



universität  
wien

## **Bachelorthesis**

**“Improving Trust in Recommender Systems  
through Context Clues”**

verfasst von Dipl.-Ing. Tobias Sippl 00820552

2023S 200061-1 Bachelorarbeit

SS 2023

Studienrichtung (lt. Studienblatt): Bachelorstudium Psychologie

LV-Leitung: Dr. Irina Nalis-Neuner

Wien, 30. Juni 2023

Tobias Sippl

00820552

## **Abstract**

Providing serendipitous recommendations in a recommender system can lead to various problems. Not only is it hard to find an item which is both novel and relevant to the user, but the user should also engage with the item instead of discarding it as irrelevant. Serendipitous items tend to appear as unusual or un-fitting at first, but turn out to be a pleasant surprise once investigated. So it is not only necessary to find a serendipitous item, but also to present it in a way which increases the likelihood that a user engages with it. An experiment was conducted in which participants interacted with a novel book recommender system and were assigned to either an experimental or a base group. The experimental group interacted with a visually enhanced version which was designed to trigger various heuristics aiming to increase user engagement and trust, while the base group used a comparatively bland interface. Trust and serendipity were then measured for each item and the recommender system itself. It was found that while it did not increase trust in the recommender system, serendipity was rated higher for books the participants did not know.

### **Keywords:**

Recommender Systems, Trust, Serendipity, Heuristics, Psychology, Antecedents

# 1 Introduction

Today it is virtually impossible to interact with the Internet, computers or smart-phones without the use of Recommender systems (RS). While RSs are often understood as search engines like Google or Bing, they also drive platforms for media recommendation like Netflix, Spotify or similar. Recommender systems also exist in the form of health recommender systems, where they can assist in smoking cessation or provide other health advice. RSs can be more noticeable in the form of autocomplete on mobile phones, or work unnoticed in the background in the form of a music platform automatically choosing the next song.

Research on recommender systems is very active, with a strong focus on improving the quality of provided recommendations. There exists a large amount of techniques to deal with this problem, with the most recent addition of large language models, known to the public by names like ChatGPT or Bard. A basic approach to evaluate results is by user satisfaction, user retention and accuracy. While it makes sense to assess results by these metrics, they ignore societal factors and human needs. A more modern approach to evaluate recommender systems is to look beyond accuracy and regard a broader set of criteria. The Digital Humanism initiative, for example, provides a set of values to consider, some of which are serendipity, diversity and fairness.

It is of interest, how the presentation of items influences their perception by the user. While an algorithm can pick a very serendipitous item, the user might ignore it, if it is not presented in the right way. A user might be interested in a more diverse set of recommendations if they are given some context or extra information instead of sticking to what they already know. A study was therefore conducted to test if the perception of serendipity and trust can be improved in the context of a book recommender system.

## 1.1 Recommender Systems

Recommender systems are a type of information filtering system providing suggestions of relevance to a specific user (Ricci et al., 2022a). They are employed in a wide variety of software-systems and take on many forms. Some of the most prominent examples are the Google search engine, the Netflix recommendation algorithm or autocorrection in office or chat applications (Ricci et al., 2022b).

In many cases, their main task is to reduce a large amount of items to smaller set, thereby solving or reducing information overload (Nikolakopoulos et al., 2021). For example, a user might search for a popular genre of music on a

audio-streaming platform. Instead of simply returning possibly thousands of results, a recommender system first filters the items according to a set of rules established by its developers.

These rules can be manifold and depend on the goals of the RS (Ricci et al., 2022b). In the case of video streaming platforms, RSs may estimate how much a user might like certain items and rank them accordingly, while also excluding previously watched items. In contrast, a music streaming service may place already familiar items within the results, as users typically also want to listen to songs they already know (Jannach et al., 2022).

To describe the quality of the results a recommender system delivers, one of the most common terms used is accuracy. A result is accurate, if the RS returned what the user intended to find. While this term is useful in the case of a search engine (e.g. Google, Bing), it is less applicable to a news recommender system. For a user to be satisfied with the recommendations of a news RS, items need to be recent, of appropriate magnitude and account for changing user preferences among other characteristics (Raza & Ding, 2022).

Research now considers other metrics under the term or framework of "beyond accuracy". This umbrella term covers many principles, some of which are human values (e.g. diversity, fairness), while others are oriented towards the users attitude about the recommender system, like trust or privacy (Liao et al., 2022; Stray et al., 2022).

## 1.2 Serendipity

The term serendipity generally refers to some form of an unexpected, positive surprise. While the term is used slightly different by each field of research, a paper on serendipity in recommender systems by Ziarani and Ravanmehr (2021) found that in literature, the term is commonly defined by the components usefulness, novelty, relevance and unexpectedness. Recommendations that fulfill all of these criteria are rare and difficult to generate.

Another important aspect in regards to serendipitous recommendations is the role of emotion. While a serendipitous recommendation might fit the users profile perfectly, their emotional state might simply not allow to receive it as such (Nguyen et al., 2018). Furthermore, some users personalities are more or less accepting of serendipitous recommendations.

The methods to generate serendipitous recommendations generally follow the same strategies also employed in other recommender systems, although

modified or configured to increase the generation of results which are novel, unexpected and relevant. Three commonly used models are collaborative filtering, content-based filtering and context-awareness filtering (Ziarani & Ravanmehr, 2021). Collaborative filtering first matches the inquiring user to users with similar preferences, and then recommends items which those similar users also liked. Content-based filtering recommends items which have similar properties to other items the inquiring user rated positively. Context-awareness approaches analyze the user context, which may be the location of the user or the time the recommendation is generated at. Models which employ more than one approach are called hybrid-models and combine multiple strategies (Ziarani & Ravanmehr, 2021).

The evaluation methods for the serendipitousness of an item can be grouped into three sets. Direct evaluation by a formula, user feedback or evaluation of serendipity components (novelty, relevance, usefulness, unexpectedness). According to Maccatrozzo et al. (2017) and Pina E Cunha et al. (2015) evaluation based on user feedback is the most accurate.

### 1.3 Trust and Recommender Systems

The interaction between a user and a recommender system is also defined by the amount of trust the user holds against it. The more a user trusts the recommender system to generate useful items, the more the user will accept the items (Harman et al., 2014). This is especially important when recommending serendipitous items, as these may appear unexpected which can lead to trust-issues (Afridi, 2019).

