. Finally, the last panel presents a list of candidates based on settings in (b) (See Fig. 2 (c)). In collecting places, we used Yelp API 2.0 in the bound box of greater Seattle area from 47.396°S, -122.440°E to 47.859°N, -122.075°W consulted from [45]. With the API, we extracted information on the five criteria for every destination we collected. Next, we filtered out locations deemed non-plausible for the high-stakes decisions. Specifically, we filtered out places with a price range below '$' and rating below 2.5. Finally, we anonymized restaurant names to reduce the possibility of bias from one's familiarity based on his/her prior experience.

4.1. S1 Methodology

In S1, we recruited 120 Turkers from the U.S. with an average age of 35.1 (SD=9.35, range: 21-70) and 47.9% females. 51% reported completing at least a bachelor's degree. In recruiting participants, we balanced the age and gender to reduce bias [46]. The average study time was 14 minutes 49 seconds and the participants were paid $1.50, roughly matching $6 per hour. To make the session reasonably short while getting comprehensive comparison results, we used a mixed study design. Specifically, participants were randomly assigned to a pair of 1D and 2D for direct comparison (within-subjects for 1D vs 2D) for one of the three WQ types (between-subjects for IND, PRO, and PAIR).

Turkers were randomly assigned to the study apparatus that showed one of the three WQ 1D vs. 2D pairs. They were given short instructions explaining the corresponding condition and the 5 criteria they will control for making a decision. This short instruction was to increase the familiarity of the conditions they will use. Then Turkers made decisions twice, first in the French restaurant scenario, and second in the Japanese restaurant scenario. But the order of 1D and 2D were counterbalanced. In this way, we nullified the learning effect of datasets while simultaneously removing the ordering effect between 1D and 2D. Alternatively,
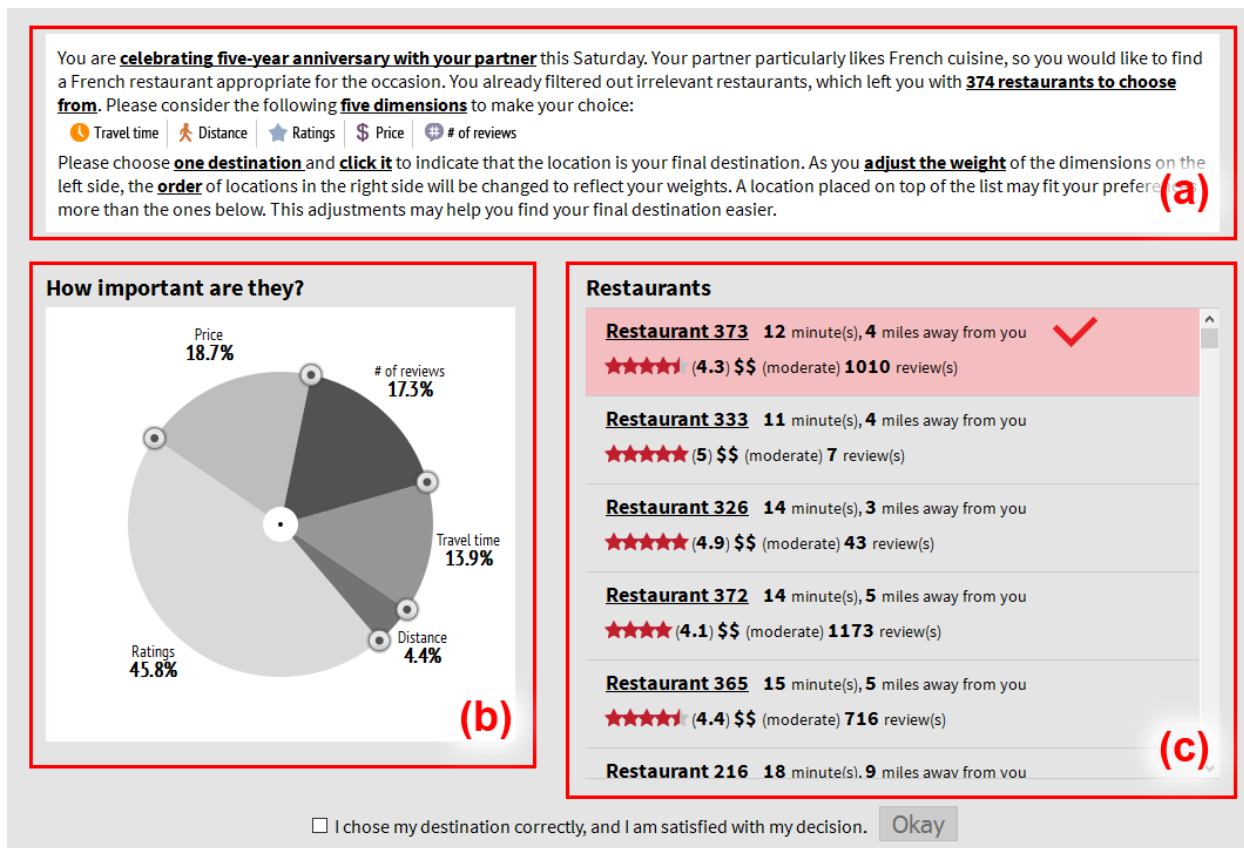
Figure 2. The web-based apparatus used in two studies which presents panels for (a) scenario description, (b) one of the UIs for MCDM, and (c) a list of recommended restaurants based on a user input made in (b).

