How to Recommend? User Trust Factors in Movie Recommender Systems

Shlomo Berkovsky, Ronnie Taib, Dan Conway
Data61, CSIRO, Australia
firstname.lastname@csiro.au

ABSTRACT
How much trust a user places in a recommender is crucial to the uptake of the recommendations. Although prior work established various factors that build and sustain user trust, their comparative impact has not been studied in depth. This paper presents the results of a crowdsourced study examining the impact of various recommendation interfaces and content selection strategies on user trust. It evaluates the subjective ranking of nine key factors of trust grouped into three dimensions and examines the differences observed with respect to users' personality traits.

Author Keywords
Recommender systems, user-system trust, presentation of recommendations, user study.

INTRODUCTION
The success of recommender systems, especially commercial ones, depends to a large extent on the user's uptake of the recommendations. There are several factors that can influence this uptake, e.g., the accuracy of the recommendations, their freshness, or their potential value for the user [10]. Although these variable have been studied extensively, factors related to *user-system trust* (will be referred hereafter as trust) have received less attention.

We argue that the degree of trust users put in the system plays an important role in the decision making process preceding the uptake of the recommendations [13]. Trust has been shown influential in the broad context of automation and interactions with decision support systems [16, 17, 11] and it has also been considered in the context of recommender systems [41]. For instance, it was found that accuracy and diversification of recommendations positively affect trust, which led to increased customer purchases [22]. Other works established that explanations [12], confidence displays [31], and system transparency [5] also contribute to user trust in recommender systems.

IUI 2017, March 13–16, 2017, Limassol, Cyprus.
© 2017 ACM ISBN 978-1-4503-4348-0/17/03 ...$15.00.
http://dx.doi.org/10.1145/3025171.3025209

It is important to note that user-system trust consists of three layers [11]. Dispositional trust reflects the user's tendency to trust systems and encompasses cultural and demographic factors. Situational trust refers to more specific factors, like the performed task, system complexity, and user's workload. Lastly, learned trust encapsulates the experiential aspects, which develop as the user interacts with the system and forms the perception of its performance. Focussing on the learned trust in the context of recommendation agents, Benbasat and Wang extended models of human-human trust and identified five constructs of user-system trust: competence, integrity, benevolence, transparency, and intention to re-use [1].

In this work, we set out to investigate the dependencies between various aspects of recommendation interfaces and user-system trust. Note that our work does not deal with the recommendation algorithms selecting the items, but rather with the ways these items are recommended. That is, we assume an existing recommendation list and we study the trustworthiness of its presentation. Hence, the insights provided here are independent of the application domain and recommendation task, and they apply to recommendations of items that can be characterised by domain features, such as 'rating', 'popularity', 'category', or 'brand'.

We synthesise prior work on factors of trust in recommender systems [34, 41] and consider three broad dimensions of recommendations that potentially can 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. We identify nine distinct, although partially interconnected, factors of trust and map them onto these three dimensions.

We report the results of a crowdsourced user study that compares the trust instilled by several variations of a movie recommender within each of the three dimensions. During the study, we present these variations to users and ask them to select their preferred one with respect to the constructs of trust. Then, we explain the mechanics of these variations to users and asked to justify their preferences. We analyse the results obtained in the user study and summarise the collected justifications, in order to identify the dominant factors impacting trust. We also present a thorough analysis of the differences

	Low-trait			*High*-trait		
	thr_l	*users*	$\overline{score}$	thr_h	*users*	$\overline{score}$
Extraversion	3.0	32	2.19	4.5	36	4.93
Agreeableness	2.5	36	2.03	4.0	33	4.30
Conscientious.	4.0	31	3.66	6.0	37	6.41
Neuroticism	2.5	38	2.33	4.0	35	5.11
Openness	4.0	30	3.68	6.0	30	6.30

Table 6. Characterisation of the low and high personality trait groups: low/high threshold, number of users, and mean trait score.

When reviewing these findings, we hypothesise that these changes can be attributed to the identical presentation of the three recommendation lists, the content of which solely reflects the prioritisation. This might not have been clear in the ranking phase, while the debrief actually attracted user attention to the subtle differences between the lists and triggered the changes in preferences. Qualitative input illustrates the reasons for the user choices during debrief. For example, a user explains their vote for familiarity through the integrity construct, "*popular recently released movies are always the first I watch*". Likewise, the substantial increase in preference towards diversity in the competence construct is evidently illustrated by comments like "*different genres allow for a wider experience with films and culture, so this leads to most knowledge*" and "*they have a broader knowledge of films, rather than picking what is necessarily the most recent or highest rated*".

