Supplemental Material: 
User Preferences for Hybrid Explanations

1 INTRODUCTION
This document serves as supplemental material for the paper “User Preferences for Hybrid Explanations”, by Pigi Kouki, James Schaffer, Jay Pujara, John O’Donovan, and Lise Getoor, published on the 11th ACM Conference on Recommender Systems [1]. In this document, we present the details of the survey presented in the paper, and we encourage the reader to read this document along with the paper. For information about the motivation and the results of the survey, please refer to the paper [1].

2 PRE-SURVEY QUESTIONNAIRES
The title of the survey was “PSLViz! A new way to explore recommendations” which was followed by a short description: “You are about to explore new ways to find interesting items!” Afterwards, we asked some demographic questions:

(1) Enter your Amazon Mechanical Turk id.
(2) What is your age range?
(3) What is your gender?
(4) What is your education level?
(5) What is 4 + 8?
(6) Are you a native English speaker?
(7) Do you have any issues with color blindness?
(8) Which animal is heavier on average, an elephant or a mouse?

Out of the 206 users that participated in the study, 95% of participants were between 18 and 50 years of age (with 5% being above 50) and were 42% male and 58% female. 64% of the users reported to hold a college degree, 33% reported to hold a Master’s or Ph.D. degree, while the rest reported to hold a highschool degree. All users replied correctly to the question number 5 and all users reported to be native English speakers. 15 of the users reported to have issues with color blindness, while 7 of the users replied wrong to the question number 8. Overall, participants spent between 10 and 30 minutes for the study, while 15 people spent 5 or less minutes for the study. We regarded outlier users to be a) users that replied color blindness, 2) users that gave a wrong answer to question number 8, or 3) users that spent 5 or less minutes for the study. For those outliers, we performed separate analysis and we discuss this analysis later in Section 5.

2.1 Familiarity with recommender systems
We asked the users to rate their agreement in a 7-point Likert scale (from “strongly agree” to “strongly disagree”) with each of the above statements about recommender systems. We clarified that recommender systems are systems that generate recommendations on websites such as Netflix, Hulu, Amazon, and so on.

- I am familiar with recommender systems.
- I frequently use recommender systems.
- I often accept recommendations from websites that have recommendation systems.
- I seriously consider recommendations when available.
- Overall, I would be likely to trust a recommender system.
- Overall, I tend to be satisfied with recommendation systems.
- Overall, I think recommendation systems are accurate.

The experimental results did not show any correlation between recommender systems familiarity and user experience.

2.2 Familiarity with data visualization
Next, we asked users to rate their agreement in a 7-point Likert scale with each of the above statements about data visualization. We clarified that data visualization tools could be Matlab, Microsoft Excel, statistics packages like R, or things like parallel coordinate plots, heatmaps, treemaps, or scatter plots.

- I am familiar with data visualization.
- I frequently tabulate data with computer software.
- I have graphed a lot of data in the past.
- I am an expert at data visualization.
- I am a novice when it comes to visualizing data.

The experimental results presented in the paper indicate that visualization familiarity predicts increased user experience in all treatments, apart from one (E-G/R). For more details, please refer to the paper.

3 TREATMENTS
We presented a total of 13 treatments to the users, 9 text-based (Figures 1-8) and 4 visual-based (Figures 9-12). The treatment without the explanation (BASE) was “We recommend that Mary likes Crudo”. We also present again Table 1 which summarizes the Dimension values for each treatment for the different types of explanations tested. For each of the 13 treatments, we showed the following header to the users: “Below, PSLViz is giving a recommendation to Mary. The reasoning process for the recommendation is given through a text explanation (or visual explanation depending on the treatment). In reference to the recommendation above, please rate your agreement with each of the following statements in a 7-point Likert scale (from Strongly Agree to Strongly Disagree).”