counterbalancing the order of dataset while fixing the order of 1D and 2D is a feasible option [47].

For each scenario, they saw a list of recommendations based on their preference settings using one of the six UIs. They were asked to select a final location that fits their preferences best and told to consider their selection carefully as they will be asked questions about their choice. Turkers had to select a location by clicking it on the list and tick the checkbox indicating they chose the best choice with care. Once Turkers made a decision, the next page asked 2 attention check questions: (1) name of the restaurant they selected and (2) scenario they went through. While they explore a list of restaurants, we logged the following user (3) behavior metrics: (3a) how many times they changed weights in UI in a UI panel (b), (3b) the number of scrolls and clicks made in a list panel, and (3c) time taken for making a decision.

After Turkers finished making the two decisions using a 1D-2D pair, they were taken to a survey. In the survey, we asked Turkers' (1) demographics, (2) the perceived value of the functionality provided by MCUIs,

and (3) direct comparison questions between the 1D vs. 2D UIs in pair (scale from -3 "Strongly prefer 1D" to 3 "Strongly prefer 2D"). We derived the questions from the RS evaluation framework ResQue [48]. Specifically, we focused on measuring user's perception related to (3a) ease of use, (3b) transparency, (3c) control-ability (i.e., users' capability of creating new recommendations based on one's intention), (3d) confidence & trust, (3e) interface adequacy (i.e., perceived attractiveness and adequacy for the task [48]), (3f) usefulness of UI, (3g) result accuracy, and (3h) result satisfaction. In analyzing the results in the AMT study, we first removed the 34 Turkers who failed attention checks. This exclusion yielded an attention failure rate of 23%, which is in line with the attention-based failure reports from past AMT studies ranging from 14% to 41% [49]. This resulted in a total of 83 participants; 30 in IND pair, 28 in PRO pair, and 25 in PAIR pair.

4.2. S2 Methodology

In S2, we conducted semi-structured interviews with 8 participants who were recruited through a

university email list. S2's purpose was to gain a deeper understanding behind the patterns we determined from S1. We started by informing the participants about the study and procedure. Then we presented three 1D UIs one by one (order counterbalanced). For each UI, participants were able to freely interact with it, and while doing so follow a think-aloud protocol. After participants used all the 1D WQs, we interviewed them with a focus on direct comparisons between the three WQ types (which we were not able to get in the AMT study). After we covered 1D WQs, we presented the three 2D UIs (again, in a counterbalanced order). Then we asked them to compare 1D and 2D for the same WQ type. The interview took on average 1 hour 47 minutes. Every audio was recorded, transcribed, and analyzed for generating themes based on iterative qualitative coding process suggested by Saldaña [50].

5. Results

Through S1, we mainly identified a set of significant differences regarding participants' attitudes depending on different external representation choices. In S2, participants provided insights into the attitudinal differences we observed between 1D and 2D and also indicated how using different WQ choices affected their internal representation of making a decision in MCDM scenarios.