Analyses of Personality Traits

We turn now to the analysis of how trust perception varies across different types of users. During the initial profiling phase, we collected data about the users' personality traits through the TIPI inventory tool and calculated the scores of the Big-Five personality traits [9]. Following this, the users were split into *high* and *low* groups with respect to each of the five traits. The thresholds for the high and low scoring groups were adjusted for each trait, in order to balance the sample sizes for the analysis. Table 6 shows the low- and high-trait group thresholds (on a 7-Likert scale of TIPI), number of users in each group, and the mean score of the relevant trait for each group.

Figures 5-9 show pairwise comparisons of the ratios of user preferences for each dimension and factor, with respect to all trust constructs, during the ranking phase only. Each pair of columns presents the low-trait group on the left and the high-trait group on the right. In the following sub-sections we analyse the differences observed between the low- and high-scoring groups for each trait. The discussion primarily focuses on the aggregated trust scores calculated across all the constructs. Significance results are based again on the χ^2 test with Benjamini-Hochberg correction at $Q_e = 0.1$, conducted between the normalised vote counts received in the two groups considered. As earlier, statistically significant differences are marked by *.

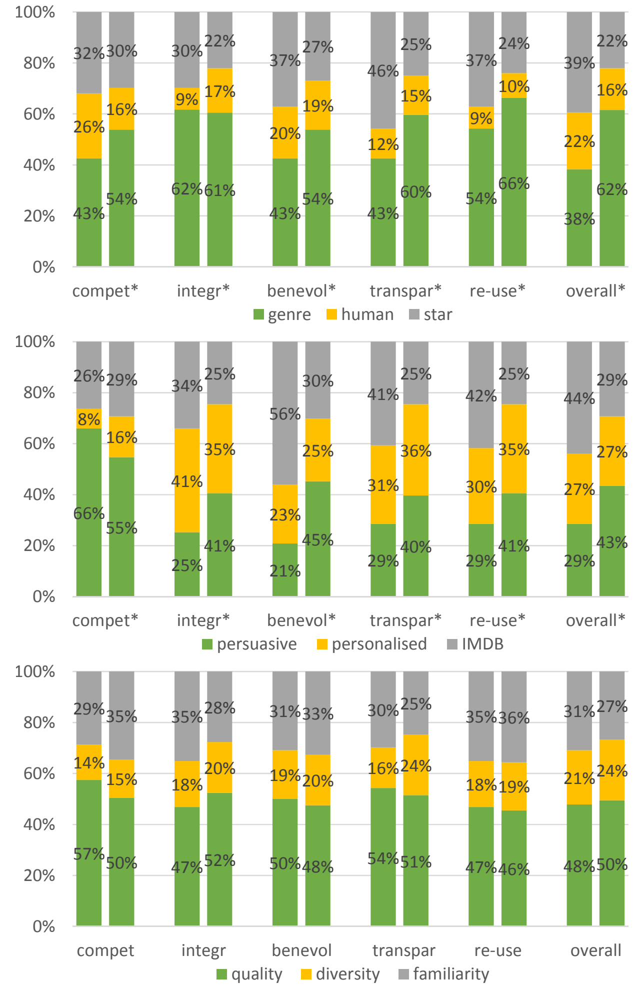

Figure 5. Low- vs high-extraversion users: presentation (top), explanation (middle), and priority (bottom). Significant differences are marked by *.

Extraversion

Looking at the top graph in Figure 5 referring to the *presentation* dimension, the main differences between the low- and high-extraversion users correspond to an aggregated 8.4% increase for the genre grouping and the corresponding 9.0% decrease for the star-rating presentation. These differences are found to be significant in all the constructs of trust. We posit that the increase of the genre grouping reflects the tendency of high-extraversion people to seek stimulation in a breadth of activities, which the genres indirectly reflect. Conversely, reliability is important for low-extraversion people, so they put more trust into the star-ranking presentation.

