A Cross-Cultural Analysis of Trust in Recommender Systems

Shlomo Berkovsky
Ronnie Taib
CSIRO Australia
Eveleigh, NSW, Australia
first.last@csiro.au

Yoshinori Hijikata
Kwansei Gakuin University
Osaka, Japan
contact@soc-research.org

Pavel Braslavski
Ural Federal University
Yekaterinburg, Russia
pbras@yandex.ru

Bart Knijnenburg
Clemson University
Clemson, SC, USA
bartk@clemson.edu

ABSTRACT

User system trust is critical to the uptake of recommendations, and several factors of trust have been identified and compared. In this paper we present a cross-cultural, crowdsourced study examining user perceptions of nine factors of trust and link the observed differences to trust development processes and cultural dimensions. While some factors consistently instil trust, others are preferred only in certain countries. Our findings and the discovered links are important for design of trusted recommender systems.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Social and professional topics** → *Cultural characteristics*;

KEYWORDS

Recommender systems; user-system trust; presentation of recommendations; cross-cultural comparison; user study.

ACM Reference Format:
Shlomo Berkovsky, Ronnie Taib, Yoshinori Hijikata, Pavel Braslavski, and Bart Knijnenburg. 2018. A Cross-Cultural Analysis of Trust in Recommender Systems. In *UMAP '18: 26th Conference on User Modeling, Adaptation and Personalization, July 8–11, 2018, Singapore, Singapore.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3209219.3209251

1 INTRODUCTION

The success of practical recommender systems largely depends on the uptake of their recommendations. A multitude of factors, associated with the system performance, accuracy, and clarity alike, can potentially affect this uptake. Although some have been studied in depth, factors related to *user-system trust* have received less attention [2, 19, 27]. We argue that the degree of trust a user puts in the recommender plays an important role in the decision making processes related to following the system's recommendations [12].

Many factors and constructs of trust are not specific to recommenders and can be traced to earlier works on user-system trust. Specifically, trust perceptions can be decomposed into the *dispositional* (cultural and demographic), *situational* (context- and task-related), and *learned* (experiential and interaction) factors [10]. Prior work on user trust in recommenders [24, 26] outlines three dimensions of recommendations that affect trust: *presentation* – how the recommendation list is presented to users; *explanation* – what text accompanies the recommended items; and *priority* – what properties of the recommended items are deemed important by the system. Nine distinct, although partially interconnected, factors of trust instantiating these dimensions were synthesised and experimentally compared in [3].

In this work, we set out to investigate the links between the users' cultural group and their preferences towards various constructs of trust. To this end, we conduct a crowdsourced experiment across four different countries, involving in the study more than 460 subjects from France, Japan, Russia, and USA. Measuring six constructs of trust, we analyse the preferences of the subjects for nine interface factors spanning the above three dimensions. While some dominant factors are found to consistently instil user trust across the countries, other factors are preferred and trusted only in certain countries. We attribute the observed preferences to Hofstede's cultural dimensions [11] and overarching trust development processes [8]. Note that our work is independent of the recommendation task and the deployed algorithm, focussing primarily on the recommendation interfaces and human-recommender trust.

Hence, our contributions are two-fold. First, we identify the features of recommendation interfaces that are trusted by users in various countries. Second, we uncover differences across the countries and link them to established cultural dimensions. Our findings are important for practical recommenders, allowing to strengthen user trust and increase the uptake of recommendations.

2 RELATED WORK

The research in human-machine trust has produced various definitions of trust, with one of the most accepted being "the attitude that an agent will help achieve a goal in a situation characterised by uncertainty and vulnerability" [17]. This definition encapsulates the primary sources of variance (user and system) and identifies uncertainty and vulnerability as the pre-conditions of trust.

Three layers of variability in human-machine trust were identified: dispositional, situational, and learned [10]. The first reflects the user's tendency to trust machines due to demographic and personal factors. The second refers to system- and task-specific factors, e.g., the complexity of machine, user's workload, and perceived risks. Finally, the learned trust encapsulates experiential aspects directly related to the system itself. Both [17] and [10] claimed that the former two are likely to be overcome by the latter when the machine exhibits a steady behaviour.