	IND pair	PRO pair	PAIR pair
Ease of use	0.27	-0.09	-0.75
Transparency	**0.82*****	**0.76*****	-0.06
Control	0.30	0.18	0.19
Trust	0.43	0.29	-0.19
Adequacy	**1.06****	**0.67****	0.30
Usefulness	**0.59***	**0.50****	0.12
Accuracy	0.36	**0.32***	0.03
Satisfaction	0.64	0.26	0.08
	PAIR-1D first	**PAIR-2D first**	
Ease of use	0.05	**-1.71*****	
Control	0.47	**-1.18****	
Trust	0.33	**-1.18***	
Adequacy	0.5	**-1.13****	
Usefulness	0.3	**-0.63***	
Accuracy	0.4	**-0.63****	
Satisfaction	0.73	**-0.88*****	

Table 2. Significance against a distribution centered around 0 using a Wilcoxon Signed-Rank test. Up: comparison between 1D and 2D groups for IND, PRO, PAIR pairs. Bottom: comparison between a group who used PAIR-1D first and PAIR-2D first. *** p<0.001, ** p<0.01, * p<0.05.

5.1. Effect of External Representation Choice

We found significant differences in several attitudinal constructs reported in Table 2, top. Perceived transparency of IND-2D (M=0.82, SD=1.38) and PRO-2D (M=0.76, SD=1.06) was rated significantly higher compared to their 1D counterparts with p<0.05 for IND-1D and p<0.001 for PRO-1D. However, there was no meaningful difference regarding this construct in PAIR 1D-2D pairs. The transparency results reflect that Turkers were able to understand their preferences in 2D of IND and PRO, the relation between them, and the impact of these preferences on the recommendation results much better using the 2D layout. Interviews confirmed these findings with 7 out of 8 participants reporting IND-2D as more transparent than IND-1D. For instance, P5 remarked: *"I think the shape helped me gauge the differences between factors easier, and that helped me expect what results will come after I set each of the handle."* Similarly, 5 considered PRO-2D more transparent than PRO-1D (3 reported them equal). PAIR-2D was also preferred over PAIR-1D by 5 interviewees, but 2 said PAIR-1D was more transparent due to the less visual complexity.

Similar patterns were observed in interaction adequacy, which captures the attractiveness required for a user to engage in a target task [48]). IND-2D (M=1.06, SD=1.39) and PRO-2D (M=0.67, SD=1.26) were rated significantly higher compared to their 1D counterparts with p<0.01 in terms of interaction adequacy. In S2, 7, 5, and 5 of participants respectively mentioned IND-2D, PRO-2D, and PAIR-2D present a more clear and visually engaging interaction experience than their 1D counterparts. Some participants mentioned that such an aspect of 2D conditions lets them spend more effort on refining their results for making a better decision. P8 mentioned: *"Really intuitive and well made [2D layouts], this is definitely more appealing than former ones (than 1D layouts). I would spend more time on it [...] perhaps I may find a better one."*

In terms of usefulness, IND-2D (M=0.59, SD=1.28) and PRO-2D (M=0.50, SD=0.90) were rated significantly higher compared to their 1D counterparts with p<0.05 for IND-1D and p<0.01 for PRO-1D. ResQue suggests that perceived usefulness impacts the perception of recommendation accuracy [48]. Aligning with the direction of ResQue, the perceived accuracy was rated as significantly higher for 2D-PRO alternative (M=0.32, SD=0.78) with p<0.05. The difference was, however, not significant for IND 1D-2D pairs.

Despite such favorable patterns observed in 2D conditions for IND and PRO, we didn't find similar patterns in PAIR pairs. In fact, it was rated as nearly

significantly worse for some constructs (e.g., "Ease of use" in Table 2, top). However, we noticed that the Turkers gave extremely different scores to the PAIR 1D-2D pair depending on the order in which the PAIR-1D and PAIR-2D were presented. Every construct except transparency showed a significantly higher rating for the group which used PAIR-2D first over the other group (See Table 2, bottom). This indicates that the novelty of the 2D-PAIR likely made the learning curve steep for new users. On the other hand, the Turkers that started with 1D appreciated the 2D significantly more. We investigated the impact of ordering in IND and PRO 1D-2D pairs as well. Interestingly, we did not find significant differences.