Turning to the middle graph about *explanation*, we observe an 8.6% increase in aggregated preference towards persuasive explanations, at the detriment of IMDb-based explanations, which drops by 11.0%. These differences are again significant across all the constructs of trust. We relate these changes to the enthusiastic and outgoing nature of high-extraversion people, which is fuelled by

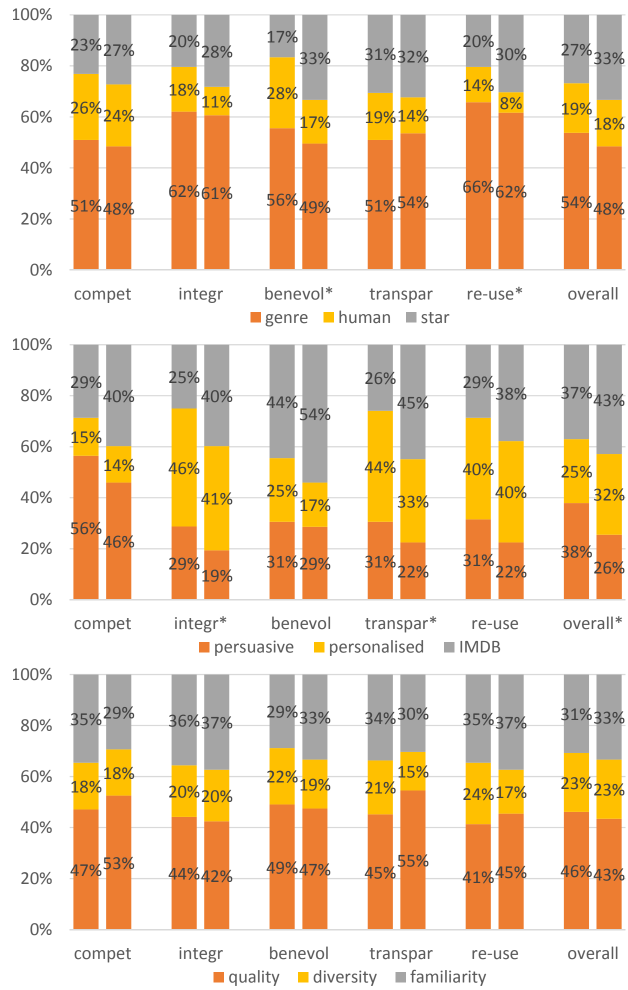

Figure 6. Low- vs high-agreeableness users: presentation (top), explanation (middle), and priority (bottom). Significant differences are marked by *.

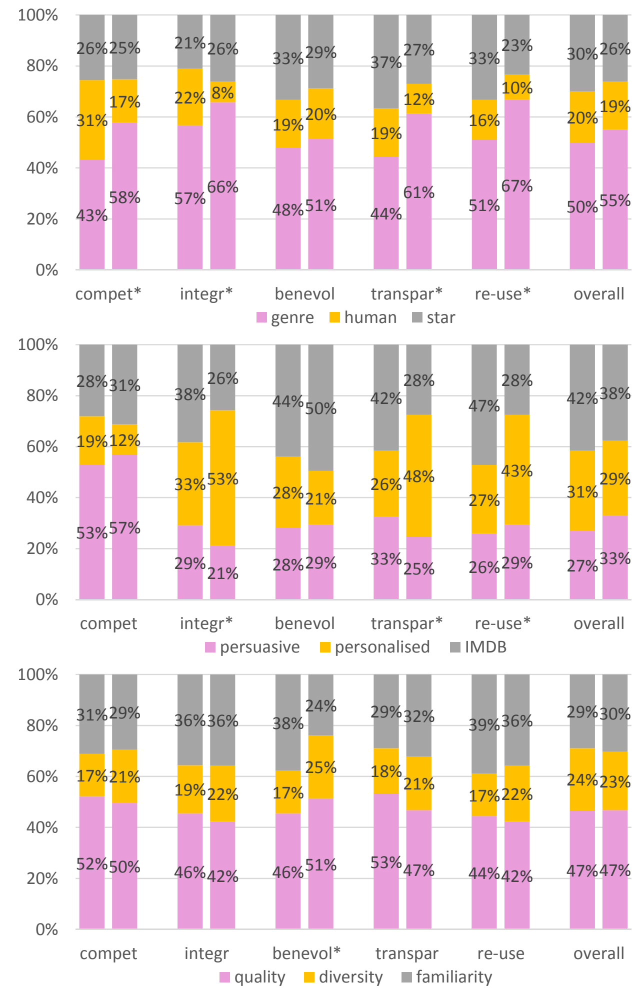

Figure 7. Low- vs high-conscientiousness users: presentation (top), explanation (middle), and priority (bottom). Significant differences are marked by *.

the exciting tone of the persuasive explanations. Again, low-extraversion people trust more the reliability of the IMDb explanations based on thousands of votes. No statistically significant differences are observed in the *priority* dimension, as shown in the bottom graph.

