Counteracting Serial Position Effects in the CHOICLA Group Decision Support Environment

Martin Stettinger
Institute for Software Technology
Inffeldgasse 16b/II
A-8010, Graz, Austria
martin.stettinger@ist.tugraz.at

Alexander Felfernig
Institute for Software Technology
Inffeldgasse 16b/II
A-8010, Graz, Austria
alexander.felfernig@ist.tugraz.at

Gerhard Leitner
Institute for Informatics Systems
Universitätsstraße 65-67
A-9020, Klagenfurt, Austria
gerhard.leitner@aau.at

Stefan Reiterer
Institute for Software Technology
Inffeldgasse 16b/II
A-8010, Graz, Austria
stefan.reiterer@ist.tugraz.at

Michael Jeran
Institute for Software Technology
Inffeldgasse 16b/II
A-8010, Graz, Austria
michael.jeran@ist.tugraz.at

ABSTRACT
Decisions are often suboptimal due to the fact that humans apply simple heuristics which cause different types of decision biases. CHOICLA is an environment that supports decision making for groups of users. It supports the determination of recommendations for groups and also includes mechanisms to counteract decision biases. In this paper we give an overview of the CHOICLA environment and report the results of a user study which analyzed two voting strategies with regard to their potential of counteracting serial position (primacy/recency) effects when evaluating decision alternatives.

Author Keywords
Recommender Systems; Group Decision Making; Group Recommendation; Decision Tasks; Decision Technologies

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Modelling Environments

INTRODUCTION
Several decisions in everyday life occur in the context of groups, for example, a decision regarding the restaurant to choose for a dinner with business partners or a decision regarding the cinema movie to watch with a group of friends. There exist various decision biases which can negatively influence the quality of group decisions [11]. Anchoring effects [9] are responsible for decisions which are biased by a shown reference value. For example, the average rating of other users in the context of rating an item in collaborative filtering scenarios [1] or the evaluation of an item by the first group member shown to other group members can trigger anchoring effects [8, 13]. Anchoring biases in collaborative filtering scenarios can be counteracted with an adaptation of the preference acquisition interface, for example, by an adaptation of the underlying rating scale visualization [2]. In the context of group decision scenarios, anchoring effects can be controlled by not completely disclosing the preferences of other group members in early stages of a decision process [8].

Another example of decision biases are decoy effects which denote the fact that irrelevant (inferior) items in an item set significantly influence the selection behavior of users [22]. In the context of recommendation scenarios, decoy effects can occur in the item selection process, i.e., if a set of candidate items is recommended to a user [21]. Decoy effects can be detected and counteracted on the basis of predictive models that estimate dominance relationships between items in a candidate (consideration) set and propose to remove items from the candidate set before showing it to the user [21].

In this paper we focus on primacy/recency effects (serial position effects) that represent situations in which items presented at the beginning and the end of the list are evaluated significantly more often than items in the middle of a list [14]. An explanation of this effect is that users are not interested in evaluating large lists to identify those items that best fit their preferences. Primacy/recency effects are also explained as a cognitive phenomenon since items at the beginning and the end of a list are also recalled more often [6]. In recommendation scenarios these effects have been investigated, for example, by Felfernig et al. [6]. The results of their studies clearly show that item properties shown to a user at the beginning and the end of a recommendation dialog are recalled significantly more often than properties in the middle of a dialog.
There exist a couple of online tools that support group decision scenarios. SMARTOCRACY provides support for voting scenarios in social network contexts where information from the social network is applied to rank recommendations [16]. DOODLE focuses primarily on the aspect of coordinating meetings and does not include additional mechanisms to determine recommendations for groups of users. Similarly, VERN [23] is a tool that supports the identification of meeting times based on the idea of unconstrained democracy where individuals are enabled to freely propose alternative dates themselves. DOTMOCRACY deals with larger groups of users and provides a method for collecting and visualizing group preferences. The system is based on the idea of participatory decision making – it’s major outcome is a graph type visualization of the group-immanent preferences. Compared to CHOICLA, these tools focus on specific scenarios, i.e., do not allow a flexible definition of decision functionalities depending on the scenario at hand. Furthermore, existing tools do not include recommendation and explanation functionalities which can help to increase trust in recommendations [7] and improve the perceived decision support quality [8].

