

# Adapting Recommendation Diversity to Openness to Experience: A Study of Human Behaviour

Nava Tintarev, Matt Dennis, and Judith Masthoff

Department of Computing Science\*  
University of Aberdeen  
{n.tintarev,m.dennis,j.masthoff}@abdn.ac.uk

**Abstract.** This paper uses a *User-as-Wizard* approach to evaluate how people apply diversity to a set of recommendations. In particular, it considers how diversity is applied for a recipient with high or low Openness to Experience, a personality trait from the Five Factor Model. While there was no effect of the personality trait on the degree of diversity applied, there seems to be a trend in the way in which it was applied. Maximal categorical diversity (across genres) was more likely to be applied to those with high Openness to Experience, at the expense of maximal thematic diversity (within genres).

**Keywords:** Diversity, Serendipity, Personality, Recommender Systems

## 1 Introduction

Recommender systems offer items for a user to try or buy based on what is known about their previous preferences or behaviours. Often these recommendations are presented as a list of items. Historically, the emphasis has been on ensuring that items are useful to the user. More recently, the need for diverse recommendations has been recognized by the recommender systems community [1–4].

There is some consensus that personality is a key aspect to consider when injecting novelty and diversity into recommendations, since attitudes towards new or diverse experiences vary considerably amongst users [5]. Such a recommender system would therefore require a model which accurately describes the personality of users. The trait model from Psychology is a fitting candidate for this. It describes a user’s personality by scores on several dimensions, called traits, which are measured via validated self-report questionnaires. Among trait models, the Five Factor Model [6] (FFM, also known as the Big 5 model) has been shown to have excellent reliability in practice [7]. The model describes the personality of an individual by five traits: Extraversion (I), Agreeableness (II), Conscientiousness (III), Emotional Stability (IV), and Openness to Experience

---

\* This research has been funded by the Engineering and Physical Sciences Research Council (EPSRC, UK), grant ref. EP/J012084/1

(V). Of the five traits, Openness to Experience seems an ideal candidate for personalising the level of novelty and diversity in recommendations, as it describes qualities such as active imagination, aesthetic sensitivity, attentiveness to inner feelings, preference for variety, and intellectual curiosity [8]. It seems plausible that users who are less open to experience would prefer to receive recommendations that are similar to previously liked items, while those that are more open to experience would prefer to receive more novel recommendations. This paper will investigate this.

It can be claimed that one of the functions of recommender systems is to help users discover new unexpected items, or to support *serendipitous* discoveries [9]. It is still unclear how to apply diversity in a way that best supports these sort of discoveries, although some exciting attempts have recently been made in the area [10]. We approach this problem by studying how people make recommendations to others, based on the other person’s preferences and openness to experience.

Section 2 outlines previous work in applying personality in recommender systems as well as recent developments in the application of diversity in sets of recommendations. Section 3 describes an experiment in which participants took the role of the system to investigate how they adapt a set of recommendations to a fictitious friend. Section 4 describes our results and Section 5 concludes with the implications of our findings.

## 2 Related Work

In relation to personalization and diversity this paper identifies two specific issues: 1) there may be a difference in preference for the degree of diversity in recommendations among users, and 2) the distinction between categorical (across categories) and thematic (within categories) diversity in recommendations has not received a great deal of weight in previous literature and would benefit from empirical testing with users.

Personality traits are “enduring patterns of perceiving, relating to, and thinking about the environment and oneself that are exhibited in a wide range of social and personal contexts” [11]. Generally it is assumed that: a) traits are relatively stable over time, b) traits differ among individuals (for instance, some people like to try new things while others prefer to stick to known options), and c) traits influence behaviour (e.g. ordering familiar food at a restaurant).

Given that traits influence behaviour, tailoring to traits may help improve the performance of a system. For example, a trait such as Openness to Experience is likely to affect a user’s overall preference for more familiar or more novel items. This is not a new idea; personality-based recommender systems are a growing area of study [7, 12, 13], and some connections between personality and people’s tastes and interests have already been found [7, 13].