Agreeableness

The differences between low- and high-agreeableness users are shown in Figure 6. In the *presentation* dimension, the main changes relate to a 6.7% increase for star-ranking and 4.8% drop for human presentations. These changes are found to be significant for benevolence and intention to re-use. This is in line with the characterisation of high-agreeableness people as cooperative, compliant, and trusting, so they would rely on the wisdom-of-the-crowds communicated through the star rating. However, note that little change is observed for the genre presentation, which remains dominant in both groups.

More substantial changes in preferences are observed in the *explanation* dimension. Here, we witness a 10.6% increase for the IMDb-based explanations, whereas both persuasive and personalised explanations drop by 6.5% and 4.1%, respectively. These changes are significant with respect to the integrity, transparency, and overall trust constructs. Similarly to the above star-ranking increase, the increase in IMDb explanation can be explained by the tendency of high-agreeableness people to get along with others, accept their opinion, and potentially compromise their own interest; hence, the drop in personalised explanations. Again, no significant differences are observed in the *priority* dimension.

Conscientiousness

Figure 7 shows the differences observed between the low- and high-conscientiousness users. The main difference in the *presentation* dimension relates to a 9.9% increase for the genre grouping, mostly on the account of a 6.7% drop in human presentation. The changes are significant with respect to competence, integrity, transparency, and intention to re-use. Our finding aligns with the organised and orderly nature of high-conscientiousness people, who appreciate the grouping of the movies according to

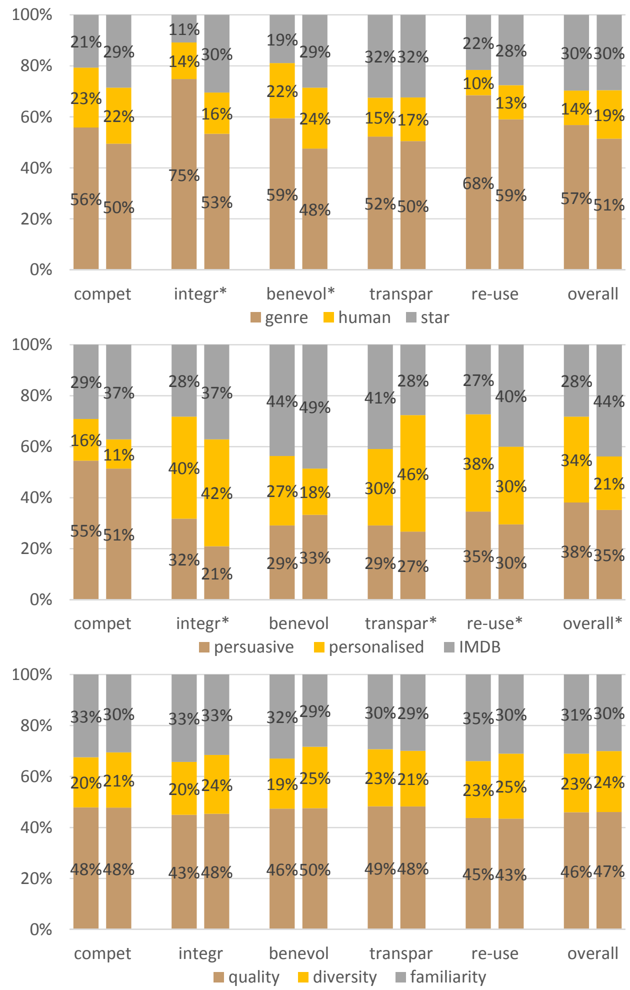

Figure 8. Low- vs high-neuroticism users: presentation (top), explanation (middle), and priority (bottom). Significant differences are marked by *.

