

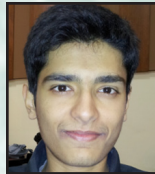


Institute for Software Research
University of California, Irvine

Experimental Materials Used in the Study on Inspectability and Control in Social Recommender Systems



Bart Piet Knijnenburg
University of California, Irvine
bart.k@uci.edu



Nikhil Rao
University of California, Irvine
nsrao@uci.edu



Alfred Kobsa
University of California, Irvine
kobsa@uci.edu

April 2012

ISR Technical Report # UCI-ISR-12-4

Institute for Software Research
ICS2 221
University of California, Irvine
Irvine, CA 92697-3455
www.isr.uci.edu

www.isr.uci.edu/tech-reports.html

Experimental Materials Used in the Study on Inspectability and Control in Social Recommender Systems

Bart P. Knijnenburg
Institute for Software Research
Department of Informatics
University of California, Irvine
bart.k@uci.edu

Nikhil Rao
Department of Informatics
University of California, Irvine
nsrao@uci.edu

Alfred Kobsa
Institute for Software Research
Department of Informatics
University of California, Irvine
kobsa@uci.edu

1. INTRODUCTION

This report has been written with the purpose of explaining the experimental materials and procedures used in the usability study of a hybrid social recommender system as described in “Inspectability and Control in Social Recommenders” by Knijnenburg et al.

2. OVERVIEW OF THE SYSTEM

The study uses a modified version of the TasteWeights system to investigate the effects of inspectability and control on the user experience of social recommender systems. TasteWeights is a social recommender system that uses a user’s Facebook profile to give music recommendations. Specifically, the system uses the overlap between the user’s “likes” and her friend’s likes and then recommends music that is liked by friends that show the most overlap (and is not yet liked by the user herself). A useful feature of TasteWeights is the “recommendation graph”, which shows how the recommendations are connected to the user’s friends, and how these friends are in turn connected to the user’s liked items.

The original TasteWeights system allows users to interactively inspect and control the recommendation graph (i.e. change the weights and inspect the graph simultaneously and iteratively). However, to investigate the effects of inspectability and control independently, we let participants in our experiment interact with the system in two stages: a control stage and an inspection stage. Moreover, in order to attain a consistent experience across study participants, our modified system limits the number of items and friends to be considered by the recommender to 10 each (with a minimum of 5 each). The number of recommendations is fixed to 10.

3. EXPERIMENTAL CONDITIONS

In the experiment we orthogonally manipulated the level of control and inspection in a 3x2 between-subjects design. The control stage is manipulated into three conditions, the inspection stage in two conditions.

3.1 Control manipulation

In the control stage, participants are assigned to one of three conditions (Figure 1):

- **No control:** Participants skip the control stage altogether. The items receive an equal weight. Based on these weights, the recommendation algorithm calculates the friend weights and ultimately the recommendations (as explained in the paper).
- **Item control:** Participants are asked to adjust the weights of the items they “like”. The items are originally weighed equally. After adjusting the weights, the recommendation algorithm again calculates the friend weights and the recommendations.
- **Friend control:** Participants are asked to adjust the weights of their friends. The friend weights are originally calculated by the recommendation algorithm, but in this condition users are allowed to change these weights. After adjusting the weights, the recommendation algorithm calculates the recommendations based on the new friend weights.

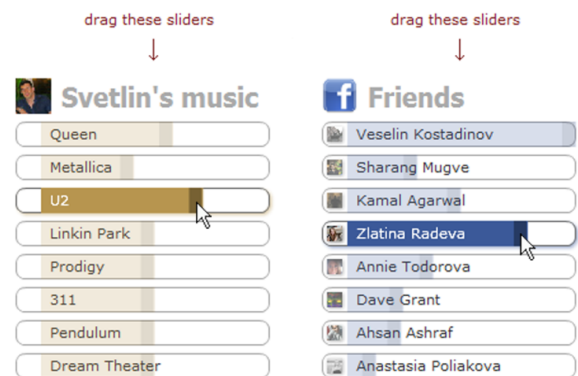


Figure 1. The control phase of item control (left) and friend control (right) conditions.