The behavioral patterns showed no significant differences between the three pairs as well as in which order participants used the two conditions. Other attitudinal constructs, such as ease-of-use, control, trustworthiness, and result satisfaction also didn't show significant differences between 1D and 2D across IND, PRO, and PAIR.

5.2. Effect of Internal Representation Choice

In terms of comparing three WQ types, we see behavioral differences will mainly be caused by the properties of UI's exterior factors rather than the way those help forming user preference structure. For instance, considering five criteria in MCDM will lead a user to spend more time and complete multiple steps in PAIR than IND and PRO due to more handles they can leverage. To determine the differences between the three, we, therefore, focus on qualitative findings in S2. Consequently, we developed the themes based on the interview data as follows:

Perceived efficiency & control granularity trade-offs: IND more efficient, yet PRO and PAIR provide better control. All 8 users reported the IND helped them search locations with less effort due to its familiarity and simplicity. P6 mentioned: *"It is intuitive, and I had no issues understanding how it works from the beginning."* On the other hand, for PRO, seven mentioned they didn't immediately get the mechanism for using it. Seven also complained about PAIR's many handles. However, despite IND being perceived as the most efficient WQ type, 6 of interviewees explicitly remarked the WQ to be least reliable. They reported IND gave them less feeling of control compared to PRO and PAIR, which made made it for them hard to find "the best result". P1 mentioned: *"It (IND) didn't retrieve the ideal place compared to other interfaces."* P3 mentioned: *"It (IND) is easy to use because you see it all the time, but I don't think you will find the result you want."* At the same time, for PRO, 6 interviewees mentioned they could precisely set the weights, which made them feel more in control. P7 remarked: *"I can really focus on one or more specific factors I care with this (PRO). I needed to understand I have to set weights up to 100. But it gives me precise control."* Similarly, 4 mentioned they liked PAIR's high granularity in terms of control they have: *"It (PW) had the highest granularity that I kind of like (P5)."*

Versatile PRO: PRO presents the best transparency among the three WQ types while being perceived as reasonably easy-to-use and efficient: Interviews also revealed that PRO was perceived as a more transparent WQ. Six interviewees mentioned that PRO helped them clearly understand how the system maps their preferences. In contrast, 5 reported IND as insufficiently presenting the relations and trade-offs between conflicting criteria: *"The trade-off is not really visible here (P3)."* Also, four reported PAIR as unclear in terms of how the preference settings affected the final recommendations. P5 mentioned: *"When I move the slider in the left side (PAIR-1D), it was difficult to imagine what I will get at the right side."*

Exploratory PAIR: Participants mentioned using PAIR triggered thinking of relations between criteria they never thought of. Six interviewees reported that using PAIR is thought-provoking and self-explanatory, which has not been reported for IND and PRO. Although PAIR was reported as the least efficient WQ in MCDM, interviewees considered PAIR to be useful when they have little experience in a decision situation and have no clear idea of which criterion is important. They remarked PW facilitates revealing their own preferences and helps in identifying their own criteria. P2 mentioned: *"Before you decide the location, you don't know what is important. The interface (PAIR) helped to understand my preferences."* Visiting every pair entails more effort compared to IND or PRO, yet this can be useful when one didn't think about preferences beforehand. *"This (PAIR) can be really useful when there are a number of important factors that I didn't think of in advance.".* The finding suggests PAIR can be useful in exploratory or high-stake MCDM scenarios.

Differences in context of use: Aligning with the three themes above, interviewees generally agreed that IND is more suitable for casual scenarios, PAIR for high-stakes and exploratory, while PRO applicable for both. IND has been reported as useful mostly for casual scenarios where making a perfect decision is not important. P6 mentioned: *"I might use this (IND) for casual and less important scenarios because it requires minimum effort."* PAIR, on the other hand, has been considered as particularly useful for

scenarios in which fine control and granularity of decision-making are necessary for achieving an order of priority. Consequently, PAIR has been considered most suited for high-stakes scenarios when a well-thought-out decision is needed, and time efficiency is less important. P5 mentioned: *"If I am buying a house, for example, something I would spend the rest of my life on, I would use this (PAIR)."* Finally, 6 interviewees mentioned that PRO is suitable for both casual and high-stakes types of scenarios, due to its ability to force thoughtful considerations while let them perceive low demand on control. In general, PRO has been perceived as a WQ that strikes the balance between interaction effort and the quality of control over the preference settings. *"The proportional seems to have a good balance between the mental effort that I need to put and the degree of granularity that I attain with high precision (P7)."*

6. Discussion

Through the two studies, we found a series of patterns that are important and useful for designing UIs for Multi-Criteria Decision-Making (MCDM).