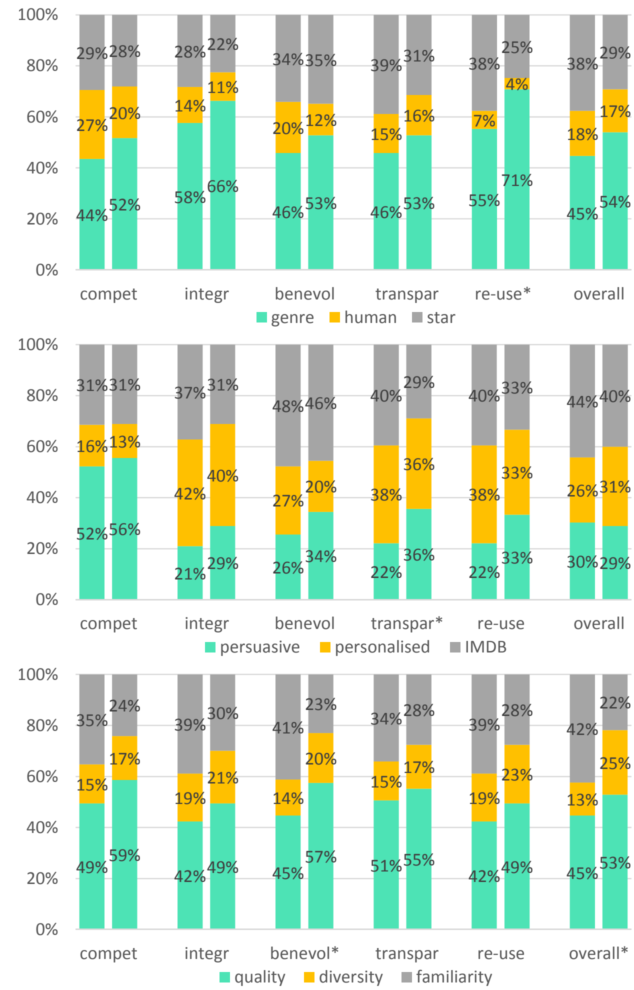

Figure 9. Low- vs high-openness users: presentation (top), explanation (middle), and priority (bottom). Significant differences are marked by *.

their genres. On the other hand, the human presentation receives less trust, presumably due to the more impulsive nature of following such recommendations, which is uncommon for high-conscientiousness people.

Surprising results are obtained in the *explanation* dimension, where personalised explanations increase by 7.4%, while IMDb-based explanations drop by 6.2%. This trend is significant for integrity, transparency, and intention to re-use. We find this result somewhat counterintuitive, as high-conscientiousness people would naturally be expected to trust the reliability of the IMDb explanations, which aggregate thousands of opinions. We hypothesise that personalised explanations, clearly linking the recommendations to past movies liked by the user, may instil more trust than the IMDb explanations.

Conscientiousness groups exhibit a significant difference in the *priority* dimension. We observe a 4.1% increase for the diversity prioritisation, with minor drops in quality- and familiarity-based prioritisations. These changes are significant for the benevolence construct, which may be related to the more balanced coverage of genres offered by the diversity prioritisation. Despite this increase, diversity remains the least trusted factor in this dimension, which is dominated in both groups by the quality prioritisation.

Neuroticism

The comparisons of low- and high-neuroticism users are shown in Figure 8. In the *presentation* dimension, we observe a 7.2% increase in preference towards star ratings, at the detriment of a 8.5% decrease in the genre presentation. This result is significant with respect to the integrity and benevolence constructs. We link this result to the worrying nature of high-neuroticism people and their general inclination to avoid frustration, boosting trust in the more reliable star-rating presentation.

More moderate, although statistically significant changes are observed in the *explanation* dimension. Here, the aggregated preference towards the IMDb-based explanations increases by 3.6%, while persuasive explanations drop by 2.9%. These findings are significant for the integrity, transparency, intention to re-use, and overall trust constructs. We explain these findings again by the more reliable nature of the IMDb explanations, which may reduce the perceived risk of frustration for high-neuroticism people. No significant differences are observed in the *priority* dimension.

Openness

Finally, Figure 9 compares between the low- and high-openness groups. We observe an increase of 7.7% for the genre grouping and drops for both star-ranking (4.4%) and human (3.3%) presentations. However, note that the change is significant for the intention to re-use construct only. The increase for genre may be able to be explained by the preference for variety and lack of focus common to high-openness people. Thus, the range of genres in this presentation can indirectly support their desire to experience diverse things. On the contrary, low-openness people are more comfortable with the more traditional star ranking of the movies.