The tailoring of recommender systems to personality has also been found to improve accuracy in two cold-start scenarios (sparse data sets and new users [14]), to help predict choices for presidential candidates [7], and to positively impact the acceptance of a system (such as the intent of users to return [15]).

However, to our knowledge, the role of a specific trait on a specific behavior of a system, such as Openness to Experience on diversity, has not yet been investigated.

Although the connection between personality and diversity has not yet been established, there is a consensus that diversity should be considered to improve user satisfaction with sets of recommendations. Diversity in recommendation sets has been suggested as a heuristic [16], and has empirically been shown to increase user satisfaction with a list of recommendations [4], and help users to find target items faster [2]. In recent years, many researchers have focused on generating diverse recommendations; aiming to reach an optimal balance between diversity (between items) and similarity (to the user profile) [1–5, 17]. Recent approaches vary from graph theoretic approaches to reranking, fuzzy logic, and fusion (based on two or more seed items) [5].

In addition, if diversity is applied effectively, exposing users to a large range of items makes them more likely to be positively surprised. *Serendipity* has been defined as the extent to which recommended items are both useful and surprising to a user [9].

So, the question is: how can diversity be applied in a way that may improve serendipity? To start, the degree of serendipity is not merely based on the number of items the user is exposed to, for example as measured by catalogue coverage (defined in [9]). Nor is it the same as having maximally different or poor recommendations – while such recommendations may be surprising, they are less likely to be useful and more likely to damage accuracy metrics. Ziegler et al. (2005) proposed a topic diversification approach based on taxonomy-based dissimilarity [4]. As may be anticipated, this also impacted accuracy negatively. An alternate set of approaches which re-rank a list of top items was found to improve diversity without a great loss in accuracy [17]. A more user-centric clustering approach found that users preferred to be exposed to items in a diversified set of clusters, but with a less diversified set of items inside each cluster [1].

This latter result links tightly with the concept of balance, or the distribution of items within categories [18]. Balance, or *thematic diversity*, can be applied with or without any category coverage requirements, which makes it distinct from *categorical diversity*. The user-centric evaluation model for recommender systems (ResQue), on the other hand also distinguishes between diversity of recommendations and *perceived* categorical diversity [19]. Other work has modelled unexpectedness as the distance from expected items and used a weighted distance for the attributes of the item (e.g. language, genre, director, etc) [20].

This paper investigates how distance from known liked items is considered by humans when making recommendations for others. That is, not only *if* they apply a diversity-increasing measure when they know someone is open to experience, but also *how* they apply it. For example, a book may have the same author, but contain very different themes, or belong to a very different genre. It seems likely that variation on one dimension may affect similarity more than variations in other dimensions – for example: if two books have the same author

and genre, but contain different themes, are they more similar than two books which have the same genre and themes, but different authors?

### 3 Experiment

The experiment was administered as an online questionnaire using Amazon’s Mechanical Turk (MT) service [21], a popular crowdsourcing tool. Building on previous studies, which found that friends of a user consistently provided better recommendations than a recommender system [22], this experiment adopts the *User-as-Wizard* method [23, 24]. The participants took the role of the system in giving three recommendations (books) to a fictitious friend. The purpose is to investigate if they adapt their set of recommendations for this friend, given some information about their personality and preferences. More specifically, this experiment investigates how Openness to Experience may affect the diversity of the recommendations given by such a participant. It also studies how this adaptation, or application of diversity is applied.

#### 3.1 Participants

On Amazon’s Mechanical Turk (MT) participants (called *workers*) can complete small tasks (called *HITs*) made available by *requesters* and are paid a set sum (usually from \$0.01 to \$1 USD) for their time. If a worker fails to complete the HIT properly, their response can be rejected by the requester, and this will be reflected in the worker’s acceptance rate. As most requesters require a high acceptance rate from workers, this can seriously affect a worker’s ability to participate in future HITs. In our experience, this causes workers on MT take their acceptance rate very seriously and endeavour to complete HITs to the best of their ability. Nevertheless, our design was such that it would make any invalid responses easy to spot and remove, as described later in this section.

