Sistemas Recomendadores: Evaluación centrada en el usuario II

Denis Parra

IIC3633

2016

Agenda Semestral

Week	Fecha semana	Clase Martes	Clase Jueves	Presentador 1	Presentador 2	Presentador 3
I	2 - 4 Ago	Intro + CF	CF + Clustering			
II	9 - 11 Ago	CF item-based	Slope One + RecSys			
III	16 - 18 Ago	Evaluacion de RecSys	Evaluacion de RecSys			
IV	23 - 25 Ago	Content-based	Tag-based			
V	30 Ag - 1 Sept	Hybrid	Factorizacion Matricial			
VI	6 - 8 Sept	Context-aware RecSys	Implicit Feedback			
VII	13 - 15 Sept	student presentation (Context, MF	RECSYS Conf	V. Dominguez	J. Schellman	P. Lopez
VIII	20 - 22 Sept	RECSYS Conf	student presentation (IF, MF)	F. Lucchini	V. Claro	V. Castillo
IX	27 - 29 Sept	Presentaciones: Proy. Final	Presentaciones: Proy. Final			
X	4 - 6 Oct	User-centric RecSys/Interfaces	student presentation	J. Lee	C. Kutscher	R. Carmona
XI	11 - 13 Oct	Active Learning/Ranking	student presentation	F. Rojos	J. Navarro	N. Morales
XII	18 - 20 Oct	Graph-based	student presentation	P. Messina	S. Martí	J. Castro
XIII	25 - 27 Oct	Applications: Social/Trust/Music	student presentation	J.M. Herrera	V. Dragicevic	L. Zorich
XIV	1 - 3 Nov	Applications: POI/Tourism	student presentation	I. Becker	T. Hepner	M. Troncoso
XV	8 - 10 Nov	Applications: Educ/Soft.Eng.	student presentation	R. Perez	P. Sanabria	J. Diaz
XVI	15 - 17 Nov	Deep Learning	student presentation	Felipe del Río	L. Pose	G. Sepulveda
XVII	29 Nov - 1 Dic	Presentacion Final	Presentacion Final			

OutLine

- Evaluación centrada en el Usuario:
  - Xiao y Benbasat: Resumen de estudios empíricos sobre "Agentes de Recomendación"
  - Framework I: Resque (Pearl Pu)
  - Framework II: Knijnenburg et al.

Frameworks de Evaluación Centrada en el Usuario

- Xiao y Benbasat (MIS Quartely paper) 2007 (act. 2012)
  - Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: use, characteristics, and impact. Mis Quarterly, 31(1), 137-209.

- Pearl Pu (ResQue) – 2011
  - Pu, P., Chen, L., & Hu, R. (2011, October). A user-centric evaluation framework for recommender systems. In Proceedings of the fifth ACM conference on Recommender systems (pp. 157-164). ACM.

- Bart Kninenburg – 2012
  - Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. User Modeling and User-Adapted Interaction, 22(4-5), 441-504.