3.2 Inspection manipulation

In the inspection stage, participants are assigned to one of two conditions (figure 2):

- **List only:** The system shows only the list of recommendations.
- **Full graph:** The system shows the “recommendation graph”, which allows participants to inspect how their items and friends are related to the recommendations.

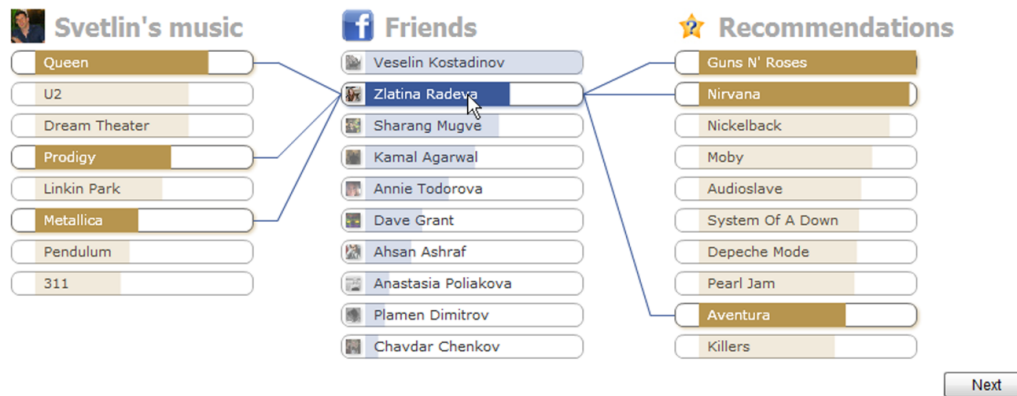


Figure 2. The inspection phase of the “full graph” condition. The “list only” condition shows the rightmost column only.

4. PARTICIPANTS AND PROCEDURE

A total of 267 participants were recruited for this study using Craigslist and Amazon Mechanical Turk. Only adult (aged 18 or older) Facebook users living in the United States were allowed to participate in the study. We also carefully inspected the time participants took to walk through the experiment, and their answers to our “trick” questions (see below) to filter out bad participants. Table 1 shows the demographical distribution of the participants.

Table 1. Distribution of demographics among the participants.

Demographic	Distribution	
Source	Craigslist	22
	Amazon Turk	245
Gender	Female	156
	Male	111
Age	18-20	28
	20-25	102
	25-30	63
	30-40	51
	40-50	19
	50-60	4
	> 60	0

The 22 Craigslist participants were given a 1-in-10 chance to win a \$10 Amazon gift certificate. The remaining 245 participants recruited through Amazon Turk received about \$1.10 for their participation¹. As an added incentive, an Amazon Kindle Fire was raffled off among all participants.

In order to provide a meaningful experience, we only allowed users to participate if their recommendation graph would show at least 5 music “likes”, showing overlap with at least 5 friends (each with an upper bound of 10), and resulting in at least 10 recommendations (we always showed exactly 10 recommendations). Denied participants were given the suggestion to populate their Facebook profile with more music “likes” and then try again.

¹ 10 participants received \$1.80 and 6 participants received \$1.00, as we tried to estimate the optimal compensation for our study.

Eligible participants first answered two “trick” questions to prevent attacks from robots as well as to filter out insincere participants:

- “To confirm your careful reading skills, please answer ‘completely agree’ below”
- “To make sure that you are not a robot, please type the name of the system you are about to evaluate below”

Participants were then asked to answer 15 questions about their personal characteristics (music expertise, trusting propensity and choice persistence). Questions were statements to which participants could agree or disagree on a five-point scale.

Participants subsequently completed the control stage (unless they were assigned to the “no control” condition), in which they were asked to adjust the weights of either their items or their friends (depending on the control condition). Next, they completed the inspection stage, where they were asked to carefully inspect the list of recommendations or the recommendation graph (depending on the inspectability condition). Finally, participants were asked to indicate whether they already knew the band/artist or not, and to rate each recommendation on a 5-star scale. To help their evaluation, they were given a short description and a link to the band/artist’s page on LastFM.

After the experiment, participants answered another 29 questions about their user experience (perceived recommendation quality, perceived control, understandability, system satisfaction) and their familiarity with recommender systems like TasteWeights.