For our experiment (HIT), participants had to be based in the United States and have an acceptance rate of 90% (to ensure a good quality of responses) and were paid \$0.50 (US). Prior to the main experiment, participants were asked to select their age from a range, their gender and how many books they read in a year (from a range). We then administered a very short Cloze Test [25] for English fluency due to the language based nature of the experiment. If participants did not pass the test, they were excluded from the rest of the experiment by the system, and instructed to return the HIT.

120 participants were used in analysis (after 128 participants, or 51.6%, were automatically excluded from the experiment due to failing the English fluency test) and were balanced between genders (57% female, 41% male, 2% undisclosed). The largest group of participants were 26-40 (46%), followed by the categories in the range of 16-25 (29%) and 41-65 (22.5%). All participants read at least one book a year - over 40% stated that they read over 20 books a year, 26% that they read up to 20, and 33% that they read up to 10 books a year.

The average score for Openness to Experience<sup>1</sup> was 5.52 (StD=1.21) out of 7, and had a range from 2.5 to 7, which is normal for the general population<sup>2</sup>. The average time taken to complete the experiment was around 5 minutes (316 seconds), although we allowed up to 30 minutes.

### 3.2 Materials

This experiment uses short stories that have been previously validated for describing people who are low or high in Openness to Experience, but keep the levels of the other four traits within the normal range for the general population [28]. Originally the stories were adapted from the NEO-IPIP 20-item scales [29] by combining the phrases into sentences to form a short story, with the addition of a very common male name, Oliver. Table 1 shows the stories used.

**Table 1.** Stories used for two levels of openness to experience, high and low.

<b>openness_low</b>	Oliver is not interested in abstract ideas, as he has difficulty understanding them. He does not like art, and dislikes going to art galleries. He avoids philosophical discussions. He tends to vote for conservative political candidates. He does not like poetry and rarely looks for a deeper meaning in things. He believes that too much tax money goes to supporting artists. He is not interested in theoretical discussions. Oliver is quite a nice person, and tends to enjoy talking with people.
<b>openness_high</b>	Oliver believes in the importance of art and has a vivid imagination. He tends to vote for liberal political candidates. He enjoys hearing new ideas and thinking about things. He enjoys wild flights of fantasy, getting excited by new ideas.

### 3.3 Experimental design

A between subject design was used with participants randomly allocated to one of three conditions: baseline (no-story), `openness_high` and `openness_low`. Reasonably, if a person is required to make a recommendation for another they are likely to factor in attributes such as the taste and personality of the other person. Therefore participants were first introduced to Oliver, with the appropriate story (and none for the baseline) from Table 1 that gave the participants an idea of his personality. The interests for Oliver were given in the same way in all three conditions: “*Oliver just read a book called *The Rose* by the author John Smith, which he really liked.*”.

<sup>1</sup> This is the mean score based on two items, of which one is phrased positively and the other negatively [26].

<sup>2</sup> Values for openness to experience are typically high in the general population (m=5.38, StD=1.07) [27]

**Meet Oliver**

*Oliver believes in the importance of art and has a vivid imagination. He tends to vote for liberal political candidates. He enjoys hearing new ideas and thinking about things. He enjoys wild flights of fantasy, getting excited by new ideas.*

Oliver just read a book called *The Rose* by the author John Smith, which he really liked.

**Task**

You want to recommend 3 books for Oliver to read next. There are many books to choose from.

You can pick a new book to recommend to Oliver by the book's author, the genre of the book compared to *The Rose* and the literary themes present in the book compared to *The Rose*.

Examples of literary themes are: global warming, love story, friendship and war

Examples of genres are: crime thriller, romance and science fiction.

---

**Recommendation 1 of 3**

I want to recommend a book with:

What made you make this recommendation?

When you are happy with your choice, click the next button.