There are many ways in which trust between users and recommender systems can be established (Liao et al., 2022). Intuitively, providing a user with relevant recommendations will establish trust over time, while providing unsatisfying recommendations will erode trust. There are also other challenges, as according to Ricci et al. (2022b, p.7) "some users do not trust recommender systems, thus they play with them to see how good they are at making recommendations" and notes that "a certain system may also offer specific functions to let the users test its behavior in addition to those just required for obtaining recommendations".

In addition to these examples, where trust is built by providing satisfying results, heuristics are typically used to evaluate situations (or in this case recommender systems). As Gilovich et al. (2002, p.XV) writes: "The core idea of the heuristics (...) is that judgment under uncertainty is often based on a limited num-

ber of simplifying heuristics rather than more formal and extensive algorithmic processing”.

For example, Liao et al. (2022) tested the effect of the bandwagon heuristic on a recommender system for movie trailers. This heuristic describes the effect by which individuals judge something based on how they believe others think of it (Sundar, 2007). In their study, Liao et al. (2022) presented users with one-sentence explanation as to why the item was recommended. When using an explanation which triggered the bandwagon heuristic, users evaluated the recommendation better compared to explanations which did not trigger the heuristic.

## 1.4 Antecedents

So far, we have looked at how users evaluate a recommender system by using heuristics (snap-judgements) and build trust over a longer timeframe by consistently receiving satisfying recommendations. How individual recommendations are evaluated depends to a large degree on the type of recommender system (movies, music, news, web search-engine, etc.), how the user perceives the recommender system itself (“is it trustworthy?”) and how the user thinks the recommendation was reached (“why was this recommended to me?”).

A model to explain how we evaluate a recommender system is through antecedents. A certain set of antecedent is relevant to a certain type of recommender system. For example, perceiving an online store as having a huge selection of items, positively influences our evaluation of the store in most cases (Jarvenpaa et al., 2000). In the context of a news recommender system, Shin (2020) found, transparency and accuracy are paramount. If users evaluate the relevant antecedents highly, they will also rate the recommender systems accordingly.

In that aspect, the serendipity or trust can be viewed as antecedents to evaluating a recommender system. And trust or serendipity themselves have again antecedents to explain how we rate these elements of a recommender system. Trust for example, can be influenced by conveying size and a positive reputation, which reduce risk perception, leading to a higher willingness to buy an item (Jarvenpaa et al., 2000).

To summarize, the type of recommender system defines which antecedents are of importance for a positive evaluation, while users evaluate the antecedents by applying heuristics and their long-term experiences with the RS. If serendipitous items should be suggested, users might have trust issues, as those items

may be perceived as unusual and therefore irrelevant. By increasing the trust into a recommender system through the use of various heuristics, this problem might be circumvented.

## 1.5 Hypothesis

The goal is to present serendipitous recommendations without creating trust issues or seeming irrelevant. This could be facilitated by adding context information and to build trust through triggering heuristics which influence the relevant antecedents. The hypothesis are:

**H1:** Trust in a recommender system can be increased by providing context clues triggering various heuristics connected to trust.

**H2:** Providing context and a visually interesting interface increases the likelihood for accepting serendipitous items as relevant.

**H3:** H1 and H2 combined, lead to a better evaluation of the recommender system.

If the hypothesis hold, this would present an opportunity to improve recommender systems which provide serendipitous items. By changing the way how items are typically represented, users would be able to appreciate a more diverse set of items.

## 2 Method

To test the hypothesis a short experiment was conducted in which participants interacted with a book recommender system. Participants were presented with one of two versions of the recommender system. The first version only presented the books and some information about the author, while the other added additional context, like user ratings and a reason as to why the item was recommended. The context information itself did not hold any relevant information about the books themselves.

### 2.1 Participants

Participants were able to take part in the study from March to April of 2023. The participants were students taking part in a class on recommender systems at the TU Wien. As a reward for taking part in the study, they received some extra points towards their final grade. In total, 110 people participated of which 9 were excluded in the results due to very short answering times. The mean age of the participants was 26 years.

### 2.2 Procedure

The study was accessible through a webpage and could be taken on any device with a modern web-browser. Participants were first presented with a survey to record their demographic data and their reading habits. Participants were then randomly assigned into either the experimental or the base group. Both groups were presented with the same list of books in the same order. The books were selected from the top rated books of the site Goodreads in the categories of technology, history, science and fantasy. Every participant rated 8 books, answering for each how well it fit their taste, how unexpected the result was and if they would be interested in buying and reading the book if time and money were not an issue. Finally an exit questionnaire captured their thoughts on the recommender system itself.

After accessing the webpage, where the study was hosted, participants were first presented with a short clarification on how the data would be used and other privacy information. After accepting, they were able to continue to the demographic survey. It was only required to enter age and gender, as per the recruitment scheme, the occupation of the participants was known. The RS should give the impression, that a book recommender system was being tested and



therefore the participants reading habits were inquired. The first part was a self-evaluation, containing items like “By comparison, I read a lot.” or “I think reading news articles, online news or similar is important”. Each of the 7 items in the self-evaluation part could be rated between 1 (not at all/fully disagree) or 100 (very much/fully agree). The items were taken from various reading habit surveys (“Forschungsdatenzentrum Bildung”, 2023) and reduced. Next up, the participants were asked what devices they read on or if they listen to books. The 6 options were “Audiobooks”, “Computer”, “E-Reader”, “Mobile Phone”, “Print” and “Tablet”. There were 2 additional fields in case users were using other devices. Next, participants were asked which genres they like to read. We gathered the top 21 genres listed in various webshops and additionally added 4 boxes to add additional genres in case we missed them. Each option was presented as a checkbox. Finally, each participant was given the possibility to list 4 books they have read recently (if any) and how many minutes they read per day on average, split into reading related to their profession, recreational reading and reading news.