- The recommendation process is clear to me.
- I would enjoy using PSLViz if it presented recommendations in this way.
- The recommendation is convincing.
- The recommendation is not well reasoned.
- A recommendation like this might make me change my mind.
- I am unconvinced by this recommendation.
We recommend Crudo to Mary because:

1. Mary’s friend Cindy likes Crudo
2. Mary likes Nopa that is Peruvian like Crudo
3. Some user with similar taste as Mary likes Crudo
4. People who like LaMar, also like Crudo & Mary likes LaMar
5. Crudo is highly rated
6. Mary tends to give high ratings

Figure 1: Text: (E-P)

We recommend Crudo to Mary because:

1. Mary’s friend Cindy likes Crudo
2. Mary’s friend Josh likes Crudo
3. Mary likes Sipan that is also Peruvian like Crudo
4. Mary likes Fresca that is also Peruvian like Crudo
5. People who like LaMar, also like Crudo & Mary likes LaMar
6. People who like Fresca also like Crudo & Mary likes Fresca
7. People who like Mancora also like Crudo & Mary likes Mancora

Figure 2: Text (E-D)

We recommend Crudo to Mary because:

1. Mary’s friends Cindy, Josh, Rosie like Crudo
2. Mary likes Sipan, Fresca, Limon, Catalana that are also Peruvian like Crudo
3. People who like LaMar, Fresca, Mancora, Cocola, Milohas, Hoya, and Milagros also like Crudo & Mary likes LaMar, Fresca, Mancora, Cocola, Milohas, Hoya, and Milagros

Figure 3: Text (E-GD)

We recommend Crudo to Mary because:

1. Mary’s friend Cindy likes Crudo
2. Mary likes Sipan that is also Peruvian like Crudo
3. People who like LaMar also like Crudo & Mary likes LaMar

Figure 4: Text (E)

We recommend Crudo to Mary because:

1. Mary’s friends Cindy and 4 others like Crudo
2. Mary likes Sipan and 3 other restaurants that are also Peruvian like Crudo
3. People who like LaMar and 6 other restaurants, also like Crudo & Mary likes LaMar and these 6 same restaurants

Figure 5: Text (E-G/R)

Since visual representations lacked context, we provided additional information to the users. For the visual depicted in Figure 9 we gave this information: “In the visual, the first column shows people. The second column shows a set of contexts (e.g. people or features), and the third column a set of items. A recommendation is a set of arrows that connects these entities.”

Similarly, for the visual depicted in Figure 10 we gave the information: “In the visual, the first column shows people. The second column shows a set of functions or “rules”, and the third column a set of items. A recommendation is a set of arrows that connects these entities. Here, green arrows represent the recommendation, and red arrows represent the reasoning, or explanation for that recommendation.”

Table 1: Dimension values for each treatment for the different types of explanations tested.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BASE</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>NA</td>
<td>no</td>
<td>NA</td>
</tr>
<tr>
<td>E-G/R</td>
<td>english</td>
<td>no</td>
<td>yes</td>
<td>low</td>
<td>yes</td>
<td>low</td>
</tr>
<tr>
<td>E-G/VD</td>
<td>english</td>
<td>no</td>
<td>yes</td>
<td>high</td>
<td>no</td>
<td>low</td>
</tr>
<tr>
<td>RULE</td>
<td>rule</td>
<td>no</td>
<td>no</td>
<td>low</td>
<td>no</td>
<td>high</td>
</tr>
<tr>
<td>E-W</td>
<td>english</td>
<td>yes</td>
<td>no</td>
<td>low</td>
<td>no</td>
<td>high</td>
</tr>
<tr>
<td>E-P</td>
<td>english</td>
<td>no</td>
<td>no</td>
<td>low</td>
<td>no</td>
<td>high</td>
</tr>
<tr>
<td>E-D</td>
<td>english</td>
<td>no</td>
<td>no</td>
<td>med</td>
<td>no</td>
<td>high</td>
</tr>
<tr>
<td>E-G/D</td>
<td>english</td>
<td>no</td>
<td>no</td>
<td>med</td>
<td>no</td>
<td>low</td>
</tr>
<tr>
<td>E</td>
<td>english</td>
<td>no</td>
<td>no</td>
<td>low</td>
<td>no</td>
<td>low</td>
</tr>
</tbody>
</table>