**Fig. 1.** A screenshot of the high openness variant of the questionnaire

Participants were instructed to recommend three books for Oliver to read next. For each recommendation, they were asked to specify how they varied the recommended book from *The Rose* (a fictional book created for this experiment), which they already knew Oliver liked. They had to specify the degree of variation on *each* of the three dimensions: author (same or different), genre (same, similar or different) and themes (almost all themes in common, some themes in common, no themes in common). To make sure participants understood these types of dimensions, examples of both literary themes (global warming, love story, friendship and war), and genres were given (e.g. crime thriller, romance and science fiction). We did not specify the actual genre of, or themes present in, *the Rose*. Participants were then asked for a justification for their recommendation. Participants were only permitted to proceed to the next recommendation once they had specified all dimensions and provided justification for these, which among other things, ensured a large degree of qualitative feedback and helped to deter and remove spurious or malignant participants. Figure 1 shows the layout of the experiment and the story for high Openness to Experience.

An alternative design could have been to ask participants to select books from a list, where the books varied according to these properties. However, since we wanted participants to be able to specify three books that all differed (or did not) in the same way – which meant there should be 54 books ( $2 \times 3 \times 3 \times 3$ ) to scroll through – the more direct approach of asking them how they would vary the recommendation was chosen.

At the end of the questionnaire, participants were shown the three recommendations that they had made along with an optional comments box to allow them to comment on the list as a whole or any other aspect of the experiment.

Participants were then asked to complete a short TIPI personality test [26] to establish their own personality according to the Five Factor Model (which includes Openness to Experience), as this could have influenced their decisions.

Our main hypothesis is that the story with high Openness to Experience will lead to more diverse recommendations compared to the story with low Openness to Experience, as well as the baseline.

## 4 Results

### 4.1 Effect of story

To calculate diversity of a particular recommendation to “the Rose”, we took the sum of the diversity on author, genre and themes. We realize that some of these attributes may be more important than others, but for simplicity we assumed they all have the same weight. To calculate the diversity of a particular recommendation to “the Rose” on author, we used 0 for books with the same author, 1 for books with another author. For genre, we used 0 for books with the same genre, 0.3 for books with a similar genre and 1 for books with a different genre. For diversity on themes, we used 0 for books with almost the same themes, 0.3 for books with many themes in common, and 1 for books with no themes in common. We realize that the 0.3 is arbitrary, but felt it was plausible it represents a minority ( $< 0.5$ ) but non-negligible degree of variation ( $> 0$ ).

**Table 2.** How diverse the given recommendations were: the average and maximum diversity across the recommended set.

Condition	DiversityAvg	DiversityMax
baseline	1.42 (0.31)	2.00 (0.42)
openness_low	1.41 (0.38)	2.08 (0.51)
openness_high	1.46 (0.30)	2.14 (0.44)
overall	1.44 (0.33)	2.08 (0.46)

We used two measures of diversity for the set of three recommendations: the average, and the maximum (of the diversity for each book, across the three books). Table 2 shows the diversity scores for the three conditions, and overall. While there is a trend for more diverse recommendations in the condition with the story for high Openness to Experience, this trend is not significant. However, we note that the diversity for all conditions is also very high. The average degree of diversity is around 1.5, which means that at least two dimensions were modified. The maximum diversity is over 2, which means either changing at least two dimensions to something completely different, and possibly changing all three dimensions. While this is not surprising for the condition with high openness to experience, it is more so for the condition with low openness to experience. Participant comments suggest that they would like to expose their friend to more diverse recommendations, even if it is only to decide they do not like certain types of books or to better understand their preferences: *“Oliver should read something completely different to find out whether or not he likes it”*; *“I would*

*do this as a test to narrow Oliver’s reading preferences the way English teacher make me narrow my essay thesis statement!”*

Given the similarity across conditions for the two diversity measures, we will only look at the average diversity (DiversityAvg), in the rest of the paper.