**Table 1**

*The books as shown in the recommender*

#	Title	First listed Author
1	1984	George Orwell
2	Clean Code: A Handbook of Agile Software Craftsmanship	Robert C. Martin
3	Sapiens: A Brief History of Humankind	Yuval Noah Harari
4	Circe	Madeline Miller
5	A Game of Thrones (A Song of Ice and Fire, #1)	George R.R. Martin
6	The Art of Statistics: How to Learn from Data	David Spiegelhalter
7	The Midwife of Auschwitz	Anna Stuart
8	Guns, Germs, and Steel: The Fates of Human Societies	Jared Diamond

## 2.3 Recommender

The recommender always presented the same 8 books in the same order. If the participants were in the base group, they were presented with a page containing the book cover, title, author(s), a book description and some information about the primary author. In addition, the experimental group also saw a logo of a respected austrian newspaper (which also operates an online-bookstore), user ratings, a reason for why the recommendation was generated, availability and a price. For a comparison see Figure 1. Both groups then had to rate each recommendation, answering 5 questions. The questions tried to ascertain if the

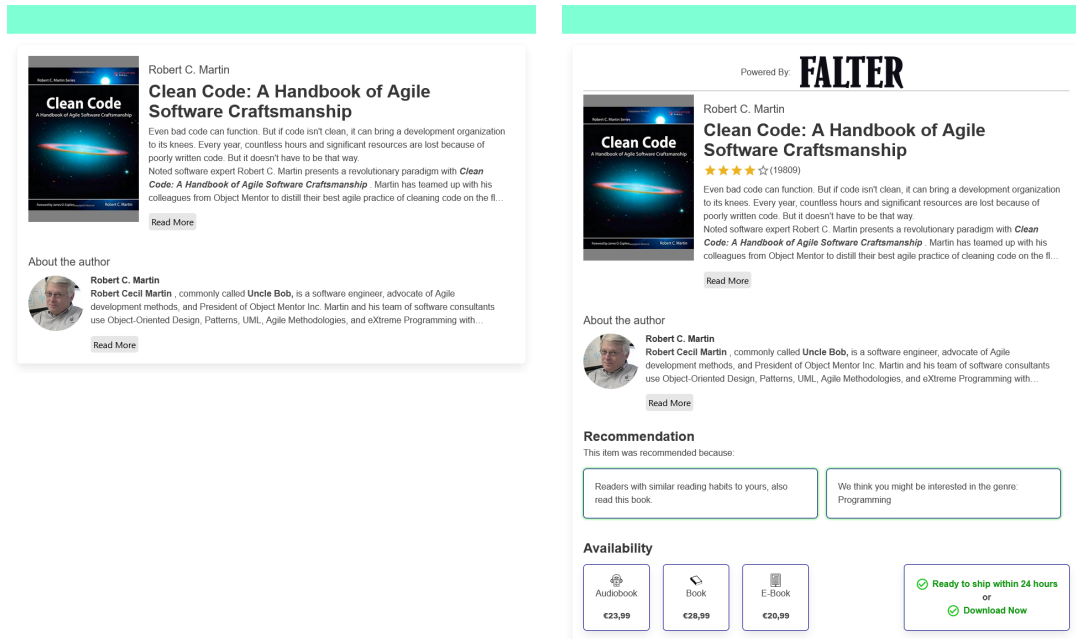
book matched the participants interest, how well participants knew the book (familiarity), if it was different from what they would typically read (novelty), if it was a pleasant surprise (serendipity) and if it was a satisfying result (satisfaction). Each of the questions could be answered with a slider ranging from 1 to 100. Finally they were asked if they would buy the book if they had ample time to read and money were not an issue. This was a simple yes or no item.

After all recommendations had been rated, participants were presented with an exit survey containing 17 questions, each of which could be answered by a slider ranging from 1 to 100. The items were chosen to capture the perceived serendipity and trust in regards to the recommendations. In addition 4 questions were asked to establish if the participants knew of the austrian newspaper of which the logo was featured in the experimental group. The questions to record serendipity were taken from Knijnenburg et al. (2012), those regarding trust from Madsen and Gregor (2000). Items that did not fit the recommender were removed and wording was generally slightly changed to better fit the context of book recommendations.

While participants rated the various recommendations and answered the questionnaires, the time spent on each page was recorded in order to expunge data sets with unreasonably quick responses.

Figure 1

*The base and experimental version of the recommender*



*Note:* This image shows a comparison of the two versions of the recommender that were shown to the groups. On the left the base group version can be seen. Only the book cover, author, book description, author image and author description are visible. The right image shows the experimental group version which also contains an explanation, price, availability and a logo of a prominent austrian newspaper.

### 3 Results

The median time for taking the study was 16 minutes and 9 datasets were rejected due to a completion time of less than 8 minutes (below lower quartile - 1.5 times the IQR). On average, participants required 43 seconds to rate a book, around 5 minutes for the entry questionnaire and 2.5 minutes for the exit questionnaire. See Table 2 for a full listing.

**Table 2***Time taken to answer the various sections of the recommender*

Section	Base (N=51)		Exp. (N=48)		p	Cohens d
	Mean	SD	Mean	SD		
Entry Questionnaire	415	537	388	247	0.625	-0.346
1984	89	54	110	67	0.044	-0.135
Clean Code: A Handbook of Agile Software Craftsmanship	88	214	142	540	0.252	-0.299
Sapiens: A Brief History of Humankind	45	30	54	33	0.070	-0.121
Circe	58	63	65	53	0.275	-0.262
Game of Thrones	38	23	57	102	0.097	0.056
The Art of Statistics: How to Learn from Data	46	36	45	22	0.610	-0.274
The Midwife of Auschwitz	44	33	54	41	0.088	-0.269
Guns Germs and Steel: The Fates of Human Societies	46	31	96	266	0.092	0.007
Exit Questionnaire	180	105	180	69	0.514	-0.207
Total	1048	605	1191	767	0.153	-0.346

*Note:* Values are mean scores, values in parentheses are the standard deviations.

In order to answer **H1**, if higher trust in a RS can be achieved by providing context clues to trigger heuristics, answers were collected by providing an exit questionnaire. The detailed results can be found in Table 3. In total, the experimental group scored slightly lower (Mean: 59.6, SD: 15.0) than the base group (Mean: 60.3, SD: 19.8) although with a smaller standard deviation. Therefore **H1** does not hold. The only subscale on the exit questionnaire, which was almost significant was P2 (liking the recommender system,  $p$ : 0.082), although with a small effect strength ( $d$ : 0.120).