In the *explanation* dimension, we observe an increase of 7.5% for the persuasive, and drops of 4.2% and 3.2% for the IMDb-based and personalised explanations, respectively. This finding is harder to explain, although the significant difference obtained in transparency hints that information in the persuasive explanations potentially resonates with the curiosity of the high-openness people. On the contrary, IMDb explanations driven by the wisdom-of-the-crowds may seem too restrictive for high-openness people and, as such, their preference towards this explanation drops.

Significant differences between the two groups are also observed in the *priority* dimension. Here, preference towards the familiarity prioritisation decreases by as much as 9.3%, mostly at the detriment of an aggregated 6.8% increase for the quality prioritisation. These changes are found to be significant for the benevolence and overall trust constructs. We attribute this finding to the desire of high-openness people to explore outside of the mainstream embodied by the familiarity prioritisation. That said, the quality of the recommendations is still important, which explains the observed trade-off.

DESIGN IMPLICATIONS AND DISCUSSION

In this work we investigated a number of recommendation presentation factors that can potentially instill user trust. We surveyed prior literature to extract nine factors of trust, which were grouped into three dimensions. We then designed and conducted a crowdsourced user study that experimentally compared the power of these factors. This paper presents a thorough analysis that highlights several dominant factors and explores differences related to users' personality.

Our study brings several operationalisable findings to the foreground. The first refers to the presentation dimension, where *genre-based grouping of the recommended items was the most trusted presentation factor*. In the ranking phase this was preferred by the users with respect to all the constructs of trust, with the ratio of votes towards genre grouping hovering around the 50-60% mark. Although the debrief pulled some votes to other presentations, genre grouping still remained the dominant factor. The users commended its organised structure, which helped them to identify desired items in an easier way. This finding re-affirm the results of [23] and suggests that, beyond movie genres, system designers should consider grouping the recommended items according the available salient domain features, in order to increase the levels user trust.

Another important finding manifests in the priority dimension. In line with earlier results of [22], *quality prioritisation of the recommendation lists was found to be the most trusted* in the ranking phase, outperforming diversity and familiarity. This dominance was observed with respect to all the studied constructs of trust, with quality prioritisation attracting between 40% and 50% of votes. However, in the debrief phase quality remained the most trust factor only in three constructs, while the preference towards diversity and familiarity increased significantly. We explain this by the subtle differences between the recommendation lists, which become apparent only if explicitly explained to the users. The lack of a clear winner in this dimension suggests that different users may prefer different prioritisations of the recommendation lists. Thus, system designers should pay attention to these preferences of the users and consider how to align them with their business goals.

In the explanation dimension, different factors were preferred for different constructs of trust. The personalised explanations were perceived to be most trusted with respect to integrity, transparency, and intention to re-use. On the contrary, IMDb-based explanation were most trusted with respect to benevolence and overall trust, and persuasive explanations were preferred with respect to the competence construct. Interestingly, exactly the same preferred factors were observed both in the ranking and debrief phases of the study. Qualitative user feedback highlights the individual nature of the personalised explanations, which naturally instils high levels of trust. We believe, the findings we present here should play an important role when considering the intended effect of the recommender system, e.g., to provide pure user-centred recommendations recommendation (benevolence is a priority) or to bring users back to the system again (intention to re-use is a priority).

Following this, we delved deeper into differences in trust perception driven by user's personality traits. We considered the traits of the Big Five model and compared between groups of users exhibiting high and low trait scores. Multiple statistically significant differences were observed in all the traits, with as much as 12 and 8 factors of trust (across all three dimensions considered) being significantly different in the extraversion and conscientiousness traits, respectively. Since the differences between the studied prioritisations were subtle, the most pronounced changes between the groups were observed in the presentation and explanation dimensions. This allows us to conclude that the levels of trust instilled by the presentation of the recommendation list and the explanation of items are user-dependent and may vary substantially across different types of users. Thus, system designers should consider boosting trust by *adjusting the presentation and explanation of recommendations to the user's personality traits*, while the prioritisation of the recommendation list requires less attention.