Xiao y Benbasat
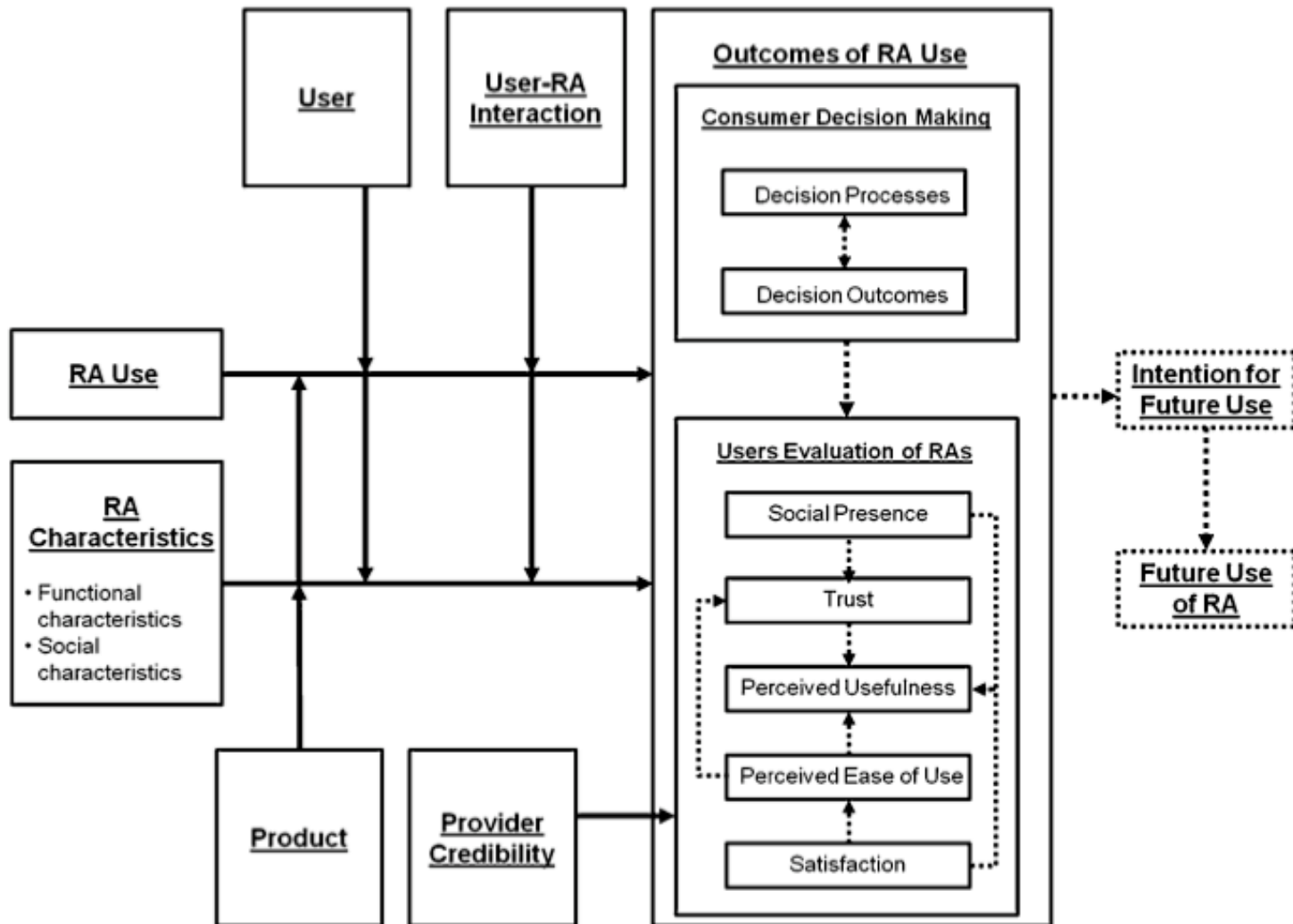
Fig. 2 Updated conceptual model

Resumen de más de 20 estudios

Paper	Type of studyType of RA	Independent variables	Dependent variables
Chang and Chin (2010)	Experiment (lab) RA for mini-notebooks	Recommendation sources: word of mouth (WOM), advertising, or recommendation systems Gender (moderator) Perceived risk (moderator)	Intention to purchase online

Major areas addressed	Major findings
RA use compared to the use of advertising or WOM	A positive recommendation by WOM led to a stronger increase in willingness to purchase online than did advertising and recommendation systems The effect of WOM, advertising, and recommendation systems on online purchase intentions was greater for female consumers, who perceived higher risks in purchasing.

Resumen de más de 20 estudios

Paper	Type of studyType of RA	Independent variables	Dependent variables
Wang and Doong (2010aa)	Experiment (lab) RA for eBooks	Argument form (claim only, claim plus data and warrant, and claim plus data and backing) Spokesperson type (Web itself, expert, customer)	Argument quality Source credibility Purchase intention

Major areas addressed	Major findings
RA output characteristics → explanation	Customers' perceptions of the argument quality and source credibility of the RA's recommendations were found to effectively influence their purchase intentions at the Webstore Customers' perceptions of argument quality and source credibility differed significantly as a result of the varied argument forms Although the various spokesperson types generated significantly different levels of source credibility, argument quality remained unchanged