**Table 3***Answers to the exit questionnaire measuring trust*

Subscale	Question	Base (N=51)		Exp. (N=48)		p	Cohens d
		Mean	SD	Mean	SD		
R1	There was enough information provided to decide if I wanted to purchase a book.	73.1	27.1	72.9	21.8	0.521	-0.026
R2	The recommendations were accurate to my reading preferences.	54.5	26.9	55.1	19.6	0.449	-0.180
T1	The system used appropriate methods to provide recommendations.	60.6	24.2	65.0	24.4	0.187	0.204
T2	The system has sound knowledge about how to provide recommendations.	60.6	24.6	55.6	23.8	0.844	0.004
T3	The provided recommendations were on a level with those a competent person could provide.	62.4	27.7	62.3	26.5	0.507	0.131
T4	The system correctly used the information I entered.	67.4	26.8	64.2	22.0	0.742	0.061
T5	The system made use of all the knowledge and information available to reach the recommendations.	61.8	24.4	60.4	23.2	0.618	0.197
U1	I know what will happen the next time I use the system because I understand how it behaves.	53.7	27.8	48.6	23.8	0.835	-0.045
U2	I understand how the system will assist me with decisions I have to make.	58.9	28.1	60.0	23.2	0.412	0.115
U3	It is easy to follow what the system does.	57.4	30.3	54.2	24.0	0.716	0.185
U4	I recognize what I should do to get the advice I need from the system the next time I use it.	60.3	31.1	55.0	25.6	0.820	-0.135
F1	When a recommendation was unusual I am confident that it would still be an interesting read.	49.2	30.6	53.0	25.9	0.252	0.171
P1	I felt that the recommender was suitable for my reading preferences.	59.2	24.6	55.1	22.6	0.802	-0.283
P2	I liked using the recommender.	66.1	27.4	72.9	20.2	0.082	0.120
S1	Among the recommended items there were some pleasant surprises.	81.4	20.5	79.0	19.5	0.724	0.140
S2	There were some unexpected items among the recommendations.	78.1	22.3	74.8	24.7	0.756	0.364
Trust		60.3	19.8	59.6	15.0	0.585	0.201

*Note.* Subscales of the Human Computer Trust questionnaire (R, T, U, F, P). R: Reliability, T: Technical Competence, U: Understandability, F: Faith, P: Personal Attachment Serendipity scale: S. A p smaller than 0.05 would be significant and show that the experimental group rated a scale significantly higher than the base group.

**H2** proposed that providing context and a visually more interesting interface for a recommendation should enhance the serendipity perception of an item. The combined scores by group and question can be found in Table 4. Overall the experimental group rated books lower on most scales. Answers to the questions asked for each book can be found in Table 5. As none of the results were significant, **H2** does not hold. Overall, the experimental group rated the books as slightly less serendipitous.

According to **H3** higher trust and more context clues should lead to a better evaluation of the recommender system. As can be seen in Table 3, the experimental group rated the recommender better than the base group (mean: 72.9 vs 66.1,  $p$ : 0.082), although the total trust score was lower. While there is a difference in mean score, the results fail to reach significance.

**Table 4**

*Answers to the five questions asked for each book*

Question	Base (N=51)		Exp. (N=48)		p	Cohens d
	Mean	SD	Mean	SD		
The book recommended to me matches my interests.	55.7	19.1	50.6	16.6	0.919	0.283
I know this book well. (Never heard of it = 1. Read it recently = 100)	25.3	14.1	22.8	12.4	0.827	0.190
The book is different from books I typically buy/read.	55.0	19.3	55.1	15.0	0.487	-0.007
The book recommended to me is a pleasant surprise.	49.5	17.7	46.5	16.1	0.815	0.181
I am satisfied with the recommendation.	56.6	17.8	53.3	15.7	0.836	0.198

**Table 5***Answers to the five questions asked for each book*

Book	Matches Interest			Familiarity			Novelty			Serendipity			Satisfaction			
	Base	Exp.	SD	Base	Exp.	SD	Base	Exp.	SD	Base	Exp.	SD	Base	Exp.	SD	
1984	62.8	34	53.7	34	50.6	36	42.6	38	50.6	36	42.6	38	50.6	36	42.6	38
Clean Code: A Handbook of Agile Software Craftsmanship	62.5	29	57.6	28	11.4	25	17.7	30	11.4	25	17.7	30	11.4	25	17.7	30
Sapiens: A Brief History of Humankind	54.5	33	53.7	36	31.1	37	33.1	38	31.1	37	33.1	38	31.1	37	33.1	38
Circe	47.5	36	34.5	27	10.9	25	6.0	14	10.9	25	6.0	14	10.9	25	6.0	14
Game of Thrones	63.1	31	49.1	36	69.2	29	53.0	32	69.2	29	53.0	32	69.2	29	53.0	32
The Art of Statistics: How to Learn from Data	74.3	28	72.2	24	15.1	28	11.2	20	15.1	28	11.2	20	15.1	28	11.2	20
The Midwife of Auschwitz	42.4	34	42.0	29	5.6	12	9.9	17	5.6	12	9.9	17	5.6	12	9.9	17
Guns Germs and Steel: The Fates of Human Societies	38.1	33	41.9	28	8.8	20	8.9	16	8.8	20	8.9	16	8.8	20	8.9	16

*Note:* The following questions were asked.

*Matches Interest:* The book recommended to me matches my interests.

*Familiarity:* I know this book well. (Never heard of it = 1. Read it recently = 100)

*Diverse:* The book is different from books I typically buy/read.

*Serendipity:* The book recommended to me is a pleasant surprise.

*Satisfaction:* I am satisfied with the recommendation. The base group consisted of 51 participants, while there were 48 participants in the experimental group.

## 4 Discussion

It is very likely, that many of the scales did not yield significant results due to the sample size being smaller than optimal. While each group consisted of around 50 participants, 90 participants per group would have been ideal. Nonetheless, some interesting pointers can be drawn from the results. Noteworthy is also, that the standard deviation of the experimental group was smaller than that of the base group across all answers in the exit questionnaire. This is likely due to the experimental group being more engaged with the recommender compared to the blander version the base group was presented with.

While the experimental group rated the recommender more favorable (see Table 3, Question P2) at an average of 72.9 vs 66.1, many of the other items in the exit questionnaire were rated worse or on a similar level. A possible explanation could be that, due to the additional context, the experimental group paid closer attention when evaluating the items, which resulted in a worse judgement of the recommender system. Looking at how well the recommended books matched the participants interests, the experimental group rated all books, except the last, worse than the base group. In the case of books 4 and 5 the difference was as large as 14 points. A similar difference could then also be found in the rating of serendipity and satisfaction of those books. While this could be an influence of the additional context clues provided by the experimental recommender system, it could also be due to chance in the distribution of the participants. After all, the familiarity rating of the same books was also lower by a similar score (for example 16 points for book 5).