Regarding the effect of choosing different external representations, we found people's perceived transparency, interaction adequacy, perceived usefulness of UIs, and perceived accuracy of results can be all improved with appropriate design of UIs for MCDM. One interesting aspect is about the constructs we didn't find differences. Despite the people's improved perceived transparency, there were no meaningful differences in control (i.e., adjusting the results in a way that a user intended) and trustworthiness. We assume this would be related to non-transparent aggregation of multiple criteria in scoring items. Every WQ in MCDM aggregates intrinsically different criteria in terms of scale and range [1], such as a user's preference (1 to 5 in Yelp) and price range (e.g., "$$" indicates moderate price in Yelp). Aligning with this aspect, the degree to which participants were satisfied with the results didn't show differences, although they felt more engaged in their exploratory information-seeking and felt the UI is more useful in 2D conditions. Such a two-sides result may indicate that 2D can improve the experience of a user's exploratory information-seeking in MCDM, but may have limited capability to make them understand the mechanism behind scoring decision candidates using multiple criteria.

Another aspect we found interesting is no significant behavioral differences between 1D and 2D conditions. We assume this pattern is closely related to the nature of the MCDM task, which is highly *exploratory*. It has been argued that the behavioral efficiency, the

degree to which participants put their effort into completing tasks, and behavioral effectiveness, how many participants were successful in completing a task, cannot be used as an ideal "measure of success" [51]. Researchers suggested a series of ways to operationalize the success metric in exploratory tasks (e.g., false discovery rate [52] and interaction rates [53]). However, this is an on-going research direction and more study is needed. Aligning with the criticism about using behavioral efficiency and effectiveness in evaluating exploratory task, some interviewees mentioned they would be willing to put more effort depending on how they are motivated and how the interface guides them.

The last aspect that we found interesting is the interaction effect between external representation and internal representation. Enhancing external representation using 2D layout was not always successful; it was successful in IND and PRO, but not PAIR. This suggests that enhancing a conventional 1D UI to a 2D needs careful consideration. In particular, choosing a visualization layout/metaphor that can successfully facilitate user's internal representation of the current situation can be crucial. In addition, when there are many criteria to be considered, applying an interactive selection mechanism can offload user's cognitive burden. For instance, when designing PAIR-2D with more than 5 criteria, allowing a user to select a subset of criteria can make PAIR-2D a feasible choice in more complex and realistic scenarios.

Finally, we believe that our findings of "efficient" IND, "versatile" PRO, and "exploratory" PAIR can present useful insights when building an interface for supporting MCDM scenarios. One interesting aspect we find missing is exploring WQ's trade-offs in small-group collaboration or other collaborative domains.

7. Conclusion

In this work, we characterized the design space of UI in MCDM scenarios and determined two design factors of *internal representation* and *external representation*. Through the studies, we gained insights regarding how different choices of the factors would make end-users' perception and ideal use-context different. We hope our findings can motivate improvement of UIs for MCDM and more broadly, UIs for exploratory tasks.

8. acknowledgement

References

[1] V. Belton and T. Stewart, *Multiple criteria decision analysis: an integrated approach*. Springer Science & Business Media, 2002.

[2] D. R. Fesenmaier, K. W. Wöber, and H. Werthner, *Destination recommendation systems: Behavioral foundations and applications*. Cabi, 2006.

[3] P. Jankowski, N. Andrienko, and G. Andrienko, "Map-centred exploratory approach to multiple criteria spatial decision making," *International Journal of Geographical Information Science*, vol. 15, no. 2, pp. 101–127, 2001.

[4] R. G. Golledge, *Spatial behavior: A geographic perspective*. Guilford Press, 1997.