The success of a recommender system largely depends on the uptake of the recommendations, highlighting the importance of trust in recommenders [19]. Several trust factors were taxonomised and compared in [3]. The factors considered were the *presentation*

UMAP '18, July 8–11, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-5589-6/18/07…$15.00
https://doi.org/10.1145/3209219.3209251

of recommendations, i.e., information that accompanies the items [21, 22, 25], *explanations*, aiming to persuade users follow the recommendations [7, 9, 24], and *prioritisation* of the recommendations, i.e., items included in the recommendation list [1, 15, 20]. It was found that quality-based prioritisation of the recommendations and their grouping by a domain feature were trusted most by the users. However, trust perceptions varied across user personalities, which resonates with the individual differences discussed in [10, 17].

In this work, we turn to another factor of the dispositional layer, *cultural differences*. To this end, we conduct the study reported in [3] again, but this time in four different countries. To the best of our knowledge, not much research looked at cross-cultural differences in recommender systems. Namely, only differences in the evaluation of recommendation interfaces [4], preferences for certain items [23], and attitudes towards mobile recommenders [5] have been studied.

3 EXPERIMENTAL METHODOLOGY

We present the factors of trust and dimensions in the context of movie recommendations and outline the design of the user study.

3.1 Dimensions and Factors

Various considerations related to the recommendation lists and their presentation may influence user trust in a recommender. We group them into three dimensions, each including three instantiations referred to as *factors* (more details can be found in [3]).

The *presentation* dimension (see Figure 1) considers three ways to present the recommendations. These are: item grouping according to a certain domain feature (here, by the genre of the recommended movies), use of a humanoid agent to present the recommendations (an image of a person and first-person text), and numeric score communicating the quality of the items (star-rating of the movies). The *explanation* dimension (Figure 2) refers to the text that accompanies the recommendations. The variations of this dimension include persuasive explanations highlighting the advantages of the items (awards or box office figures), personalised explanations listing the reasons for recommendations (list of similar movies liked by the user), and factual explanation (score and number of votes on IMDb). Lastly, the *priority* dimension (Figure 3) deals with the properties of the recommendation list that the system deems important. These are quality (top-scoring IMDb movies), diversity (movies covering many genres), and familiarity (recent movies).

3.2 Study Design and Participants

We conducted a crowdsourced user study comparing these dimensions and factors. The study was divided into two stages. First, the subjects' demographic data was collected and they selected movies they had already watched and liked, to be used for personalised explanations. Second, the subjects were shown nine pages with three lists of movies generated by recommenders denoted A, B, and C. Each page covered a single dimension with its three factors embodied by the recommenders. Sample pages for the three dimension (and three factors each) are shown in Figures 1-3.

The subjects went through nine pages, i.e., three iterations for each dimension. To counter-balance potential order effects, the order of the dimensions and factors on each page was randomised, implementing the factorial design [6]. On each page, the subjects

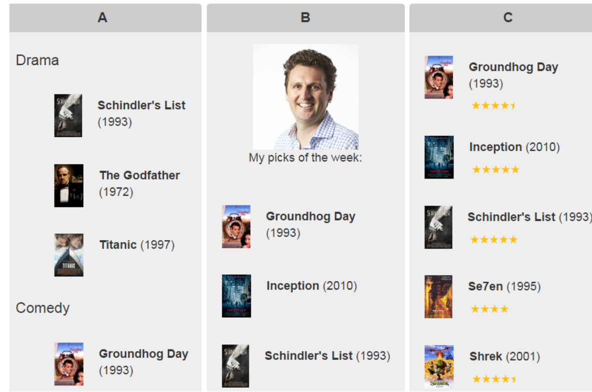

Fig. 1: *Presentation*: genre (A), human (B), star (C).

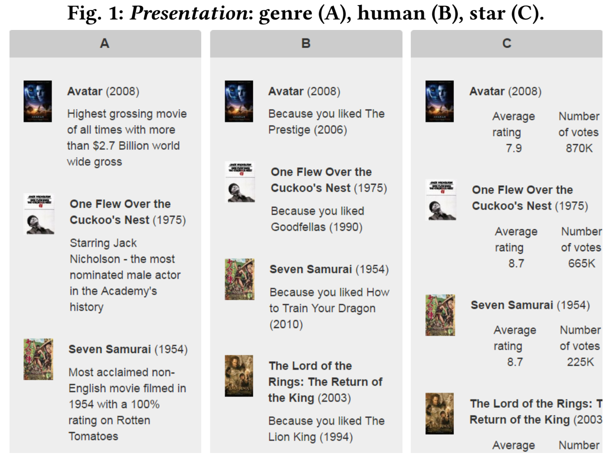

Fig. 2: *Explanation*: persuasive (A), personalised (B), IMDb (C).