In summary, this work establishes the levels of trust instilled by various presentation, explanation, and prioritisation strategies in movie recommender systems. Being collected from a large pool of participants across multiple countries and supported by qualitative feedback, these results provide solid evidence for recommender systems researchers and practitioners alike, informing the design of future systems. Our study was independent of the underlying recommendation method; thus, it assumed that the recommendation lists already existed and dealt with the trust instilled by the presentation of the list. Although our study considered the domain of movies only, we believe that similar findings may be obtained in other domains, where the items can be characterised by well-defined features. For example, movie genre grouping may be replaced by presentation of cameras according to their manufacturer or IMDb explanation could translate into the number of hotel reviews on TripAdvisor instead.

The conducted analysis, which relies of six distinct construct of trust, uncovered that the notion of user trust in recommender systems is complex and multi-dimensional. As such, system designers may need to steer their choices based on the desired effect of the system. It is reasonable to assume that higher levels of trust indirectly boost the uptake of the generated recommendations. That said, other perceived aspects of the system, e.g., integrity or transparency, may be manipulated by the designers through a careful application of strategies impacting user trust. In a practical recommender system, performance metrics affected by these aspects may be as important as the recommendations themselves. For example, a restaurant recommender may be willing to be seen objective and free of vendor biases, such that it may prioritise the benevolence construct of trust. Likewise, a research paper recommender may wish to be seen knowledgable and, therefore, prioritise the competence construct. Not only does our work show practical ways to boost benevolence and competence of a recommender systems, but it also highlights how these properties of the system are perceived by different types of users.

Several open questions that require further attention arise from our work. The first refers to the dependencies between the trust factors examined. For example, consider a recommender that prioritises items according to their quality, groups them by genre, and also provides personalised explanations to users. How would the combined trust in such a recommender compare to the trust levels instilled by its individual factors? The second is about the intricate relationship between trust and recommendation uptake. We assumed that these are directly related, but this correlation may vary across recommendation tasks and application domains. Hence, a deeper look into this assumption would be beneficial. The third refers to the generalisation of our findings. As the results were obtained in the domain of movies, further studies will be required to establish their validity in other application domains. Finally, while we endeavoured to synthesise as wide as possible range of prior works, the list of factors investigated in this work may not be exhaustive. In the future, we plan to design, implement, and evaluate novel factors, or even dimensions, aiming to instil trust in recommender system users.

REFERENCES

1. I. Benbasat and W. Wang. Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, 6(3), 2005.
2. Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.
3. L. M. Collins, J. J. Dziak, and R. Li. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychological methods*, 14(3):202, 2009.
4. P. T. Costa and R. R. McCrae. Four ways five factors are basic. *Personality and individual differences*, 13(6):653–665, 1992.
5. H. S. M. Cramer, V. Evers, S. Ramlal, M. van Someren, L. Rutledge, N. Stash, L. Aroyo, and B. J. Wielinga. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5):455–496, 2008.
6. A. Dillon and C. Watson. User analysis in HCI - the historical lessons from individual differences research. *International Journal of Human-Computer Studies*, 45(6):619–637, 1996.
7. A. Felfernig and B. Gula. 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*, page 37, 2006.
8. A. N. Finnerty, B. Lepri, and F. Pianesi. Acquisition of personality. In *Emotions and Personality in Personalized Services*, pages 81–99. 2016.
9. S. D. Gosling, P. J. Rentfrow, and W. B. Swann. A very brief measure of the big-five personality

domains. *Journal of Research in personality*, 37(6):504–528, 2003.

10. A. Gunawardana and G. Shani. Evaluating recommender systems. In *Recommender Systems Handbook*, pages 265–308. 2015.

11. K. A. Hoff and M. Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3):407–434, 2015.

12. D. Holliday, S. M. Wilson, and S. Stumpf. User trust in intelligent systems: A journey over time. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*, pages 164–168, 2016.

13. A. Jameson, M. C. Willemsen, A. Felfernig, M. de Gemmis, P. Lops, G. Semeraro, and L. Chen. Human decision making and recommender systems. In *Recommender Systems Handbook*, pages 611–648. 2015.

14. J. D. Johnson, J. Sanchez, A. D. Fisk, and W. A. Rogers. Type of automation failure: The effects on trust and reliance in automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 2163–2167, 2004.

15. S. Y. Komiak and I. Benbasat. The effects of personalization and familiarity on trust and adoption of recommendation agents. *Management Information Systems Quarterly*, pages 941–960, 2006.