## 4.2 Diversity in a sequence of recommendations

**Table 3.** Change of diversity across the three recommended items

Condition	Book1	Book2	Book3
baseline	1.04 (0.58)	1.56 (0.49)	1.67 (0.67)
openness_low	1.18 (0.68)	1.35 (0.66)	1.72 (0.74)
openness_high	1.33 (0.63)	1.38 (0.61)	1.68 (0.72)
overall	1.18 (0.64)	1.43 (0.60)	1.69 (0.71)

Sets of recommendations differ from single recommendations in that diversity within the list is also an important factor. A less studied factor is how this diversity changes (or should change) across a sequence of recommendations. We investigated what happened with the diversity of the recommendations in the set, and found that it increases. Book2 is more different from the seed item than Book1, and Book3 is more different than Book2. These differences are significant at  $p < 0.01$  (Bonferroni corrected). The values for diversity for the three books across conditions can be seen in Table 3.

It is worth noting that in the condition with the story for low openness, the diversity for Book3 is comparable to the other conditions, even if the diversity starts lower for the first two books. Participant comments give the impression that they are aiming to win the confidence of their friend by starting with familiar recommendations, but then making an effort to broaden their horizons: *Book1: “This is familiar ground for Oliver; he will probably like it just as well as the last book he read by this author. If not, however, he knows that he doesn’t need to read any more of that author’s books.”; Book2: “This allows him to see someone else’s take on the same genre and themes as the book he originally enjoyed. It helps him determine whether or not he would like to read more books of that genre in the future.”; Book3: “This would be something just for variety, something to break him out of his usual pattern and maybe get him interested in something he perhaps wouldn’t think he’d like. He may find something he really enjoys.”.*

These findings bring into dispute previous results, which found that users thought that items recommended by friends mostly served as reminders of previously identified interests compared to those by online recommender systems which were “new” and “unexpected” [22]. However, participants made a clear effort to give familiar recommendations for the first item in particular, as was seen in the user comments: *“I think it’ll be good for Oliver to broaden his hori-*



*zons a little bit, but still be in his comfort zone. (low openness)”; “He liked the author so another one by him would be the next logical step. (high openness)”*

### 4.3 Effect of participant personality

In this experiment participants scored highly on Openness to Experience which is comparable to the norm for the general population. There was no correlation between the aggregated TIPI score on openness and the diversity of the recommendation set.

Also, when we controlled for effect of participant personality there was no significant correlation between the conditions and the diversity of the recommendations given. It seems like there was no significant influence of participant personality on the recommendations given.

### 4.4 What is diversified

We investigated if people applied diversity to the same extent to all three dimensions: author, theme and genre. Table 4 shows the average diversity of the three dimensions. While none of the differences are statistically significant, it is interesting that the trend for the low openness condition is the opposite as for the high openness condition. For low openness, participants varied themes more than genre, whilst for high openness they varied genre more than themes. If we survey the ranges for the conditions, all conditions cover the full range (0-3), except for high openness. For high openness, the range for genre only starts from 0.3 (i.e. no 0 diversity) and the range for themes is 0-2.3 (i.e. nobody applied maximum theme diversity for all three books). We can also see that the average genre diversity is largest for high openness. This is reflected in participant comments where genre diversity is often mentioned: *“Oliver has a vivid imagination; therefore, he would enjoy many different authors and genres of books. ”*; *“This same author happens to have written another book, but it’s in a completely different genre. Nevertheless, Oliver might like it given that it’s the same writing style and has many themes in common with the book he liked.”*. Similarly, the average theme diversity is lowest for the high openness condition. Again, participants’ comments reflect that they are willing to consider recommendations with no common themes: *“Oliver may have enjoyed The Rose because of the author’s writing style and would enjoy other books written by that same author, although with no similar themes in common.”* Finally, there is also a difference in the range of diversity for authors across participants. In the low openness condition, participants occasionally did not introduce any variation in author for any of the three books (range=0-3), while at least one author (across the three books) was different in the other two conditions (range=1-3).