Typical CHOICLA scenarios range from industrial settings (e.g., selection of conference locations, selection of new employees, and evaluation of project proposals) to decision scenarios in private settings (e.g., selection of a restaurant for a dinner with friends, selection of a hotel for a holiday trip with friends, and the selection of a movie to watch with friends in a cinema). The contributions of this paper are the following. (1) we present CHOICLA which is a novel domain-independent decision support tool for groups of users and (2) we show how to counteract primacy/recency effects occurring in the evaluation of decision alternatives.

The remainder of this paper is organized as follows. In the Section Choicla Environment we provide insights to the CHOICLA design process where users (creators of decision apps) can model decision tasks from scratch. In Section Choicla Decision Apps we provide an overview of the intelligent management of already created decision apps. In Section User Study we report the results of an empirical study which focused on (1) usability aspects and (2) possibilities to counteract serial position (primacy/recency) effects in the evaluation of decision alternatives. Finally, we discuss issues for future work and conclude the paper.

CHOICLA ENVIRONMENT

Configurability of Decision Apps

Since decision scenarios differ in terms of their process design, a variety of parameters (features) is needed to make decision processes configurable (see Figure 1). CHOICLA offers a variety of features that represent user interface elements (and corresponding functionalities) that can be included in a decision process. For example, the creator of a new CHOICLA decision app can select (set) the feature \( f_{21} \) which indicates that in the new decision app preferences of other users will be shown in terms of a basic summary (individual user preferences will not be displayed).

In order to make decision processes (and corresponding user interface elements) configurable in an intelligent fashion, a configuration model is needed that expresses the set of features and their relationships [4, 20]. This configuration model is an integrative part of CHOICLA. An example of such a relationship is the requires constraint between the features \( f_{14} \) and \( f_{15} \), expressing the fact that the selection of the first one also requires the selection of the latter. An example fragment of the CHOICLA feature model is depicted in Figure 1. We will not discuss this feature model in detail but rather explain major properties on the basis of selected features. An example screenshot of the CHOICLA user interface for defining the parameters of a decision app is depicted in Figure 2.

CHOICLA includes different types of group recommendation heuristics that can be selected to be used within the scope of a group decision process. Figure 1 includes the heuristics Majority and Maut (see [5]). The first one is a basic implementation of majority voting, the second one implements Multi Attribute Utility Theory (MAUT) for group-based settings. Both approaches to aggregate group preferences are discussed in Subsection Recommendation Support. Solution alternatives can be managed in different ways, for example, if feature \( f_9 \) has been selected, only the creator of the corresponding CHOICLA decision app is allowed to add decision alternatives. The scope of a decision app (see also Section Choicla Decision Apps) can be public or private. If it is private, the app cannot be reused by other other users, otherwise it can be reused.

Recommendation Support

CHOICLA includes different group recommendation heuristics (e.g., majority voting and group-based MAUT) that can foster consensus in group decision making (see, e.g., [8, 12]). The degree of perceived decision quality of the participants can be significantly increased by using such aggregation functions – see, for example, [12, 19]. In the following we exemplify group recommendation heuristics (aggregation functions) that can be selected by the creator of a decision app during the corresponding configuration process.

For the explanation of CHOICLA aggregation functions we use a working example which is related to the task of restaurant selection by a group of friends. Table 1 represents the individual restaurant ratings of the participants. For these ratings, group recommendations can be derived on the basis of the CHOICLA group decision heuristics (see Table 2).

<table>
<thead>
<tr>
<th>restaurant</th>
<th>Martin</th>
<th>Rene</th>
<th>Philip</th>
<th>Stefan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphrodite</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Zeus</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Hermes</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Poseidon</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1. Examples of individual user ratings with regard to the available decision alternatives (restaurants).