Framework I - ResQue

- Identifica qué variables (constructos) definen la experiencia de un usuario con un sistema recomendador
- Desarollado en base a modelos existentes para evaluar (TAM y SUMI) y a resultados de estudios relacionados
  - TAM: perceived ease of use of a system, its perceived usefulness and users' intention to use the system
  - TAM v2 (UTAUT): performance expectancy, effort expectancy, social influence, and facilitating conditions
  - SUMI (Software Usability Measurement Inventory) : efficiency, affect, helpfulness, control, learnability

Framework I - ResQue

- Identifica qué variables definen la experiencia de un usuario con un sistema recomendador
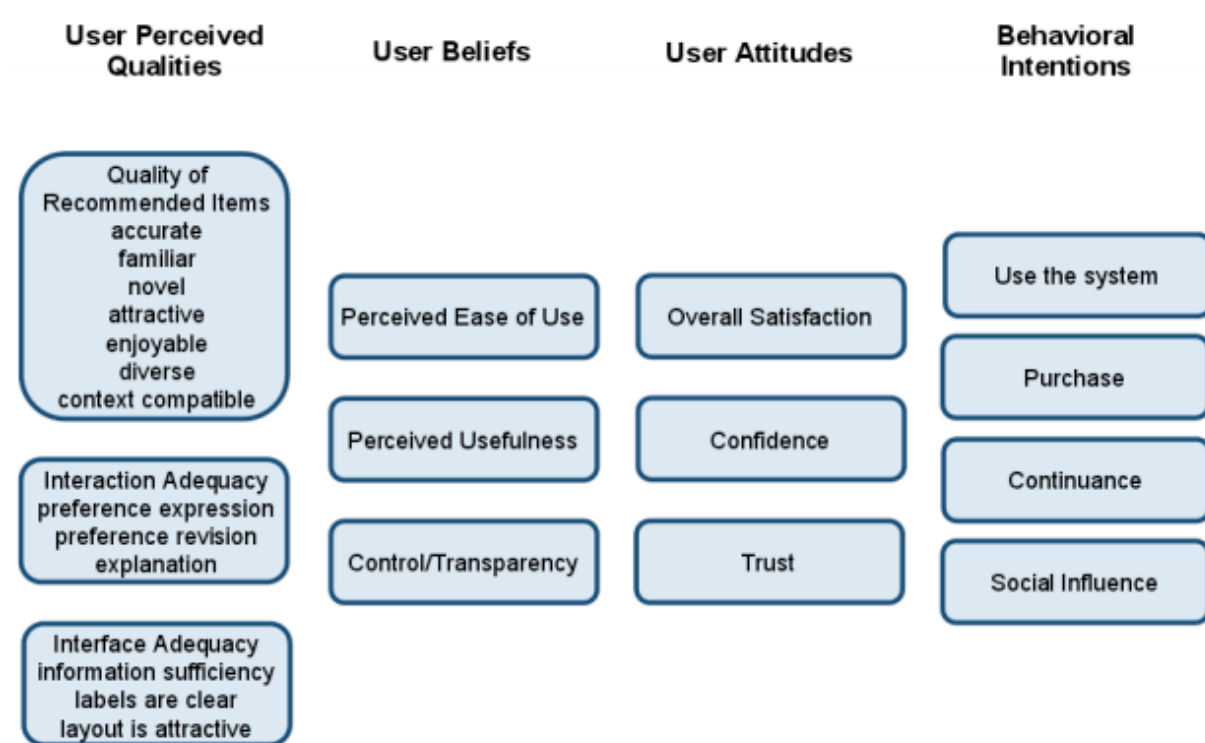
Figure 1: Constructs of an Evaluation Framework on the Perceived Qualities of Recommenders (ResQue).

Encuesta

A1. Quality of Recommended Items

A.1.1 Accuracy

- The items recommended to me matched my interests.*
- The recommender gave me good suggestions.
- I am not interested in the items recommended to me (reverse scale).

A.1.2 Relative Accuracy

- The recommendation I received better fits my interests than what I may receive from a friend.
- A recommendation from my friends better suits my interests than the recommendation from this system (reverse scale).