The expectation was to find an increase of serendipity when using the experimental recommender, which means that people would be more willing to read a book they do not yet know. While the combined rating of serendipity of the experimental group was lower than that of the base group (46.5 vs 49.5, see Table 4), some individual books may have skewed the results. For example, the book "A Game of Thrones" was rated very low by the experimental group (see Table 5). It appears that if a book is well known, different from what the participant typically reads and does not match their interest, then serendipity and satisfaction ratings are lower. This logic applies to well known examples in this study, namely "1984", "A Game of Thrones" and "Sapiens: A Brief History of Humankind". For the lesser known books in the study, the experimental recommender was rated similar or better than the base version with the exception of "Circe". In the cases of "The Midwife of Auschwitz" and "Guns Germs and Steel: The Fates of Human Societies", the experimental recommender was rated around 10 points higher in



serendipity and satisfaction, while the other scales were rated very similarly.

If the assumption that the experimental group engaged more thoroughly with the recommender than the base group is correct, it might be an explanation as to why they rated well known books less serendipitous. According to the results (see Table 2) the experimental group spent more time on all books but one and less time on the entry and exit questionnaires.

Overall the trust rating of the experimental recommender did not improve in comparison to the base version. This somewhat conflicts with other research which found that merely providing an explanation should increase the trust rating of recommenders (Liao et al., 2022). On the other hand, the participants spent more time with the experimental version, which suggests they engaged more thoroughly with the items, even if this does not reflect in a higher trust rating in the exit (trust) questionnaire.

As the experimental version of the recommender did not provide any information that would give extra insight into the books, the results imply that designing recommenders in a way to be more appealing to its users, also increases their engagement with the items themselves. Similar results have been found by Afridi (2019) or Afridi (2018).

## Limitations

As the participants were recruited from a course on recommender systems for computer science students at the TU Vienna, the individuals of the sample were very similar in demographic terms. In addition, they were promised extra points for participation. Participants who rushed through the study were filtered out, but some might have taken time while not paying attention. The group of participants was also smaller than optimal, which resulted in non-significant results where studying a larger group might have led to different results. The participants were able to take the study on their devices, which means that it was impossible to control the setting of the study.

Due to pre-selecting mostly popular books, the possibility for serendipitous items was reduced, which might have negatively influenced the results of the study. A future study should aim for a larger group of participants and use an actual recommender instead of pre-selected items. It would be interesting to see, how a different set of items would fare in the eyes of the participants. According to Table 5, the participants of both groups were familiar with around half of the books presented.

The Human Computer Trust scale (Madsen & Gregor, 2000) was used, of

which the items only partially fit. Many of the items had to be discarded or reworded in order to fit our case, which speaks for a scale that only fits a narrow case. Ideally a more fitting or general scale could be developed for future investigations of the subject matter.

## **Conclusion**

Adding an explanation on why items were recommended, triggering various heuristics and providing more context information did not increase the overall trust rating in the experimental recommender. Still, the recommender was rated better than a more bland version. However, a higher serendipity scores was seen with less familiar or publicly known items. This suggests that providing a more enjoyable user experience, which grabs user attention, also leads to a deeper engagement with the content. This, in turn, lead to a stronger opinion on individual items.

## **Future Work**

Participants were familiar with many of the items presented to them by the experimental recommender, which possibly led to a worse evaluation of the system overall. Conducting a study only with lesser known items could provide further insight into how participants perceive a recommender system. If there was a large enough group of participants, enabling and disabling individual items for many test groups might yield interesting results about which parts influence trust and engagement specifically.

## References

- Afridi, A. H. (2018). User Control and Serendipitous Recommendations in Learning Environments. *Procedia Computer Science*, 130, 214–221. <https://doi.org/10.1016/j.procs.2018.04.032>
- Afridi, A. H. (2019). Transparency for beyond-accuracy experiences: A novel user interface for recommender systems. *Procedia Computer Science*, 151, 335–344. <https://doi.org/10.1016/j.procs.2019.04.047>
- Forschungsdatenzentrum Bildung. (2023, May). Retrieved May 12, 2023, from [https://www.fdz-bildung.de/konstrukt.php?construct%5C\\_id=565](https://www.fdz-bildung.de/konstrukt.php?construct%5C_id=565)
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002, July). *Heuristics and Biases: The Psychology of Intuitive Judgment* (First). Cambridge University Press. <https://doi.org/10.1017/CBO9780511808098>
- Harman, J. L., O'Donovan, J., Abdelzaher, T., & Gonzalez, C. (2014). Dynamics of human trust in recommender systems. *RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems*, 305–308. <https://doi.org/10.1145/2645710.2645761>
- Jannach, D., Quadrana, M., & Cremonesi, P. (2022). Session-Based Recommender Systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 301–334). Springer US. [https://doi.org/10.1007/978-1-0716-2197-4\\_8](https://doi.org/10.1007/978-1-0716-2197-4_8)
- Jarvenpaa, S. L., Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. *Information Technology and Management*, 1(1/2), 45–71. <https://doi.org/10.1023/A:1019104520776>
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>