Fig. 3: *Priority*: quality (A), diversity (B), familiarity (C).

were asked to select to indicate their preferred list – A, B, or C – with respect to each of the six constructs of trust: *competence* ("recommender most knowledgeable about movies"), *benevolence* ("recommender best reflecting my interests"), *integrity* ("recommender providing most unbiased suggestions"), *transparency* ("I understand the best the reasons for the recommendations"), *intention to re-use* ("for selecting my next movie to watch, I would use"), and *overall trust* ("out of these, the most trustworthy recommender"). The constructs were phrased in a simple language shown in brackets, inspired by the operationalisation of [2].

We instructed the subjects to select their preferred list relying solely on the presentation of the recommendations and disregard the content of the lists, which may naturally contain more or less liked movies. Also, some of the factors listed in Section 3.1 may be dependent, e.g., high-scoring movies may cover multiple genres, boosting quality and diversity. As the lists were pre-compiled, we controlled for this. For example, movies listed in list A in Figure 3 have a higher IMDb rating than movies in lists B and C, while the number of genres and the year of release are comparable [3].

The study was conducted using the CrowdFlower platform. Crowdsourcing increases the risk of collecting spurious data; hence, strict quality assurance policies were implemented [18]. Specifically, responses were rejected on the basis of short completion times (under 5 minutes, half of those in [3]), repeating answers (AA.. or ABAB..), and inconsistent answers (AA.. and then BB.. for the same question).

We conducted the study in four countries, each with the content translated and fully adapted: France (FR, 112 subjects), Japan (JP, 110), Russia (RU, 123), and USA (US, 117). The reported sample sizes represent the data retained after the quality assurance. For the recruitment, we used the location and language of the subjects as inclusion criteria and a coarse-grain proxies of cultural homegeneity. Although some of these countries are multi-cultural on their own, we consider the subjects from each country as the relevant cultural group. Across the countries, we observed a comparable distribution of subjects in terms of gender, age, and IT literacy.

	Compet.	Benev.	Integr.	Transp.	Re-use	Overall
FR subjects (N=112)						
Genre	**60**	**57**	**51**	**55**	**51**	**51**
Human	5	1	7	10	5	3
Star	37	49	43	39	49	49
Persuasive	**88**	32	32	34	34	34
Personalised	4	27	11	28	23	13
IMDb	15	**48**	**63**	**41**	**49**	**55**
Quality	**54**	**55**	**53**	**52**	**52**	**51**
Diversity	13	10	12	7	9	11
Familiarity	18	23	24	23	20	16
JP subjects (N=110)						
Genre	**48**	**55**	**54**	39	**54**	**53**
Human	29	10	12	18	9	11
Star	27	33	38	**47**	44	49
Persuasive	**84**	33	**43**	28	**49**	**61**
Personalised	11	**59**	16	**57**	32	19
IMDb	14	11	44	15	24	24
Quality	**54**	**46**	**41**	**49**	**50**	**49**
Diversity	16	11	19	9	18	18
Familiarity	19	30	25	24	24	24
RU subjects (N=123)						
Genre	**80**	**67**	**55**	**62**	**64**	**65**
Human	5	11	6	11	7	6
Star	31	33	54	43	47	49
Persuasive	**77**	**50**	52	**52**	**53**	**54**
Personalised	14	31	14	28	24	13
IMDb	31	37	**56**	36	43	51
Quality	**72**	**72**	**73**	**73**	**71**	**70**
Diversity	9	8	9	8	9	9
Familiarity	27	22	25	23	28	29
US subjects (N=117)						
Genre	**57**	**60**	**58**	**54**	**62**	**55**
Human	17	16	15	16	12	13
Star	30	28	31	38	34	37
Persuasive	**59**	**40**	36	34	**35**	37
Personalised	16	38	21	**35**	**35**	29
IMDb	27	24	**53**	31	30	**38**
Quality	**55**	**42**	**49**	**51**	**38**	**42**
Diversity	13	11	14	18	15	17
Familiarity	25	30	27	21	37	24

Table 1: User preferences for trust constructs (dominant in bold).