Finally, participants answered a number of demographics questions (gender, age, city, country of birth) and were given the option to provide their email address in order to enter in our raffle for the Kindle Fire.

5. QUESTIONNAIRES

Table 2 shows the questions used in the pre- and post-experimental questionnaires of the study. The order is the same as the order of presentation in the experiment. Questions that were dropped from our analysis are displayed in gray, and do not have a factor loading. As can be seen, choice persistence was not used at all. For the resulting factors that were used in our Structural Equation Model (see the paper), Cronbach’s Alpha and the average variance extracted (AVE) are listed. Correlations between factors are displayed in Table 3. Finally Figure 3 shows the correlation between the factors and the 3x2 experimental conditions.

Table 2. Questionnaire items used to measure participants' personal characteristics and user experience. Items that contribute to a modeled factor have a factor loading.

Considered Aspects	Items	Factor Loading	Response Frequencies				
			-2	-1	0	1	2
Music Expertise Alpha: 0.74 AVE: 0.627	Compared to my peers I listen to a lot of music.	0.871	0	16	33	114	104
	Compared to my peers I am an expert on music.	0.773	7	36	64	108	52
	I only know a few bands/artists.	-0.658	138	89	10	22	8
	I am a music lover.	0.847	0	0	5	64	198
Trusting Propensity Alpha: 0.80 AVE: 0.657	In general, people really do care about the well-being of others.	0.814	4	24	48	151	40
	The typical person is sincerely concerned about the problems of others.	0.820	5	52	52	145	13
	Most of the time, people care enough to try to be helpful, rather than just looking out for themselves	0.797	5	37	55	151	19
	Most people are honest in their dealings with others		7	49	80	119	12
	There aren't many people you can really trust		11	83	56	95	22
	My typical approach is to trust new acquaintances		13	74	53	100	27
Choice persistence	I am willing to examine product attributes very carefully in order to make sure that a product fits my preferences perfectly.		0	7	11	108	141
	I am not easily satisfied with a product.		8	69	73	87	30
	I waste as little time as possible comparing products.		59	113	45	37	13
	When shopping, I have a hard time finding clothing that I really love.		16	40	28	110	73
	When I am in the car listening to the radio, I often check other stations to see if something better is playing even if I'm satisfied with what I'm listening to.		18	49	27	94	79
Perceived Recommendation Quality Alpha: 0.90 AVE: 0.738	I liked the artists/bands recommended by the TasteWeights system.	0.949	6	30	27	125	79
	The recommended artists/bands fitted my preference.	0.950	10	30	24	123	80
	The recommended artists/bands were well chosen.	0.943	10	35	26	101	95
	The recommended artists/bands were relevant.	0.804	4	18	14	120	111
	TasteWeights recommended too many bad artists/bands.	-0.698	104	88	45	20	10
	I didn't like any of the recommended artists/bands.	-0.776	174	61	16	14	2
Perceived Control Alpha: 0.84 AVE: 0.643	I had limited control over the way TasteWeights made recommendations.	0.699	13	52	48	112	42
	TasteWeights restricted me in my choice of music.	0.858	40	90	45	76	16
	Compared to how I normally get recommendations, TasteWeights was very limited.	0.912	36	86	53	68	24
	I would like to have more control over the recommendations.	0.719	8	27	38	130	64
	I decided which information was used for recommendations.		42	82	50	79	14
Understandability Alpha: 0.92	The recommendation process is not transparent.		24	77	76	68	22
	I understand how TasteWeights came up with the recommendations.	0.892	8	41	17	127	74
	TasteWeights explained the reasoning behind the recommendations.		28	59	46	91	43
	I am unsure how the recommendations were generated.	-0.882	71	90	28	62	16
	The recommendation process is clear to me.	0.945	14	65	23	101	64
System satisfaction Alpha: 0.92 AVE: 0.708	I would recommend TasteWeights to others.	0.888	9	32	47	128	51
	TasteWeights is useless.	-0.885	99	106	29	27	6
	TasteWeights makes me more aware of my choice options.	0.769	11	43	56	125	32
	I can make better music choices with TasteWeights.	0.822	12	50	70	95	40
	I can find better music using TasteWeights.	0.890	14	45	62	109	37
	Using TasteWeights is a pleasant experience.	0.785	0	11	38	130	88
	TasteWeights has no real benefit for me.	-0.846	56	91	49	53	18
Familiarity with recommender systems Alpha: 0.86 AVE: 0.794	I am familiar with online recommender systems.	0.842	21	33	24	133	56
	I have occasionally followed the advice of a recommender system.	0.938	25	29	25	127	61
	I have heard of systems similar to TasteWeights.		36	54	48	96	33
	I have never used anything like TasteWeights before.		56	79	24	61	47
	I regularly use systems similar to TasteWeights.		81	88	38	48	12
	Using TasteWeights was a completely new experience for me.		46	67	40	63	51