A.1.3 Familiarity

- Some of the recommended items are familiar to me.
- I am not familiar with the items that were recommended to me (reverse scale).

Encuesta II

A.1.4 Attractiveness

🎬 The items recommended to me are attractive.

A.1.5 Enjoyability

🎬 I enjoyed the items recommended to me.

A.1.6 Novelty

🎬 The items recommended to me are novel and interesting.*

🎬 The recommender system is educational.

🎬 The recommender system helps me discover new products.

🎬 I could not find new items through the recommender (reverse scale).

A.1.6 Diversity

🎬 The items recommended to me are diverse.*

🎬 The items recommended to me are similar to each other (reverse scale).*

Encuesta III

A.1.7 Context Compatibility

🎬 I was only provided with general recommendations.

🎬 The items recommended to me took my personal context requirements into consideration.

🎬 The recommendations are timely.

A2. Interaction Adequacy

🎬 The recommender provides an adequate way for me to express my preferences.

🎬 The recommender provides an adequate way for me to revise my preferences.

🎬 The recommender explains why the products are recommended to me.*

A3. Interface Adequacy

🎬 The recommender's interface provides sufficient information.

🎬 The information provided for the recommended items is sufficient for me.

🎬 The labels of the recommender interface are clear and adequate.

🎬 The layout of the recommender interface is attractive and adequate.*

Encuesta IV

A4. Perceived Ease of Use

A.4.1 Ease of Initial Learning

I became familiar with the recommender system very quickly.

🎬 I easily found the recommended items.

🎬 Looking for a recommended item required too much effort (reverse scale).

A.4.2 Ease of Preference Elicitation

🎬 I found it easy to tell the system about my preferences.

🎬 It is easy to learn to tell the system what I like.

🎬 It required too much effort to tell the system what I like (reversed scale).

A.4.3 Ease of Preference Revision

🎬 I found it easy to make the system recommend different things to me.

🎬 It is easy to train the system to update my preferences.

🎬 I found it easy to alter the outcome of the recommended items due to my preference changes.

🎬 It is easy for me to inform the system if I dislike/like the recommended item.

🎬 It is easy for me to get a new set of recommendations.

Encuesta V

A.4.4 Ease of Decision Making

- 🎬 Using the recommender to find what I like is easy.
- 🎬 I was able to take advantage of the recommender very quickly.
- 🎬 I quickly became productive with the recommender.
- 🎬 Finding an item to buy with the help of the recommender is easy.*
- 🎬 Finding an item to buy, even with the help of the recommender, consumes too much time.

A5. Perceived Usefulness

- 🎬 The recommended items effectively helped me find the ideal product.*
- 🎬 The recommended items influence my selection of products.
- 🎬 I feel supported to find what I like with the help of the recommender.*
- 🎬 I feel supported in selecting the items to buy with the help of the recommender.

Encuesta VI

A6. Control/Transparency

- I feel in control of telling the recommender what I want.
- I don't feel in control of telling the system what I want.
- I don't feel in control of specifying and changing my preferences (reverse scale).
- I understood why the items were recommended to me.
- The system helps me understand why the items were recommended to me.
- The system seems to control my decision process rather than me (reverse scale).

A7. Attitudes

- Overall, I am satisfied with the recommender.*
- I am convinced of the products recommended to me.*
- I am confident I will like the items recommended to me. *
- The recommender made me more confident about my selection/decision.
- The recommended items made me confused about my choice (reverse scale).
- The recommender can be trusted.

Encuesta VII

A8. Behavioral Intentions

A.8.1 Intention to Use the System

🎬 If a recommender such as this exists, I will use it to find products to buy.

A.8.2 Continuance and Frequency

🎬 I will use this recommender again.*

🎬 I will use this type of recommender frequently.