[5] T. D. Wilson, "Models in information behaviour research," *Journal of documentation*, vol. 55, no. 3, pp. 249–270, 1999.

[6] B. Dervin, "From the mind's eye of the user: The sense-making qualitative-quantitative methodology," *Sense-making methodology reader*, 1992.

[7] K. A. Cook and J. J. Thomas, "Illuminating the path: The research and development agenda for visual analytics," tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.

[8] P. Pirolli and S. Card, "Information foraging.," *Psychological review*, vol. 106, no. 4, p. 643, 1999.

[9] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," in *Proceedings of international conference on intelligence analysis*, vol. 5, pp. 2–4, McLean, VA, USA, 2005.

[10] C. Ahlberg and B. Shneiderman, "Visual information seeking: Tight coupling of dynamic query filters with starfield displays," in *The craft of information visualization*, pp. 7–13, Elsevier, 2003.

[11] S. Bostandjiev, J. O'Donovan, and T. Höllerer, "Tasteweights: a visual interactive hybrid recommender system," in *Proceedings of the sixth ACM conference on Recommender systems*, pp. 35–42, 2012.

[12] W. Liu, S. Xiao, J. T. Browne, M. Yang, and S. P. Dow, "Consensus: Supporting multi-criteria group decisions by visualizing points of disagreement," *ACM Transactions on Social Computing*, vol. 1, no. 1, pp. 1–26, 2018.

[13] S. Hong, M. Suh, N. Henry Riche, J. Lee, J. Kim, and M. Zachry, "Collaborative dynamic queries: Supporting distributed small group decision-making," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[14] J. Matejka, M. Glueck, T. Grossman, and G. Fitzmaurice, "The effect of visual appearance on the performance of continuous sliders and visual analogue scales," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5421–5432, 2016.

[15] P. Pu, L. Chen, and P. Kumar, "Evaluating product search and recommender systems for e-commerce environments," *Electronic Commerce Research*, vol. 8, no. 1-2, pp. 1–27, 2008.

[16] M. Batet, A. Moreno, D. Sánchez, D. Isern, and A. Valls, "Turist@: Agent-based personalised recommendation of tourist activities," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7319–7329, 2012.

[17] R. Sinha and K. Swearingen, "The role of transparency in recommender systems," in *CHI'02 extended abstracts on Human factors in computing systems*, pp. 830–831, 2002.

[18] J. McGrenere and W. Ho, "Affordances: Clarifying and evolving a concept," in *Graphics interface*, vol. 2000, pp. 179–186, 2000.

[19] D. A. Norman, "Affordance, conventions, and design," *interactions*, vol. 6, no. 3, pp. 38–43, 1999.

[20] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[21] J. J. Gibson, *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.

[22] B. Loepp, K. Herrmanny, and J. Ziegler, "Blended recommending: Integrating interactive information filtering and algorithmic recommender techniques," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 975–984, 2015.

[23] M. Raubal and C. Rinner, "Multi-criteria decision analysis for location based services," in *Proceedings of the 12th International Conference on Geoinformatics*, pp. 47–53, Citeseer, 2004.

[24] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*, pp. 1–35, Springer, 2011.

[25] B. Loepp, T. Hussein, and J. Ziegler, "Choice-based preference elicitation for collaborative filtering recommender systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3085–3094, 2014.

[26] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, pp. 285–295, 2001.

[27] B. Shneiderman, "Dynamic queries for visual information seeking," *IEEE software*, vol. 11, no. 6, pp. 70–77, 1994.

[28] Z. Liu and J. Stasko, "Mental models, visual reasoning and interaction in information visualization: A top-down perspective," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 999–1008, 2010.

[29] R. Savolainen, "Information behavior and information practice: reviewing the "umbrella concepts" of information-seeking studies," *The library quarterly*, vol. 77, no. 2, pp. 109–132, 2007.

[30] P. Ingwersen, "Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory," *Journal of documentation*, vol. 52, no. 1, pp. 3–50, 1996.

[31] S. French, *Decision theory: an introduction to the mathematics of rationality*. Halsted Press, 1986.