### 4.5 Same, same but different

One of our concerns was that participants should be able to pick several books with the same properties. So for example, if a participant wanted to recommend three very similar books this should be possible. There are instances in

**Table 4.** Comparison of average diversity of the set along three dimensions: author, genre and theme. A comparison of the three levels of story and overall are given.

Condition	Author	Genre	Theme
baseline	1.92 (0.70)	1.22 (0.70)	1.12 (0.69)
openness_low	1.83 (0.71)	1.17 (0.63)	1.25 (0.78)
openness_high	1.88 (0.75)	1.45 (0.58)	1.06 (0.49)
overall	1.88 (0.72)	1.28 (0.64)	1.14 (0.66)

the data where two or even three books are selected with both a lot of variety (different, different, none) and very little variety (same, similar, many).

#### 4.6 Discussion

Our results are in line with previous work which distinguishes between diversity across categories (implemented as genre in our experiment) and balance within each category (or thematic diversity in our experiment) [18]. A previous study found that a majority of users preferred recommendations based on an approach which exposed them to items in a diversified set of clusters, which could be analogous to categorical diversity, rather than within clusters (thematic diversity) [1]. The preference by a majority for categorical rather than thematic diversity in that study may reflect the fact that the general population normally scores relatively highly on Openness to Experience.

#### 4.7 Implications for recommender system design

One potential implication is that if it is possible to detect the Openness to Experience of users the recommender system can apply more thematic variation for these users. An open question is whether this can be detected from the types of items users have already rated – those who are less open to experience are likely to have explored fewer different categories. For users that are new to the system, it may still be possible to ascertain if they are early adopters or use many different applications using a unified login (such as OpenID). An alternative would be to conduct a short personality test for all users, or to apply more categorical diversity by default (given the majority of people are more open to experience) and increase the degree of thematic diversity when users have low Openness to Experience (or if they rate many consecutive recommendations as poor). Despite the overlap with previous findings, these are speculative implications, since one limitation of our study is that we surveyed the behaviour of *Users-as-Wizards* rather than the behaviour of the recipient of such recommendations. Participants seemed to use a different strategy when recommending for a friend who has high compared to low Openness to Experience. While it is not certain if this is also an optimal way to adapt, it gives support to the theory that this is an adaption that is worthwhile to test with recipients as well. Secondly, distinguishing these

two types of diversity (categorical and thematic) may have different roles on increasing serendipity for different types of users. We have not studied the effect of diversified lists on recipients, however creating a measure of diversity as is used by humans and that overlaps with current algorithms may be a first step towards tailoring for serendipity. Finally, for a set of recommendations, it seems like a good strategy to start with a familiar item and increasingly diversify as the set becomes larger. We caution however that further studies in other domains may be required to establish a more general theory for the application of diversity according to user personality. In future work we plan to investigate items that vary along the dimensions of degree of investment (e.g. movie vs holiday) and objectivity (e.g. movie vs camera) [30, 31].

## 5 Conclusion

This paper investigated the role of Openness to Experience on a set of recommended items. Participants acted as the system and gave recommendations for a fictitious friend for whom they knew an item they liked and were given a story about their personality. While we did not find a significant effect of story on the degree of diversity applied, there seemed to be a difference in how it was applied. Categorical (or genre) diversity was likely to be applied to greater extent to a friend who was open to experience, at the expense of thematic (or inter-genre) diversity. This is in line with recent findings in recommender systems which have found that users are more likely to prefer recommendations from different clusters rather than varied within a cluster. Despite the fact that users of recommender systems may be particularly high on the Openness to Experience personality trait, there may be room for tailoring. Participants who are low on the trait may benefit from more inter-theme (thematic) rather than inter-genre (categorical) variation. We also found that participants increased the diversity of their recommendations within the set of three items. While the sequence started with the most familiar items, in the low openness condition, the final item was highly diverse from the seed item (more so than in the high openness condition). This suggests that it would be good practice to start with familiar items to win the user's confidence, but to quickly diversify thereafter.