🎬 I prefer to use this type of recommender in the future.
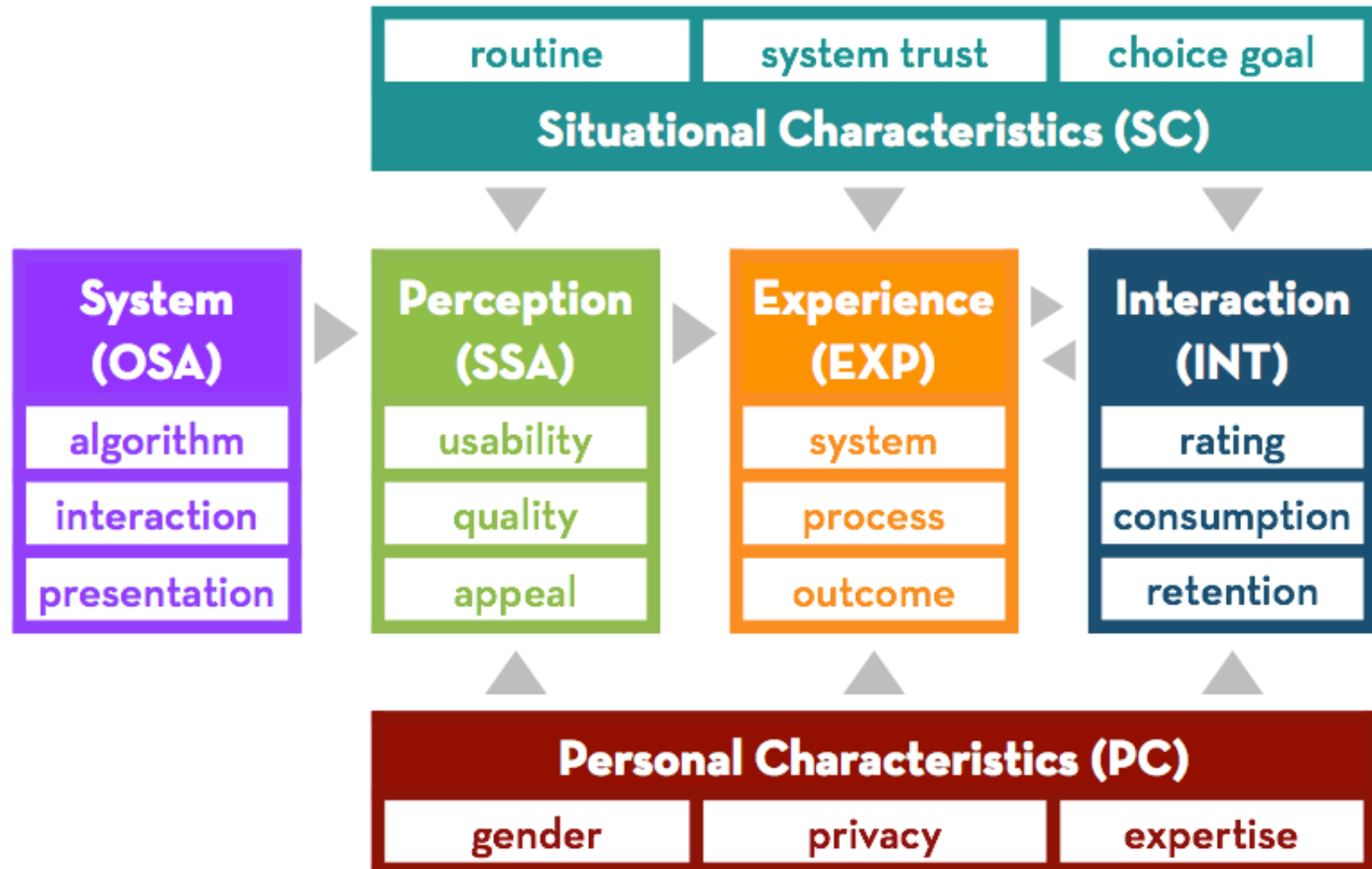
A.8.3 Recommendation to Friends

🎬 I will tell my friends about this recommender.*

A.8.4 Purchase Intention

🎬 I would buy the items recommended, given the opportunity.*

Framework II

Fig. 1 An updated version of the User-Centric Evaluation Framework [61].

Knijnenburg et al.

- En este modelo, el evaluador debe identificar las variables específicas y a qué dimensiones y/o categorías de aspectos corrrespondan.