Table 3. Correlations between the modeled factors.

	Understandability	Control	Quality	Satisfaction	Familiarity	Trusting
Expertise	0.161*	0.171*	0.117	0.006	0.307***	-0.081
Trusting	0.000	-0.056	0.012	0.125	0.052	
Familiarity	0.189**	0.198**	0.081	-0.003		
Satisfaction	0.313***	-0.762***	0.687***			
Quality	0.263***	-0.646***				
Control	-0.326***					

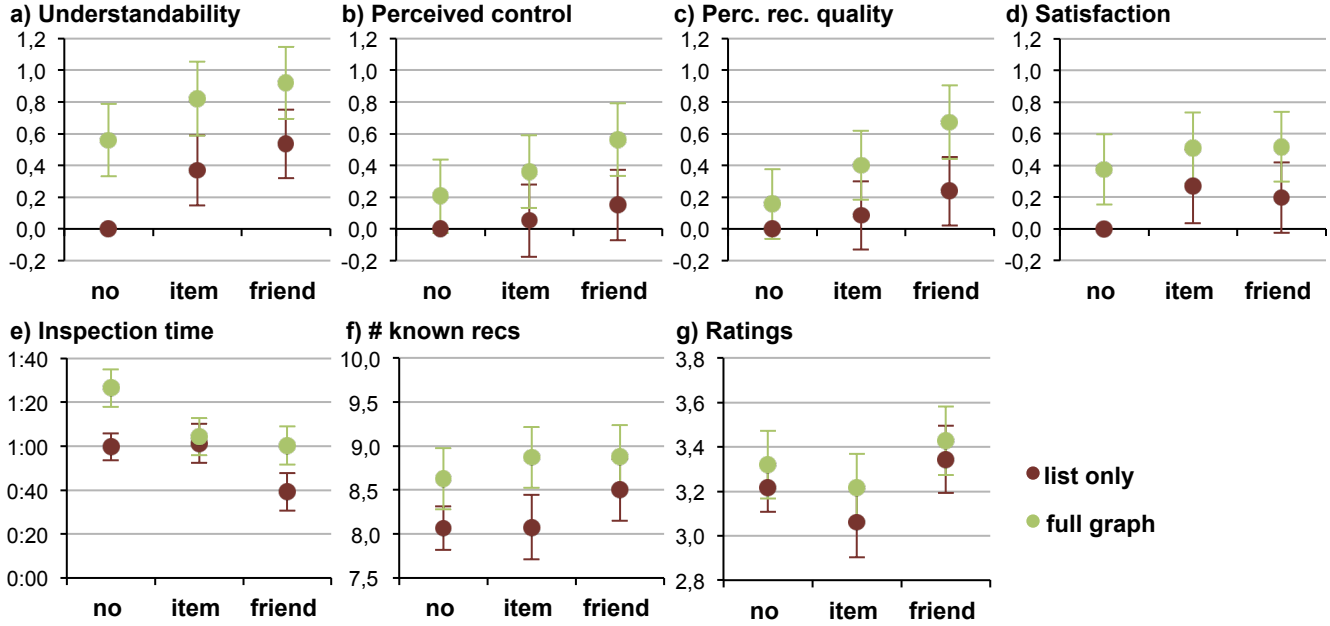


Figure 3. Marginal effects of inspectability and control on the subjective factors (top) and on behaviors (bottom). For the subjective factors, the effects of the “no control, list only” condition is set to zero, and the y-axis is scaled by the sample standard deviation.