Majority Voting (see Formula 1) determines the value \( d \) that a majority of the users selected as voting for a specific solution \( s \) where \( \text{eval}(u, s) \) denotes the rating for solution \( s \)

More details on the CHOICLA feature model can be found in [18].
defined by user \( u \). For example, the majority of votings for \textit{Aphrodite} is 4 (see Table 2).

\[
MAJ(s) = \max_{d \in \{1..5\}} \left( \# \left( \bigcup_{u \in \text{Users}} \text{eval}(u, s) = d \right) \right)
\]

(1)

\textit{Least Misery} (see Formula 2) returns the lowest voting for solution \( s \) as group recommendation. For example, the LMIS value for \( s = \text{Aphrodite} \) is 3.
ternatives in terms of there is a need for a more detailed evaluation of decision making scale can be used to evaluate alternatives. In some cases Figure 3 depicts a $C_{\text{predefined}}$ which is different in a typical MAUT scenario. Alternatives – a typical example of such a scenario is the group-based decision regarding a restaurant or a hotel [10]. (3) All decision makers are allowed to add decision alternatives – a typical example of such a scenario is the group-based decision regarding a restaurant or a hotel [10]. (4) Only external users are allowed to add alternatives, i.e., users who do not participate in the decision process. For example, job applicants should be able to add a bundle consisting of the application documents as decision alternative to a (personnel) decision task (the application itself is interpreted as a new alternative - see [17]). Another related example is the selection of next year’s conference location where proposers submit their material in a similar fashion. MAUT-based majority voting for $s = \{\text{MAJ, LMIS, MPLS, GDIS}\}$. For example, the ensemble-based majority voting for $s = \text{Aphrodite}$ is 4.

$$MAUT(s) = \sum_{u \in Users} \sum_{d \in \text{Dimensions}} \text{eval}(u, d) \times \text{weight}(d) / |\text{dimensions}|$$ (6)

If we look at the individual ratings of the dimensions of the restaurant $\text{Zeus}$ in Figure 4 we notice the values 9, 7, 8, and 10. For simplification purposes we assume in our example that all dimensions have the same weight ($wd1 = wd2 = wd3 = wd4 = 5$). The current user’s individual MAUT value of the restaurant $\text{Zeus}$ is $(9 \times 5 + 7 \times 5 + 8 \times 5 + 10 \times 5) / 4$ which results in 42.5. To display the MAUT-based evaluation of a solution (restaurants in our working example), utility (MAUT) values are transformed to a 5-star scale.

CHOICLA also offers a modified version of the Group Distance algorithm (see Formula 8). This modified version of the Group Distance algorithm ($GD'$) takes into account past decision outcomes. A user-weighting $w(u)$ is used to ensure that users whose individual ratings often differ from related decisions (represented by $distance(u)$), are privileged. This user-weighting $w(u)$ considers the total sum of all the distances between the individual ratings of a user to the “final decision” in the particular decision situation (see Formula 7). If a history of past decision outcomes is available, the recommendation is influenced by these user-weightings.

$$w(u) = \sum distance(u) + 1$$ (7)

$$GD'(s) = \min arg_{d \in \{1..5\}} (\sum_{u \in Users} w(u) \times |\text{eval}(u, s) - d|)$$ (8)

### Decision Alternatives

In CHOICLA, there are different ways in which decision alternatives can be entered into the system. (1) only the creator of a decision task can add/modify decision alternatives – this is needed if a person is interested in the opinions of his/her friends about the given alternatives (e.g., alternative candidates for the next digital camera). Another example are so-called “Micro-Polls” where the creator is only interested in knowing the preference distribution of a large group of users. (2) all decision makers are allowed to add decision alternatives – a typical example of such a scenario is the group-based decision regarding a restaurant or a hotel [10]. (3) only external users are allowed to add alternatives, i.e., users who do not participate in the decision process. For example, job applicants should be able to add a bundle consisting of the application documents as decision alternative to a (personnel) decision task (the application itself is interpreted as a new alternative - see [17]). Another related example is the selection of next year’s conference location where proposers submit their material in a similar fashion.

### Evaluation of Decision Alternatives

As shown in Figure 4 we used Ambience, Price, Quality and Location as dimensions in our working example. The stars
in Figure 4 show the average evaluation based on the values given for the different dimensions of the decision alternatives. To keep the screen in Figure 4 comprehensible, per default only the accumulated ratings of the decision alternatives are shown – the detailed evaluation of the dimensions is hidden and only displayed if users click on the corresponding alternative. The tab Places contains, if available, the geographical information of the decision alternatives. The third tab Group Preferences only shows up if a (predefined) threshold in terms of the number of participations is reached. It displays the actual group recommendation based on the selected recommendation heuristic (see Section 2) in terms of a bar chart. This threshold prevents the participants from statistical inferences to the individual user preferences, especially if no complete preference visibility is wanted at all. Figure 5 visualizes the current group recommendation.