- Liao, M., Sundar, S. S., & Walther, J. B. (2022). User Trust in Recommendation Systems: A comparison of Content-Based, Collaborative and Demographic Filtering. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3491102.3501936>
- Maccatrozzo, V., Terstall, M., Aroyo, L., & Schreiber, G. (2017). SIRUP: Serendipity In Recommendations via User Perceptions. *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, 35–44. <https://doi.org/10.1145/3025171.3025185>
- Madsen, M., & Gregor, S. (2000). Measuring Human-Computer Trust.
- Nguyen, T. T., Maxwell Harper, F., Terveen, L., & Konstan, J. A. (2018). User Personality and User Satisfaction with Recommender Systems. *Information Systems Frontiers*, 20(6), 1173–1189. <https://doi.org/10.1007/s10796-017-9782-y>
- Nikolakopoulos, A. N., Ning, X., Desrosiers, C., & Karypis, G. (2021). Trust your neighbors: A comprehensive survey of neighborhood-based methods for recommender systems. <https://doi.org/10.48550/ARXIV.2109.04584>
- Pina E Cunha, M., Rego, A., Clegg, S., & Lindsay, G. (2015). The dialectics of serendipity. *European Management Journal*, 33(1), 9–18. <https://doi.org/10.1016/j.emj.2014.11.001>
- Raza, S., & Ding, C. (2022). News recommender system: A review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, 55(1), 749–800. <https://doi.org/10.1007/s10462-021-10043-x>
- Ricci, F., Rokach, L., & Shapira, B. (2022a). Recommender Systems Handbook Preface. In *Recommender Systems Handbook* (Third, pp. vii–viii). Springer US. <https://doi.org/10.1007/978-1-0716-2197-4>
- Ricci, F., Rokach, L., & Shapira, B. (2022b). Recommender Systems: Techniques, Applications, and Challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 1–35). Springer US. [https://doi.org/10.1007/978-1-0716-2197-4\\_1](https://doi.org/10.1007/978-1-0716-2197-4_1)

- Shin, D. (2020). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, 109, 106344. <https://doi.org/10.1016/j.chb.2020.106344>
- Stray, J., Halevy, A., Assar, P., Hadfield-Menell, D., Boutilier, C., Ashar, A., Beat-  
tie, L., Ekstrand, M., Leibowicz, C., Sehat, C. M., Johansen, S., Kerlin, L.,  
Vickrey, D., Singh, S., Vrijenhoek, S., Zhang, A., Andrus, M., Helberger,  
N., Proutskova, P., ... Vasan, N. (2022). Building Human Values into Rec-  
ommender Systems: An Interdisciplinary Synthesis. <https://doi.org/10.48550/ARXIV.2207.10192>
- Sundar, S. S. (2007). The MAIN Model: A Heuristic Approach to Understanding  
Technology Effects on Credibility. *Digital Media*.
- Ziarani, R. J., & Ravanmehr, R. (2021). Serendipity in Recommender Systems:  
A Systematic Literature Review. *Journal of Computer Science and Tech-  
nology*, 36(2), 375–396. <https://doi.org/10.1007/s11390-020-0135-9>

## 5 Additional Data

While not of interest for answering the hypothesis, the entry questionnaire data might be relevant to some readers. In the self-assessment part of the entry questionnaire, participants answered 7 questions about their reading attitudes. The results can be found in table 6.

**Table 6**

Answers to the self-evaluation of reading habits. Answers on a scale from 1 (not at all) to 100 (fully agree).

Question	Average	SD
By comparison, I read a lot	51.7	25.3
I think reading texts related to your profession/studies/work is important.	76.3	22.4
I think reading texts recreationally (e.g. novels) is important.	57.2	27.5
I think reading news articles, online news or similar is important.	62.8	27.1
I think reading educational texts, unrelated to your work/profession/studies is important.	58.8	24.4
I only read if I have to.	25.0	30.0
I like visiting bookstores or browsing online shops for books.	42.8	36.0

**Peer-Gutachten zur Bachelorarbeit**

Gutachter/in der Bachelorarbeit: Stefan Pflieger
Verfasser/in der Bachelorarbeit: Tobias Sippl
Titel der Bachelorarbeit: Improving Trust in Recommender Systems through Context Clues

Bewertung: 0 = erfüllt die Anforderungen nicht; 4 = erfüllt die Anforderungen in hohem Maße

**I. ALLGEMEINE ASPEKTE**

1. Aufbau und Gliederung; „roter Faden“	0	1	2	3	4
2. Klarheit und Verständlichkeit	0	1	2	3	4
3. Strukturierung & Organisation der Informationsdarstellung	0	1	2	3	4
4. Formale Korrektheit (APA-Richtlinien; Korrektheit Zitate)	0	1	2	3	4
5. Korrektheit der sprachlichen Fassung (Rechtschreibung, Grammatik)	0	1	2	3	4
6. Gestaltung (Layout, Tabellen)	0	1	2	3	4

**II. THEORIETEIL**

1. Präzision der Darstellung	0	1	2	3	4
2. Argumentationsstruktur	0	1	2	3	4
3. Angemessenheit der berücksichtigten Literatur	0	1	2	3	4
4. Begründung der theoretischen Fragestellung	0	1	2	3	4
5. Formulierung der Hypothesen	0	1	2	3	4

**III. METHODENTEIL**

1. Qualität des Untersuchungsplans	0	1	2	3	4
2. Qualität der Erhebungsinstrumente	0	1	2	3	4
3. Qualität der Operationalisierungen	0	1	2	3	4
4. Angemessenheit der Beschreibung der Stichprobe	0	1	2	3	4
5. Dokumentation des Ablaufs der Datenerhebung	0	1	2	3	4

**VI. LITERATURLISTE**

1. Korrektheit der Zitierung	0	1	2	3	4
2. Umfang	0	1	2	3	4

**VII. ZUSÄTZLICHE PUNKTE**

1.	0	1	2	3	4
2.	0	1	2	3	4
3.	0	1	2	3	4

**FREIE GESAMTBEURTEILUNG (inkl. besonders positive Aspekte, Verbesserungspotential, offene Fragen; Achtung: Beurteilungen mit „0“ oder „1“ sollten in jedem Fall begründet werden!):**

Wie erwartet eine ausgezeichnete Arbeit, in der es von meiner Seite auf formaler Ebene kein Verbesserungspotenzial aufzuweisen gibt! Ein paar Flüchtigkeitsfehler habe ich dir im Dokument

angestrichen.

Beim Lesen erschließen sich mir sämtliche Aspekte der Bachelorarbeit, der Lesefluss gewährleistet das.

Wie in der vorangegangenen Fachliteraturarbeit sind sämtliche Variablen, Hypothesen und Begriffe auf eine Weise definiert und erklärt, dass sich auch ein\*e Thematikfremde\*r gut zurechtfindet.

Dasselbe trifft auch auf die Methodenbeschreibung zu. Stichprobe und Erhebungsmethode werden schlüssig beschrieben.

Ergebnisse und Diskussion sind korrekt getrennt, bieten einen guten Einblick in die erhobenen Daten und lassen Rückschluss auf einen reflektierten Umgang damit zu. Das betrifft die Interpretation der Ergebnisse, wie auch die zukünftigen Implikationen in der Thematik.