4 RESULTS

We present the results and analyse the differences observed across the countries. The results are summarised in Table 1. Note that we aggregate the results using the "majority voting" of every subject. As mentioned, every dimension was iterated thrice, which is treated as three votes. In the analysis, we consider only the users who preferred one of the factors twice or more. Users, who preferred every factor exactly once, are discarded. For example, out of the 112 subjects in the FR group, 60 subjects preferred twice or more the genre grouping, 5 subjects – human presentation, and 37 subjects – star presentation. Hence, we are left with 102 subjects.

4.1 Preference towards Constructs of Trust

Presentation. The presentation factors are dominated by the genre grouping presentation. In the FR, RU, and US groups, genre grouping achieves the majority of votes for all the constructs. In the JP group, it achieves the majority for five constructs, being inferior only to the star presentation for transparency. Hence, the grouping of movies according to their genres is clearly seen by the subjects as the most trusted presentation of recommendations.

Explanation. On the contrary, few trust votes converge for the explanation dimension. The results agree across the four groups only for the competence construct, where persuasive explanations are preferred. For integrity, two factors dominate: IMDb-based explanations are preferred by the FR, RU, and US subjects, while persuasive explanations are preferred by the JP subjects. Likewise, the JP and RU subjects prefer overall persuasive explanations, while the FR and US subjects prefer IMDb-based explanations. For the remaining constructs of benevolence, transparency, and intention to re-use, no clear trends are observed, with each of the three explanations being preferred in at least one group. In summary, the type of explanation used by a recommender depends on the target construct and on the target population. For example, we posit that persuasive explanations sustain competence, personalised explanations are linked to benevolence and transparency, while IMDb-based explanations promote integrity. Independently of this, persuasive explanations are trusted by the JP and RU subjects, while IMDB-based explanations are preferred by the FR subjects.

Priority. In the priority dimension, like in the presentation dimension, we observe a steady dominance of one factor – quality-based prioritisation of the recommendation lists is preferred over the diversity and familiarity prioritisations. This result is observed across the four groups and with respect to all the six constructs of trust. This indicates that the IMDb scores are perceived by the users to be the most trusted criterion for recommending movies.

4.2 Differences across Countries

We tested whether the subjects' preference within the three dimensions differed across the four countries. The differences were tested for significance using a series of chi-square tests. Where significant differences were found, specific deviations were tested by interpreting the standardised residuals. Given the exploratory nature of our analysis, only significant effects with $p < 0.01$ are reported.

Presentation. Although the genre grouping presentation is generally the most preferred, there are significant differences in the least preferred human presentation factor across the countries. The differences are observed for the competence ($\chi^2(6) = 37.7$, $p < 0.001$) and integrity ($\chi^2(6) = 19.6$, $p = 0.003$) constructs. Specifically, JP subjects are more likely ($p < 0.001$) and RU subjects are less likely ($p = 0.009$) to prefer the human presentation for competence. FR subjects are less likely to prefer the human presentation method for integrity ($p = 0.005$).

Explanation. There are also significant differences in the preferred explanations. These refer to the competence ($\chi^2(6) = 22.3$, $p = 0.001$), intention to re-use ($\chi^2(6) = 19.3$, $p = 0.004$), integrity ($\chi^2(6) = 45.9$, $p < 0.001$), transparency ($\chi^2(6) = 35.1$, $p < 0.001$), and overall trust ($\chi^2(6) = 33.5$, $p < 0.001$) constructs. Here, the differences refer to IMDb-based and personalised explanations. Specifically, JP subjects are more likely to prefer personalised explanations for the integrity ($p < 0.001$) and transparency ($p < 0.001$) constructs, but less likely to prefer IMDb-based explanations for integrity ($p < 0.001$), transparency ($p = 0.008$), and overall trust ($p = 0.008$). FR subjects are more likely to prefer IMDb-based explanations for integrity ($p < 0.001$), while US subjects are more likely to prefer personalised explanations for overall trust ($p = 0.010$).