**Explanations**

To increase the trust of the participants in the outcome of a decision process, explanations are indispensable [7]. During the design of a CHOICLA decision app explanations can be selected as a feature (see Figure 1).
CHOICLA supports explanations by allowing the creator of a decision task to attach argumentations as to why an alternative has been selected as group decision. If this feature is selected, the creator has to enter an explanatory text, if not, the entering of such a text remains just an option. If the creator selects the recommended alternative as the "final decision", CHOICLA generates an explanation according to the selected recommendation function. For example, if majority voting is selected, CHOICLA generates the following explanatory text for the group recommendation Zeus: 5-star evaluations were selected more often than other possible evaluations.

Preference Visibility

The insight to the individual preferences of all participants involved in a decision process can have a significant impact on decision quality (see [10] and [15]). There exist decision tasks where the detailed insight into the preferences of all participants is an advantage and of course others where an opposite effect can occur. If, for example, the decision task deals with finding an appointment for a management meeting it is essential to find a date where all heads of different departments can attend the meeting and therefore it is important to know the individual preferences of the participants.

In some decision scenarios full preference visibility can lead to disadvantages for some participants but some kind of transparency of the preferences is helpful to achieve a reasonable decision. In such cases CHOICLA can support the decision makers by offering the possibility to display solely a summary of all given preferences regarding an alternative. A summary prevents all participants from statistical inferences but still can help participants who are unsure about their preferences. All the mentioned features (see Figure 1) can be configured during the modeling process of a decision app. The interface of a CHOICLA app is automatically generated conform the selected features (see Figure 6).

CHOICLA DECISION APPS

After all features relevant for a new CHOICLA decision app have been selected, the corresponding decision app can be generated automatically. Figure 6 depicts the personal home screen of a CHOICLA user who created four group decision apps (selecting a location, determining an appointment, selecting new hardware, and deciding about new employees). Within each of these apps a user can create new decision tasks (instances) and invite (potential) group members (participants) via email. After a participant has accepted the invitation, the decision app also shows up on the personal home screen of the participant.

The tab ‘DecisionApp Store’ (see Figure 6) contains all publicly available decision apps. Decision apps can be searched and directly installed to the individual home screen. This mechanism can save a lot of effort for very common decision tasks because the creator can reuse/modify an already created decision app instead of creating a new one from scratch. This kind of reuse makes the interaction more intuitive and can also reduce barriers for using CHOICLA. The third tab ‘Create DecisionApp’ offers the possibility to create a completely new decision app from scratch.

USER STUDY

We conducted a user study with the goals (1) to evaluate the usability of the current version of CHOICLA and (2) to investigate whether specific evaluation methods for decision alternatives can counteract serial position (primacy/recency) effects. In this context, the term ‘item’ refers to individual arguments/explanations in descriptions (the difference between descriptions shown to user is the ordering of argumentations, i.e., the same individual argumentations were included in both descriptions). In this study we made CHOICLA accessible to N=44 persons outside our university who were interested in applying a group decision support software but were not aware of the fact that the system has been developed at our university. The participants (34\% female, 66\% male) had no computer science oriented technical background knowledge.

The participants of the study were organized into groups of 5–6 persons who had the task of evaluating restaurants they would like to visit for a dinner. For this study, the restaurants
were made anonymous, i.e., the names used in our working example (Hermes, Aphrodite, Poseidon, and Zeus) were not disclosed. N=23 participants (groups 1–4: 3 groups with 6 participants and 1 group with 5 participants) had to articulate their preferences on the basis of a 5-star scale, the remaining participants (N=21, groups 5–8: 1 group with 6 and 3 groups with 5 participants) rated the restaurants with regard to the predefined MAUT dimensions (Ambience, Price, Quality and Location). We anonymized all restaurants in terms of their name and also made no statement on the type of food (for example “italian”, “asian”, “steak”, ...) to prevent a bias triggered by favorite food tastes. Our major goal was to figure out whether the evaluation based on MAUT [5] can counteract primacy/recency effects (if such effects exist). Before evaluating an alternative (restaurant), each participant had to read the corresponding description which focused on the four mentioned interest dimensions (see Figure 4).