- Una vez identificadas y medidas, se cotejan con el modelo estructural para ver si corresponden.

# Ejemplo de Aplicación

- Estudio de TasteWeights: Inspectability & Controlability

Figure 1. The TasteWeights system as used in the online user experiment. This is the inspection phase of the "full graph" condition. Users can click on items, friends and recommendations to see the links between them. The inspection phase of the "list only" condition shows the rightmost list (recommendations) only.

Inspectability & Controlability

- Condiciones de Control
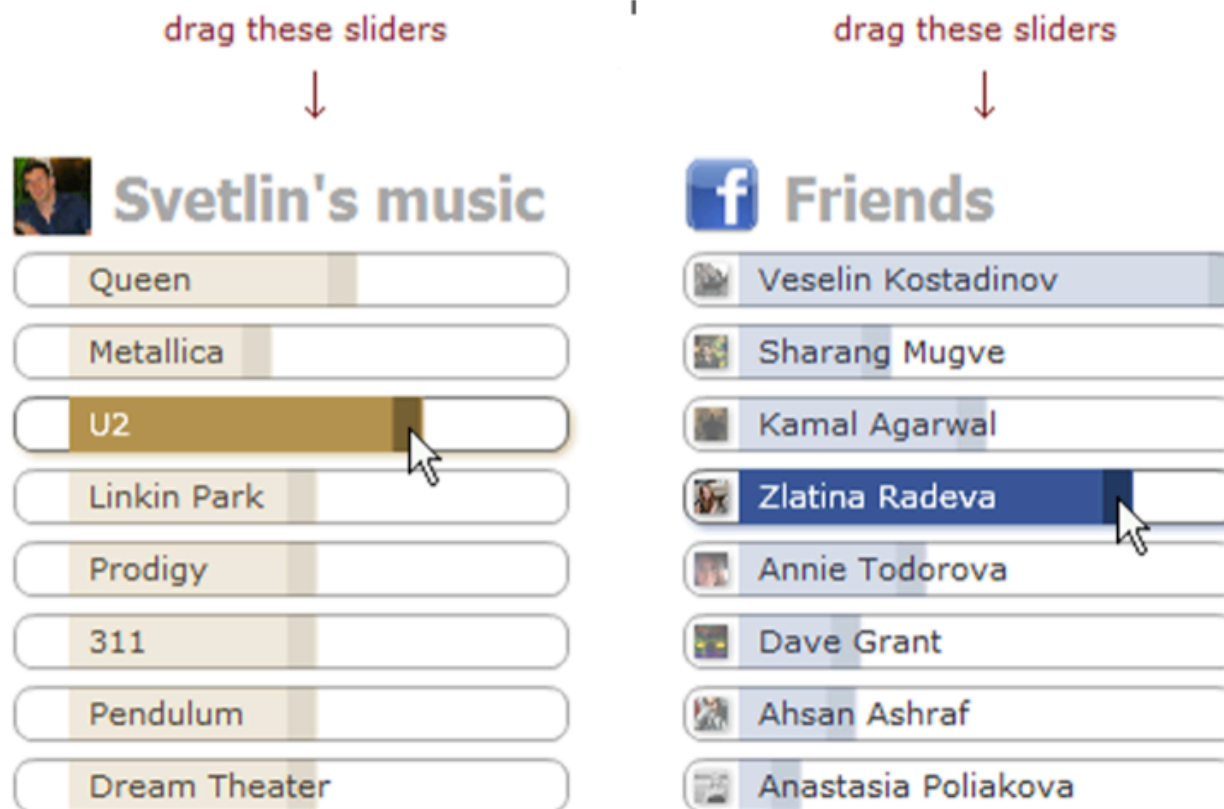
Figure 2. The control phase of item control (left) and friend control (right) conditions.
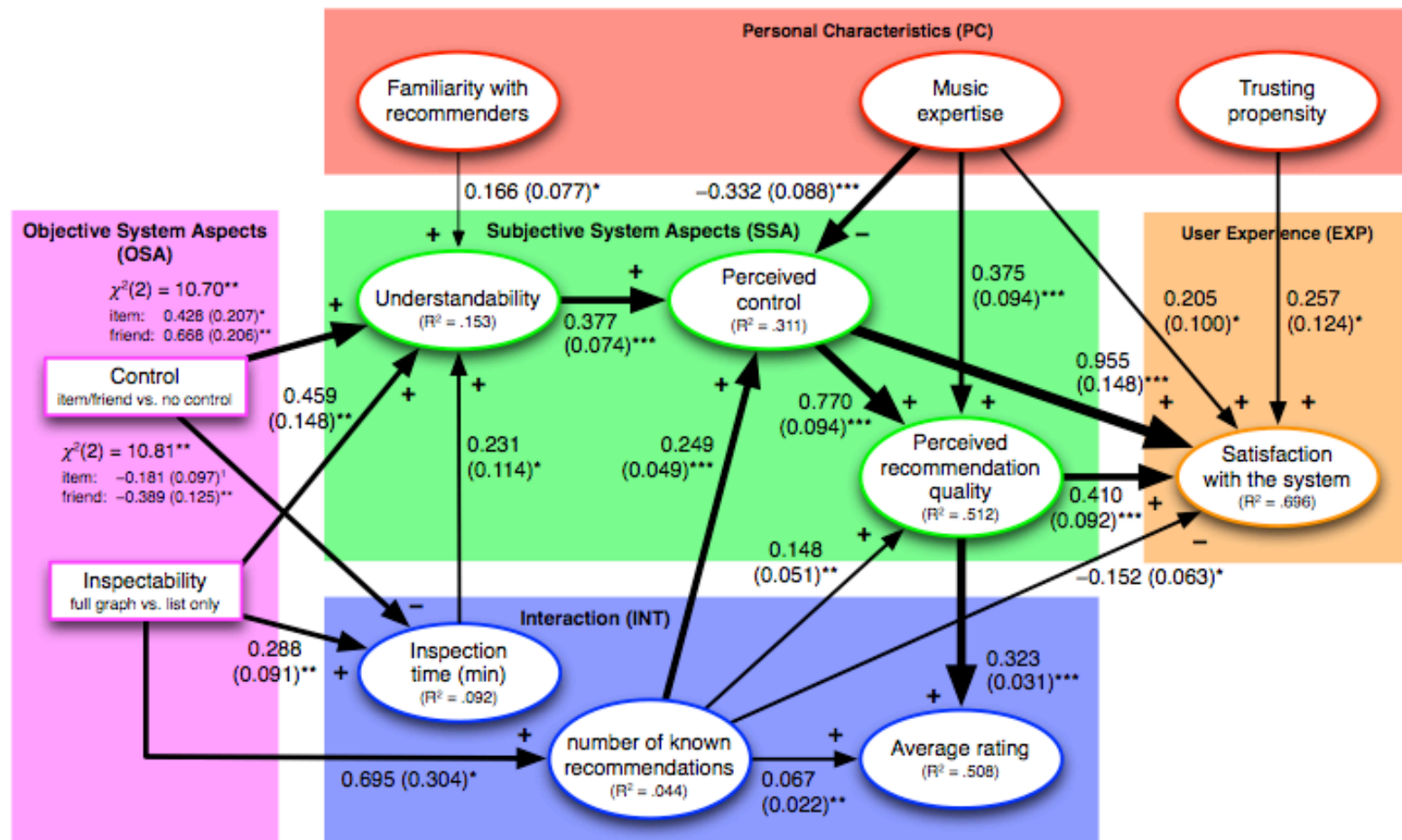
Inspectability & Controlability