Priority. In the priority dimension, significant differences are observed across the countries only for intention to re-use ($\chi^2(6) = 16.9$, $p = 0.010$). Subjects in the four countries prefer quality prioritisation, but US subjects are relatively more likely to prefer familiarity prioritisation, although not at the $p < 0.01$ level ($p = 0.040$).

5 DISCUSSION AND CONCLUSIONS

This work compared the perceptions of nine factors of trust in recommender systems across four countries. Some findings were consistent across the countries, e.g., in the presentation and prioritisation dimensions. However, in the explanation dimension the results varied substantially. We will discuss the observed trends and posit potential reasons underpinning the differences.

Presentation. Subjects in all the four countries were most likely to prefer the genre grouping presentation, for all trust factors. However, significant differences were observed in the share of users preferring the human presentation. Specifically, very few FR and RU subjects, and unexpectedly many JP subjects preferred this presentation for competence. Why did we find such differences for competence? National culture may have played a role. According to Hofstede's cultural dimensions [11], Japan scores high on Masculinity (95), while Russia (36) and France (43) score low[1]. People in cultures scoring high on Masculinity are inclined to derive trust from inferred capabilities [8], such that recommender system users may overestimate the capability of a system represented by a humanoid agent [14]. This may potentially make JP subjects more likely, and RU and FR subjects less likely, to trust human presentation. We found that the cultures' Masculinity scores were strongly correlated with the subjects' preference for human presentation for the competence construct ($r = 0.989$, $t(2) = 9.583$, $p = 0.011$), indicating that human presentation of recommendations seems to be trusted most in Masculine cultures.

Explanation. Although in all the countries the subjects were likely to prefer persuasive explanations for competence, preferences varied for the other constructs. In particular, the FR subjects preferred IMDb-based explanations for the other criteria, RU subjects mainly preferred persuasive explanations, JP subjects preferred personalised explanations for integrity and transparency, while US subjects were split among these options. Again, we posit that cultural mechanisms may underpin these preferences. People in cultures scoring high on Masculinity are inclined to derive trust based on a calculative, self-serving process [8]. Hence, personalised explanations may instil more trust in people from Masculine cultures like Japan. Indeed, Masculinity scores of the countries strongly correlated with the subjects' preference for personalised explanations for integrity ($r = 0.969$, $t(2) = 5.475$, $p = 0.031$) and transparency ($r = 0.979$, $t(2) = 6.824$, $p = 0.021$), indicating that personalised explanations appear to be trusted most in Masculine cultures. In contrast, subjects in cultures scoring low on Masculinity are inclined to derive trust based on a transference process [8]. Hence, IMDb-based explanations that present the rating and number of votes by many others, may potentially instill trust in people from Feminine cultures like France and Russia. We found that Masculinity scores of the four countries were strongly negatively correlated with the subjects' preference for IMDb-based explanations for the overall trust ($r = -0.968$, $t(2) = -5.462$, $p = 0.032$), indicating that such rating-based explanations may potentially be trusted most in Feminine cultures.

Priority. Subjects in all the four countries are by far most likely to prefer quality prioritisation for all the trust constructs. Hence, there are no significant differences between the countries, except for the intention to re-use construct.

Implications for Recommenders. These findings are valuable for designers of recommender systems and recommendation interfaces. It is well-known that recommendations should be personalised to their users. This work resonates with previous works that demonstrated that the recommendation interface should also be tailored [13, 16, 21]. We establish that such tailoring should not only be attributed to individual and personality differences [3], but also consider the users' cultural characteristics. Our analysis raises the multi-dimensional nature of trust in recommender systems [27],

[1]The most pronounced correlations referred to Hofstede's dimension of Masculinity. Due to space limitations, the discussion focuses on this dimension only.

which may be interpreted in the context of the desired effect of the recommendations. For example, a thought-through application of presentation, explanation, and priority factors may strengthen the targeted construct of trust, e.g., factual explanations will increase competence and personalised explanations – benevolence. We show how to apply these factors to instil trust and highlight how trust perceptions vary across different groups of users. It is important to note that our work does not fully bridge the gap between individual and cultural features, and this is yet to be addressed both in research and in the design of practical recommender systems.

REFERENCES

[1] Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. 2015. Optimal Greedy Diversity for Recommendation. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI*. 1742–1748.

[2] Izak Benbasat and Weiquan Wang. 2005. Trust In and Adoption of Online Recommendation Agents. *Journal of the Association for Information Systems* 6, 3 (2005).