Visual Style

<table>
<thead>
<tr>
<th>COL</th>
<th>visual</th>
<th>COLumns + pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR</td>
<td>visual</td>
<td>Columns + Pathways with Reasoning</td>
</tr>
<tr>
<td>VEN</td>
<td>visual</td>
<td>VENa diagram</td>
</tr>
<tr>
<td>CC</td>
<td>visual</td>
<td>Concentric Circles</td>
</tr>
</tbody>
</table>

Figure 6: Text: High Info (E-G/VD)

We recommend Crudo to Mary because

1. Mary’s friend Cindy, Josh, Rosie, George, Michael like Crudo
2. Mary likes Sipan, Fresca, Limon, Catalana that are also Peruvian like Crudo
3. People who like LaMar, Fresca, Mancora, Cocola, Milohas, Hoya, and Milagros also like Crudo & Mary likes LaMar, Fresca, Mancora, Cocola, Milohas, Hoya, and Milagros

Figure 7: Text (E-W)
Next, for the visual depicted in Figure 11 we also provided the following explanation text: “The visual is a Venn diagram showing intersection between hybrid recommendation information sources. One is similar users, another is a set of popular items, and the other is a set of predictions from similar users. The recommendation common to all three sources is “Cheviche””.

Finally, for the visual depicted in Figure 12 we presented the additional explanation: In the visual, concentric circles show one active user, receiving a recommendation, with layers around that user showing profile items (inner layer), similar users and recommendations (outer layer).

### 4 RANKING

Additionally, we also studied whether users have a specific preference over the ranking of the different explanation styles. The ranking question was as follows: “Suppose Mary is recommended the restaurant “Crudo” by PSLViz. Below, a list of possible explanations from the recommender system are given (assume all are correct). Please rank the following explanations with the one you prefer the most at the top.”

1. Mary’s friend Cindy likes Crudo.
2. Mary likes Nopa. Nopa and Crudo are Peruvian.
3. Some users with similar tastes to Mary like Crudo.
4. People who like La Mar also like Crudo, and Mary likes La Mar.
5. Crudo is highly rated.
6. Mary tends to give high ratings.

The results of the survey are presented in the main paper [1].
5 RESULTS

The results of the analysis are presented in the paper [1]. Here, we make some additional remarks based on an analysis for the users that we considered to be our outliers in the study. More specifically, for the users who replied wrong to the question number 8 (Section 2) and the users who spent equal or less than 5 minutes for the study, we performed an analysis which showed that their replies do not affect the outcomes of the statistical tests presented in the paper so we did not remove them from the analysis. For the users who reported color blindness, we also performed a separate analysis and concluded that they liked interface q3 slightly more than the mean, with marginal significance (say something more concrete here). The text in that color was shown in a bright purple. Most people may have been slightly turned off by the hue of the text, but color blind individuals would not have experienced this.

6 POST-SURVEY QUESTIONNAIRES

At the end of the survey, we asked users to give us any comments (in free text format) regarding the study. Some of the most interesting answers were the following:

• “I found the diagrams more difficult to understand than the texts, because that is not my background.”
• “I found the recommendation system to be interesting and nice.”
• “I know more details gathered from this survey. It is good experience to know the recommendation.”
• “I would like to have this system very much.”
• “I preferred the visualizations for recommendation systems.”
• “The part where it pointed out mary gives high scores was kinda off putting.”
• “The preference of friend is seen everywhere but friends could have different tastes.”

It is obvious from the above comments that there is a variation on the comments based on personal tastes. For example, some users preferred the visual explanation styles while others preferred the text-based explanations. Additionally, there are some users that do like hybrid explanations since they like to be more involved in the recommendation process. However, there are users that found hybrid explanations difficult to interpret as well as users that do not like to be given the reasons for a proposed recommendation. There were also a couple of interesting comments: one regarding the fact that the explanation based on the user’s rating history was not informative and another one regarding the fact that friends can sometimes have dissimilar tastes. Our next goal is to propose a personalized solution that first distinguish the personal preference of each user and then presents different explanations to each user, e.g. present visual explanations to users that prefer visuals.

REFERENCES