[32] P. Bogetoft and P. Pruzan, *Planning with multiple criteria: investigation, communication and choice*. Handelshøjskolens Forlag, 1997.

[33] A. Correia *et al.*, "How do tourists choose? a conceptual framework.," *Tourism (Zagreb)*, vol. 50, no. 1, pp. 21–29, 2002.

[34] C. C. Kuhlthau, "The influence of uncertainty on the information seeking behavior of a securities analyst," in *Proceedings of an international conference on Information seeking in context*, pp. 268–274, 1997.

[35] D.-N. Chen, P. J.-H. Hu, Y.-R. Kuo, and T.-P. Liang, "A web-based personalized recommendation system for mobile phone selection: Design, implementation, and evaluation," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8201–8210, 2010.

[36] M. Raubal, "Cognitive engineering for geographic information science," *Geography Compass*, vol. 3, no. 3, pp. 1087–1104, 2009.

[37] S. Nadi and M. R. Delavar, "Multi-criteria, personalized route planning using quantifier-guided ordered weighted averaging operators," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 3, pp. 322–335, 2011.

[38] D. Parra, P. Brusilovsky, and C. Trattner, "See what you want to see: visual user-driven approach for hybrid recommendation," in *Proceedings of the 19th international conference on Intelligent User Interfaces*, pp. 235–240, 2014.

[39] S. Hong, M. Suh, T. S. Kim, I. Smoke, S. Sien, J. Ng, M. Zachry, and J. Kim, "Design for collaborative information-seeking: Understanding user challenges and deploying collaborative dynamic queries," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–24, 2019.

[40] M. Jugovac and D. Jannach, "Interacting with recommenders—overview and research directions," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 3, pp. 1–46, 2017.

[41] T. L. Saaty, "How to make a decision: the analytic hierarchy process," *European journal of operational research*, vol. 48, no. 1, pp. 9–26, 1990.

[42] A. Hiniker, S. Hong, Y.-S. Kim, N.-C. Chen, J. D. West, and C. Aragon, "Toward the operationalization of visual metaphor," *Journal of the Association for Information Science and Technology*, vol. 68, no. 10, pp. 2338–2349, 2017.

[43] M. Stafoggia, A. Lallo, D. Fusco, A. P. Barone, M. D'Ovidio, C. Sorge, and C. A. Perucci, "Spie charts, target plots, and radar plots for displaying comparative outcomes of health care," *Journal of clinical epidemiology*, vol. 64, no. 7, pp. 770–778, 2011.

[44] D. Skau and R. Kosara, "Arcs, angles, or areas: Individual data encodings in pie and donut charts," in *Computer Graphics Forum*, vol. 35, pp. 121–130, Wiley Online Library, 2016.

[45] S. Hong, M.-J. Yoo, B. Chinh, A. Han, S. Battersby, and J. Kim, "To distort or not to distort: Distance cartograms in the wild," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[46] K. Oyibo, R. Orji, and J. Vassileva, "Investigation of the persuasiveness of social influence in persuasive technology and the effect of age and gender.," in *PPT@ PERSUASIVE*, pp. 32–44, 2017.

[47] E. Jonas, S. Schulz-Hardt, D. Frey, and N. Thelen, "Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information.," *Journal of personality and social psychology*, vol. 80, no. 4, p. 557, 2001.

[48] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, pp. 157–164, 2011.

[49] D. M. Oppenheimer, T. Meyvis, and N. Davidenko, "Instructional manipulation checks: Detecting satisficing to increase statistical power," *Journal of experimental social psychology*, vol. 45, no. 4, pp. 867–872, 2009.

[50] J. Saldaña, *The coding manual for qualitative researchers*. Sage, 2015.

[51] C. Plaisant, "The challenge of information visualization evaluation," in *Proceedings of the working conference on Advanced visual interfaces*, pp. 109–116, ACM, 2004.

[52] E. Zgraggen, Z. Zhao, R. Zeleznik, and T. Kraska, "Investigating the effect of the multiple comparisons problem in visual analysis," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[53] L. Battle, R. Chang, and M. Stonebraker, "Dynamic prefetching of data tiles for interactive visualization," in *Proceedings of the 2016 International Conference on Management of Data*, pp. 1363–1375, 2016.