[3] Shlomo Berkovsky, Ronnie Taib, and Dan Conway. 2017. How to Recommend?: User Trust Factors in Movie Recommender Systems. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*. 287–300.

[4] Li Chen and Pearl Pu. 2008. A cross-cultural user evaluation of product recommender interfaces. In *Proceedings of the ACM Conference on Recommender Systems, RecSys*. 75–82.

[5] Jaewon Choi, Hong Joo Lee, Farhana Sajjad, and Habin Lee. 2014. The influence of national culture on the attitude towards mobile recommender systems. *Technological Forecasting and Social Change* 86 (2014), 65 – 79.

[6] Linda M Collins, John J Dziak, and Runze Li. 2009. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychological methods* 14, 3 (2009), 202.

[7] Henriette S. M. Cramer, Vanessa Evers, Satyan Ramlal, Maarten van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob J. Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455–496.

[8] Patricia M Doney, Joseph P Cannon, and Michael R Mullen. 1998. Understanding the influence of national culture on the development of trust. *Academy of Management Review* 23, 3 (1998), 601–620.

[9] Alexander Felfernig and Bartosz Gula. 2006. An Empirical Study on Consumer Behavior in the Interaction with Knowledge-based Recommender Applications. In *Proceedings of the International Conference on E-Commerce Technology, CEC*. 37.

[10] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57, 3 (2015), 407–434.

[11] Geert Hofstede. 2003. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage Publications.

[12] Anthony Jameson, Martijn C. Willemsen, Alexander Felfernig, Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, and Li Chen. 2015. Human Decision Making and Recommender Systems. In *Recommender Systems Handbook*. 611–648.

[13] Bart P. Knijnenburg, Niels J. M. Reijmer, and Martijn C. Willemsen. 2011. Each to his own: how different users call for different interaction methods in recommender systems. In *Proceedings of the ACM Conference on Recommender Systems, RecSys*. 141–148.

[14] Bart P. Knijnenburg and Martijn C. Willemsen. 2016. Inferring Capabilities of Intelligent Agents from Their External Traits. *ACM Transactions on Interactive Intelligent Systems* 6, 4 (2016), 28:1–28:25.

[15] Sherrie YX Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *Management Information Systems Quarterly* (2006), 941–960.

[16] Branislav Kveton and Shlomo Berkovsky. 2016. Minimal Interaction Content Discovery in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems* 6, 2 (2016), 15:1–15:25.

[17] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80.

[18] Stefanie Nowak and Stefan M. Rüger. 2010. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the International Conference on Multimedia Information Retrieval, MIR*. 557–566.

[19] John O'Donovan and Barry Smyth. 2005. Trust in recommender systems. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*. 167–174.

[20] Umberto Panniello, Michele Gorgoglione, and Alexander Tuzhilin. 2016. Research Note - In CARSs We Trust: How Context-Aware Recommendations Affect Customers' Trust and Other Business Performance Measures of Recommender Systems. *Information Systems Research* 27, 1 (2016), 182–196.

[21] Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*. 93–100.

[22] Guy Shani, Lior Rokach, Bracha Shapira, Sarit Hadash, and Moran Tangi. 2013. Investigating confidence displays for top-*N* recommendations. *Journal of the Association for Information Science and Technology* 64, 12 (2013), 2548–2563.

[23] T. Y. Tang, P. Winoto, and R. Z. Ye. 2011. Analysis of a multi-domain recommender system. In *Proceedings of the International Conference on Data Mining and Intelligent Information Technology Applications*. 280–285.

[24] Nava Tintarev and Judith Masthoff. 2015. Explaining Recommendations: Design and Evaluation. In *Recommender Systems Handbook*. 353–382.

[25] Ye Diana Wang and Henry H Emurian. 2005. An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior* 21, 1 (2005), 105–125.

[26] Kyung Hyan Yoo, Ulrike Gretzel, and Markus Zanker. 2015. Source Factors in Recommender System Credibility Evaluation. In *Recommender Systems Handbook*. 689–714.

[27] Kun Yu, Shlomo Berkovsky, Ronnie Taib, Dan Conway, Jianlong Zhou, and Fang Chen. 2017. User Trust Dynamics: An Investigation Driven by Differences in System Performance. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*. 307–317.