Anmerkungen:

Ich weiß, in einem veröffentlichten Paper nicht üblich und auch eine stilistische Angelegenheit. Aber in einer Bachelorarbeit würde sich vielleicht, vor allem bei deinen strukturierten Überschriften, ein Inhaltsverzeichnis anbieten.

Und die „Wir“-Thematik wollte ich auch noch hier erwähnen. Entsteht der ab und an ausgebliebene Passiv und die Wir-Form durch die Kooperation mit der TU?

Ort, Datum: Weyregg am Attersee, 10.06.2023

Unterschrift Gutachter/in:



**Peer-Gutachten zur Bachelorarbeit**

Gutachter/in der Bachelorarbeit: Stefan Pflieger
Verfasser/in der Bachelorarbeit: Tobias Sippl
Titel der Bachelorarbeit: Improving Trust in Recommender Systems through Context Clues

Bewertung: 0 = erfüllt die Anforderungen nicht; 4 = erfüllt die Anforderungen in hohem Maße

**I. ALLGEMEINE ASPEKTE**

1. Aufbau und Gliederung; „roter Faden“	0	1	2	3	4
2. Klarheit und Verständlichkeit	0	1	2	3	4
3. Strukturierung & Organisation der Informationsdarstellung	0	1	2	3	4
4. Formale Korrektheit (APA-Richtlinien; Korrektheit Zitate)	0	1	2	3	4
5. Korrektheit der sprachlichen Fassung (Rechtschreibung, Grammatik)	0	1	2	3	4
6. Gestaltung (Layout, Tabellen)	0	1	2	3	4

**II. THEORIETEIL**

1. Präzision der Darstellung	0	1	2	3	4
2. Argumentationsstruktur	0	1	2	3	4
3. Angemessenheit der berücksichtigten Literatur	0	1	2	3	4
4. Begründung der theoretischen Fragestellung	0	1	2	3	4
5. Formulierung der Hypothesen	0	1	2	3	4

**III. METHODENTEIL**

1. Qualität des Untersuchungsplans	0	1	2	3	4
2. Qualität der Erhebungsinstrumente	0	1	2	3	4
3. Qualität der Operationalisierungen	0	1	2	3	4
4. Angemessenheit der Beschreibung der Stichprobe	0	1	2	3	4
5. Dokumentation des Ablaufs der Datenerhebung	0	1	2	3	4

**VI. LITERATURLISTE**

1. Korrektheit der Zitierung	0	1	2	3	4
2. Umfang	0	1	2	3	4

**VII. ZUSÄTZLICHE PUNKTE**

1.	0	1	2	3	4
2.	0	1	2	3	4
3.	0	1	2	3	4

**FREIE GESAMTBEURTEILUNG (inkl. besonders positive Aspekte, Verbesserungspotential, offene Fragen; Achtung: Beurteilungen mit „0“ oder „1“ sollten in jedem Fall begründet werden!):**

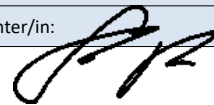
Es gibt dem ersten Feedback nichts mehr anzufügen.  
Angesprochene Punkte wurden in Betracht gezogen und bei Notwendigkeit auch umgesetzt.

Peergutachten zur Bachelorarbeit SE Bachelorseminar

Damit bleibt mir nur noch übrig zur ausgezeichneten Bachelorarbeit zu gratulieren und dir alles Gute auf deinem weiteren Weg zu wünschen, wie er auch immer mit Psychologie in Berührung kommen möge 😊

Ort, Datum: Wien, 17.06.2023

Unterschrift Gutachter/in:

A handwritten signature in black ink, appearing to be 'A. R.', written over the signature line.

## Peer-Gutachten zur Bachelorarbeit

Gutachter/in der Bachelorarbeit: Linda Basso
Verfasser/in der Bachelorarbeit: Tobias Sippl
Titel der Bachelorarbeit: "Improving Trust in Recommender Systems through Context Clues"

Bewertung: 0 = erfüllt die Anforderungen nicht; 4 = erfüllt die Anforderungen in hohem Maße

### I. ALLGEMEINE ASPEKTE

1. Aufbau und Gliederung; „roter Faden“	0	1	2	3	4
2. Klarheit und Verständlichkeit	0	1	2	3	4
3. Strukturierung & Organisation der Informationsdarstellung	0	1	2	3	4
4. Formale Korrektheit (APA-Richtlinien; Korrektheit Zitate)	0	1	2	3	4
5. Korrektheit der sprachlichen Fassung (Rechtschreibung, Grammatik)	0	1	2	3	4
6. Gestaltung (Layout, Tabellen)	0	1	2	3	4

### II. THEORIETEIL

1. Präzision der Darstellung	0	1	2	3	4
2. Argumentationsstruktur	0	1	2	3	4
3. Angemessenheit der berücksichtigten Literatur	0	1	2	3	4
4. Begründung der theoretischen Fragestellung	0	1	2	3	4
5. Formulierung der Hypothesen	0	1	2	3	4

### III. METHODENTEIL

1. Qualität des Untersuchungsplans	0	1	2	3	4
2. Qualität der Erhebungsinstrumente	0	1	2	3	4
3. Qualität der Operationalisierungen	0	1	2	3	4
4. Angemessenheit der Beschreibung der Stichprobe	0	1	2	3	4
5. Dokumentation des Ablaufs der Datenerhebung	0	1	2	3	4

### VI. LITERATURLISTE

1. Korrektheit der Zitierung	0	1	2	3	4
2. Umfang	0	1	2	3	4

### VII. ZUSÄTZLICHE PUNKTE

1.	0	1	2	3	4
2.	0	1	2	3	4
3.	0	1	2	3	4

**FREIE GESAMTBEURTEILUNG (inkl. besonders positive Aspekte, Verbesserungspotential, offene Fragen; Achtung: Beurteilungen mit „0“ oder „1“ sollten in jedem Fall begründet werden!):**

Die Arbeit ist wirklich sehr gelungen, spannend, gut recherchiert und gründlich ausgearbeitet.  
Ich finde den Aufbau in der Einleitung sehr passend, indem du die Leser\*innen grob in das Thema einführst

-und auch einen Alltagsbezug für alle herstellt durch das Erwähnen der für jedermann bekannten Plattformen wie Netflix und Spotify- und dann die Theorie immer detaillierter behandelst. Ich finde den Einsatz der zwei direkten Zitate sehr passend gewählt, um die Theorie augenscheinlicher zu berichten und den Lesefluss abwechslungsreicher zu gestalten.