<table>
<thead>
<tr>
<th>solution</th>
<th>5-star scale</th>
<th>MAUT dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>groups 1–2</td>
<td>ns (Description 1)</td>
<td>-</td>
</tr>
<tr>
<td>groups 3–4</td>
<td>ps (Description 2)</td>
<td>-</td>
</tr>
<tr>
<td>groups 5–6</td>
<td>-</td>
<td>ns (Description 1)</td>
</tr>
<tr>
<td>groups 7–8</td>
<td>-</td>
<td>ps (Description 2)</td>
</tr>
</tbody>
</table>

Table 3. Study setting: assignment of groups to different types of preference definition support (5-star scale vs. MAUT dimensions) and descriptions of alternatives (ns: negative salient = Description 1 vs. ps: positive salient = Description 2).

For each restaurant, two descriptions were available which only differ in terms of the ordering of individual argumen-
Evaluations (scale: [1..5]) of the decision alternative Zeus on basis of MAUT dimensions. The dotted lines show the average voting value of the alternative. Description 1 states the negative aspects at the very beginning and the end (negative salient) and Description 2 states the positive aspects at the very beginning and the end (positive salient).

An example of a positive salient description (restaurant Zeus) is depicted in Table 4, Table 5 shows the corresponding negative salient description. With the help of these alternative descriptions we wanted to figure out whether primacy/recency effects exist when evaluating decision alternatives. We hypothesized that the higher effort which comes with an evaluation using MAUT helps to counteract primacy/recency effects. This leads to better decision outcomes because biases such as the primacy/recency effects can be significantly reduced if evaluations are performed more accurately.

As shown in Figure 5, every decision alternative is represented by a bar. To make an easy comparison to the individual preferences possible, the group recommendation displays the alternatives in the same order as the individual preferences (see Figure 3). The bars of the chart in Figure 5 contain also the individual rating information of all participants (displayed by moving the cursor over a corresponding bar). This information is only available if the preference visibility feature has been set correspondingly when creating the decision app.

Results of the User Study
The study shows that the restaurants with negative aspects at the beginning and at the end of the description are rated lower than those with positive aspects at the beginning and the end of the description. This was the case when 5-star ratings were used (on an average 1.5 stars less – see Figure 7). The outcome of a two-sample t-test confirms that the two sample sets have different corresponding mean evaluations (p < 0.05). With 5-star rating scales, in 83.2% of the cases restaurants with a positive salient description were chosen.

If MAUT dimensions were used to evaluate the alternatives there was no significant difference between the two versions in terms of the average evaluation (see also Figure 8). In this case 53.6% of the participants selected a restaurant with a positive salient description (46.4% selected a restaurant with a negative salient description).

Consequently, a MAUT-based evaluation of the decision alternatives can counteract serial position (primacy/recency) effects. Figures 7 and 8 show the results of one decision alternative (Zeus) based on a 5-star scale (Figure 7) and on a MAUT dimension evaluation (Figure 8).

After the participants finished their evaluation we asked them whether or not they felt comfortable during the evaluation process on the basis of the systems usability scale (SUS) [3]. This is an important information for us because the MAUT-based evaluation needs more effort and time because all dimensions have to be evaluated separately (see Figure 4).

The usability feedback was very positive and provides a good motivation to continue our work on the extension and improvement of the CHOICLA functionalities. A detailed description of the results and the questionnaire for the evaluation can be found in Figure 9. We did not detect significant differences in the SUS results between the 5-star version and the MAUT-based version. The results indicate that participants spend more time for evaluating alternatives and thus are less susceptible to serial position effects.

CONCLUSIONS AND FUTURE WORK
In this paper we gave an overview of the CHOICLA environment which supports different types of group decision scenarios. Decision apps can be defined on the basis of a feature model that supports the flexible configuration (definition) of decision apps. Within the scope of a user study we detected
primacy/recency effects in the evaluation of decision alternatives and figured out that a MAUT-based item evaluation approach can help to counteract these effects. Our future work will include the development of group decision technologies for complex products and services, the development of further group decision heuristics which especially focus on fostering consensus among group members, and the investigation of further decision biases that play a role in the context of group decision making.
REFERENCES