Figure 3. The structural equation model for the data of the experiment. Significance levels: *** $p < .001$, ** $p < .01$, 'ns' $p > .05$. R^2 is the proportion of variance explained by the model. Numbers on the arrows (and their thickness) represent the β coefficients (and standard error) of the effect. Factors are scaled to have an SD of 1.

Resultados

- Control e Inspectability tienen un efecto positivo sobre "Comprensión del Sistema" (understandability)
- "Comprensión del Sistema" influye a la vez sobre la "Percepción de Control" (PC) y la "Percepción de Calidad de las Recomendaciones" (PQR)
- PC y PQR influyen sobre la satisfacción final con el sistema
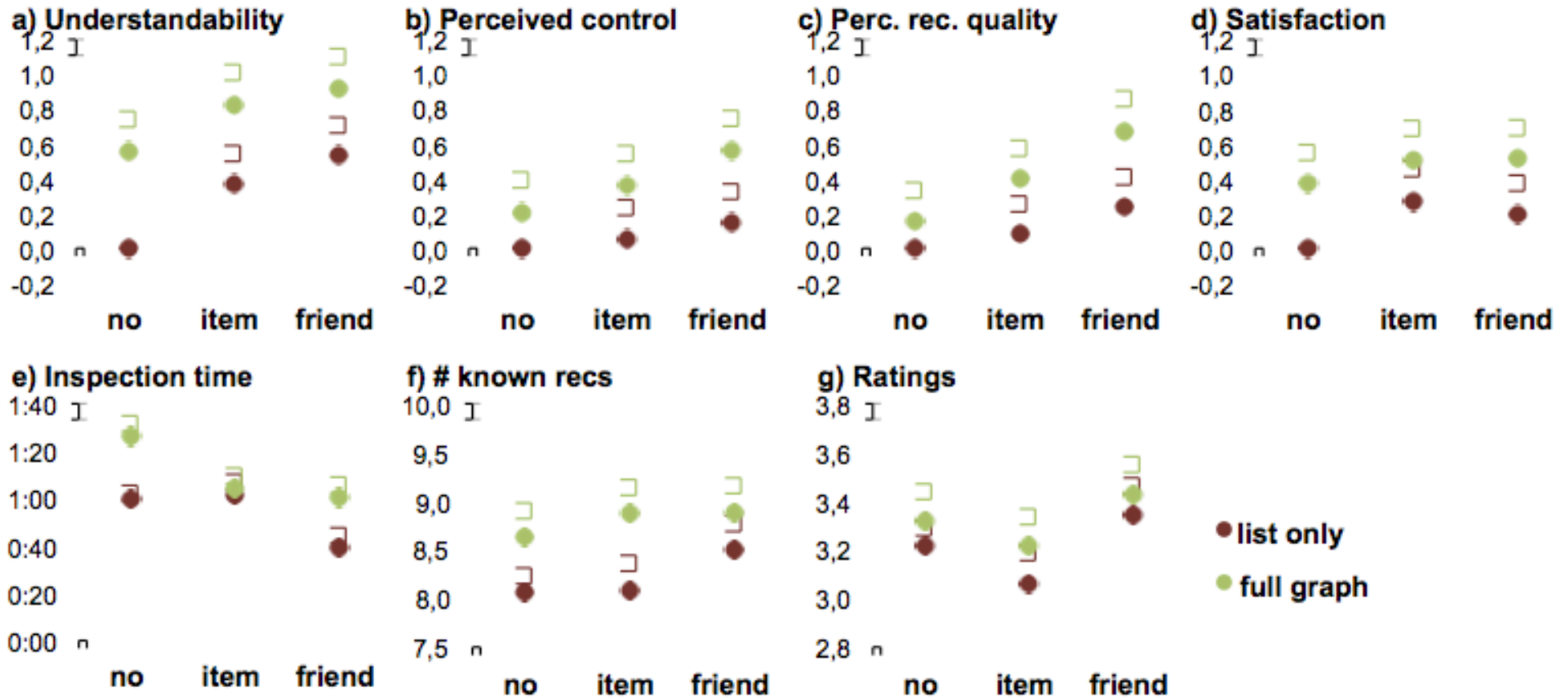
Efectos Marginales

Figure 4. 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.

Resumen de Resultados

- Visualización tipo "grafo de recomendación" mejora la experiencia del usuario al dejarlo inspeccionar las recomendaciones:
  - Comprensión, percepción de control, percepción de calidad de recomendación, satisfacción con el sistema
- Control sobre los pesos de "amigos" produce mayor efecto que control sobre los "items"
- Inspección y control son sumativos: puede incrementar escrutabilidad.