## References

1. Abbassi, Z., Mirrokni, V.S., Thakur, M.: Diversity maximization under matroid constraints. Technical report, Department of Computer Science, Columbia University (2012)
2. Bridge, D., Kelly, J.P.: Ways of computing diverse collaborative recommendations. In: Adaptive Hypermedia and Adaptive Web-based Systems. (2006) 41–50
3. Smyth, B., McClave, P.: Similarity vs. diversity. In: 4th International Conference on Case-Based Reasoning. (2001)
4. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: WWW'05. (2005) 22–32

5. Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011). (2011)
6. Goldberg, L.: The structure of phenotypic personality traits. *American Psychologist* **48** (1993) 26–34
7. Nunes, M.A.S.N.: Recommender Systems based on Personality Traits. PhD thesis, Universite Montpellier 2 (2008)
8. Costa, P.T., McCrae, R.R.: NEO personality Inventory professional manual, Odessa, FL: Psychological Assessment Resources. (1992)
9. Herlocker, J.L., Konstan, J.A., Terveen, L., Riedl, J.T.: Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **22**(1) (2004) 5–53
10. Said, A., Fields, B., Jain, B.J., Albayrak, S.: User-centric evaluation of a k-furthest neighbor collaborative filtering recommender algorithm. In: CSCW. (2013)
11. APA: Diagnostic and Statistical Manual of Mental Disorders. 4th edn. American Psychiatric Association (2000)
12. Dunn, G., Wiersema, J., Ham, J., Aroyo, L.: Evaluating interface variants on personality acquisition for recommender systems. In: UMAP. (2009)
13. Lin, C.H., Mcleod, D.: Exploiting and learning human temperaments for customized information recommendation. In: IMSA. (2002)
14. Hu, R., Pu, P.: Enhancing collaborative filtering systems with personality information. In: Recsys. (2011)
15. Hu, R., Pu, P.: Acceptance issues of personality-based recommender systems. In: Recsys. (2009)
16. Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the users perspective: survey of the state of the art. *UMUAI* **22** (2012) 317–355
17. Adomavicius, G., Kwon, Y.: Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering* **24** (2011) 896–911
18. Golbeck, J., Hansen, D.L.: A framework for recommending collections. In: Workshop on Novelty and Diversity in Recommender Systems in conjunction with Recsys. (2011)
19. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Recsys. (2011)
20. Adamopoulos, P., Tuzhilin, A.: On unexpectedness in recommender systems: Or how to except the unexpected. In: Workshop on Novelty and Diversity in Recommender Systems in conjunction with Recsys. (2011)
21. MT: Amazon mechanical turk. <http://www.mturk.com>
22. Sinha, R., Swearingen, K.: Comparing recommendations made by online systems and friends. In: Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries. (2001)
23. Masthoff, J.: The user as wizard: A method for early involvement in the design and evaluation of adaptive systems. *Fifth Workshop on User-Centred Design and Evaluation of Adaptive Systems* **1** (2006) 460–469
24. Paramythis, A., Weibelzahl, S., Masthoff, J.: Layered evaluation of interactive adaptive systems: framework and formative methods. *UMUAI* **20** (2010) 383–453
25. Taylor, W.L.: Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, **30** (1953) 415433
26. Gosling, S.D., Rentfrow, P.J., Jr., W.B.S.: A very brief measure of the big five personality domains. *journal of research in personality.* **37** (2003) 504–528
27. Goz-lab: Tipi normal values. <http://tiny.cc/9otwqw>
28. Dennis, M., Masthoff, J., Mellish, C.: The quest for validated personality trait stories. In: IUI. (2012)

29. Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., Gough, H.G.: The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality* **40**(1) (2006) 84 – 96
30. Tintarev, N., Masthoff, J.: Over- and underestimation in different product domains. In: *Workshop on Recommender Systems in conjunction with the European Conference on Artificial Intelligence*. (2008) 14–19
31. Tintarev, N., Masthoff, J.: Designing and evaluating explanations for recommender systems. In Kantor, P.B., Ricci, F., Rokach, L., Shapira, B., eds.: *Recommender Systems Handbook*, Springer (2010)