16. J. D. Lee and N. Moray. Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40(1):153–184, 1994.

17. J. D. Lee and K. A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004.

18. S. Matz, Y. W. F. Chan, and M. Kosinski. Models of personality. In *Emotions and Personality in Personalized Services*, pages 35–54. 2016.

19. N. Moray, T. Inagaki, and M. Itoh. Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, 6(1):44, 2000.

20. S. Nowak and S. M. Rüger. 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*, pages 557–566, 2010.

21. J. O'Donovan and B. Smyth. Trust in recommender systems. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*, pages 167–174, 2005.

22. U. Panniello, M. Gorgoglione, and A. Tuzhilin. 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):182–196, 2016.

23. P. Pu and L. Chen. Trust building with explanation interfaces. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*, pages 93–100, 2006.

24. P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the ACM Conference on Recommender Systems, RecSys*, pages 157–164, 2011.

25. W. B. Rouse. Adaptive aiding for human/computer control. *Human Factors*, 30(4):431–443, 1988.

26. A. Said, S. Berkovsky, E. W. D. Luca, and J. Hermanns. Challenge on context-aware movie recommendation: Camra2011. In *Proceedings of the ACM Conference on Recommender Systems, RecSys*, pages 385–386, 2011.

27. J. Sauer, A. Chavaillaz, and D. Wastell. Experience of automation failures in training: effects on trust, automation bias, complacency and performance. *Ergonomics*, pages 1–14, 2015.

28. K. E. Schaefer and D. R. Scribner. Individual differences, trust, and vehicle autonomy a pilot study. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 786–790, 2015.

29. J. Schrammel, C. Köffel, and M. Tscheligi. Personality traits, usage patterns and information disclosure in online communities. In *Proceedings of the British Computer Society Conference on Human-Computer Interaction, BCS-HCI*, pages 169–174, 2009.

30. C. L. Scott. Interpersonal trust: A comparison of attitudinal and situational factors. *Human Relations*, 33(11):805–812, 1980.

31. G. Shani, L. Rokach, B. Shapira, S. Hadash, and M. Tangi. Investigating confidence displays for top-N recommendations. *Journal of the Association for Information Science and Technology*, 64(12):2548–2563, 2013.

32. R. R. Sinha and K. Swearingen. Comparing recommendations made by online systems and friends. In *Proceedings of the Workshop on Personalisation and Recommender Systems in Digital Libraries*, 2001.

33. K. Swearingen and R. Sinha. Interaction design for recommender systems. In *Designing Interactive Systems*, volume 6, pages 312–334, 2002.

34. N. Tintarev and J. Masthoff. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*, pages 353–382. 2015.

35. M. Tkalcic and L. Chen. Personality and recommender systems. In *Recommender Systems Handbook*, pages 715–739. 2015.

36. P. Victor, M. D. Cock, and C. Cornelis. Trust and recommendations. In *Recommender Systems Handbook*, pages 645–675. 2011.

37. A. Vinciarelli and G. Mohammadi. A survey of personality computing. *IEEE Transactions on Affective Computing*, 5(3):273–291, 2014.

38. Y. D. Wang and H. H. Emurian. An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior*, 21(1):105–125, 2005.

39. W. Wu, L. Chen, and L. He. Using personality to adjust diversity in recommender systems. In *Proceedings of the Conference on Hypertext and Social Media, HT*, pages 225–229, 2013.

40. S. Xiao and I. Benbasat. The formation of trust and distrust in recommendation agents in repeated interactions: a process-tracing analysis. In *Proceedings of the International Conference on Electronic Commerce, ICEC*, pages 287–293, 2003.

41. K. H. Yoo, U. Gretzel, and M. Zanker. Source factors in recommender system credibility evaluation. In *Recommender Systems Handbook*, pages 689–714. 2015.

42. K. Yu, S. Berkovsky, D. Conway, R. Taib, J. Zhou, and F. Chen. Trust and reliance based on system accuracy. In *Proceedings of the Conference on User Modeling Adaptation and Personalization, UMAP*, pages 223–227, 2016.

43. K. Yu, S. Berkovsky, R. Taib, D. Conway, J. Zhou, and F. Chen. User trust dynamics: An investigation driven by differences in system performance. In *Proceedings of the International Conference on Intelligent User Interfaces, IUI*, 2017.