Auf S. 7 erwähnst du das „more diverse set of items“. Hier könnte man noch kurz den erwähnten diversity-Begriff in den Kontext einbetten und etwas genauer erläutern, was mit „diverse-set“ gemeint sei. Die Grafiken und Figuren sind sehr passend gewählt und unterstützen die Leser\*inne bei einem klareren Verständnis der Studiendurchführung. Du beschreibst die Studie sehr ausführlich und klar, sodass das Studiendesign sehr verständlich und nachvollziehbar vermittelt wird. Auch die Ergebnisdarstellung ist sehr übersichtlich und sauber ausgearbeitet.

Vielleicht könnte man den Discussion-Teil in einem Satz noch etwas allgemeiner einleiten und dann erst die möglichen Erklärungen der Ergebnisse behandeln.

Ich habe nun im folgenden einige Anmerkungen zur Sprache, kleine Flüchtigkeitsfehler und der sprachlichen Form herausgearbeitet.

Im Abstract hat sich ein kleiner Flüchtigkeitsfehler eingeschlichen: bei „thereforE“ fehlt das „e“.

Auf S. 3 schreibst du „We are interested in how the presentation of items influences how they are perceived“. Um einen vielleicht etwas „saubereren“ Sprachgebrauch zu verwednen, könnte man auf den doppelten Gebrauch von „how“ in einem Satz verzichten und möglicherweise eine Formulierung verwenden wie. „We are interested in how the presentation of items influences their perception by the user.“

Auf S. 3 ganz unten, bei “they take ON many (various) forms“ fehlt ein “on“.

Auf S. 4 fehlt für das korrekte Zitat „(Jannach et al., 2022)“ die Klammer.

Auf S. 8 ganz unten wird ein Mal in das Präsens gewechselt („...that we ARE testing a book-recommender“), während der Rest der Studienbeschreibung in der Vergangenheit erzählt wird.

Das auf S. 14 verwendete Verb “boost“ in: „should boost the serendipity perception“ wird normalerweise eher umgangssprachlich verwednet. Ein mögliches Synonym in dem Kontext könnte vielleicht „increase“ oder „enhance“.

Auf S. 17 bei den limitations fehlt ein „p“ bei „samPles“.

Auf S. 18 ist der letzte Absatz („Future Work“) etwas unverständlich formuliert; als Leserin fällt es etwas schwer, diesem Absatz in verständlicher Weise gut folgen zu können. Möglicherweise könnte eine kleine Umformulierung das Verständnis erleichtern.

Ich finde deine Bachelorarbeit wirklich sehr gelungen, klar verständlich, gut strukturiert und ausgearbeitet.

**Peer-Gutachten zur Bachelorarbeit**

Gutachter/in der Bachelorarbeit: Linda Basso
Verfasser/in der Bachelorarbeit: Tobias Sippl
Titel der Bachelorarbeit: "Improving Trust in Recommender Systems through Context Clues"

Bewertung: 0 = erfüllt die Anforderungen nicht; 4 = erfüllt die Anforderungen in hohem Maße

**I. ALLGEMEINE ASPEKTE**

1. Aufbau und Gliederung; „roter Faden“	0	1	2	3	4
2. Klarheit und Verständlichkeit	0	1	2	3	4
3. Strukturierung & Organisation der Informationsdarstellung	0	1	2	3	4
4. Formale Korrektheit (APA-Richtlinien; Korrektheit Zitate)	0	1	2	3	4
5. Korrektheit der sprachlichen Fassung (Rechtschreibung, Grammatik)	0	1	2	3	4
6. Gestaltung (Layout, Tabellen)	0	1	2	3	4

**II. THEORIETEIL**

1. Präzision der Darstellung	0	1	2	3	4
2. Argumentationsstruktur	0	1	2	3	4
3. Angemessenheit der berücksichtigten Literatur	0	1	2	3	4
4. Begründung der theoretischen Fragestellung	0	1	2	3	4
5. Formulierung der Hypothesen	0	1	2	3	4

**III. METHODENTEIL**

1. Qualität des Untersuchungsplans	0	1	2	3	4
2. Qualität der Erhebungsinstrumente	0	1	2	3	4
3. Qualität der Operationalisierungen	0	1	2	3	4
4. Angemessenheit der Beschreibung der Stichprobe	0	1	2	3	4
5. Dokumentation des Ablaufs der Datenerhebung	0	1	2	3	4

**VI. LITERATURLISTE**

1. Korrektheit der Zitierung	0	1	2	3	4
2. Umfang	0	1	2	3	4

**VII. ZUSÄTZLICHE PUNKTE**

1. Ergebnisdarstellung	0	1	2	3	4
2. Diskussion	0	1	2	3	4
3.	0	1	2	3	4

**FREIE GESAMTBEURTEILUNG (inkl. besonders positive Aspekte, Verbesserungspotential, offene Fragen; Achtung: Beurteilungen mit „0“ oder „1“ sollten in jedem Fall begründet werden!):**

**Titelblatt:**

Da das Titelblatt bis auf den Titel deiner Arbeit in Deutsch gehalten ist, würde ich vorschlagen den deutschen Begriff „Bachelorarbeit“ zu verwenden, oder das gesamte Titelblatt auf Englisch zu verfassen.

**Allgemein:**

Ich finde deine sehr aktuellen Anmerkungen bezüglich ChatGPT und co. sehr gelungen, dadurch schaffst du mit deiner Arbeit einen klaren Bezug zur gegenwärtigen Realität der user\*innen und weckst direkt Interesse zum Weiterlesen.

Allgemein ist der Lesefluss im ersten Teil der Introduction nun sehr schlüssig und die Übergänge auch noch „geschmeidiger“, was den Leser\*innen das klare Verständnis erleichtert und der Text flüssig gelesen werden kann.

**Theorieteil:**

Die Argumentationslinie ist sehr klar, nachvollziehbar und gut strukturiert. Die einzelnen Absätze von Abschnitt „1.1 RS“ könnten durch das Einsetzen von einzelnen Füllwörtern wie „moreover“, oder „in conclusion“ noch flüssiger miteinander verbunden werden, wie du es in den weiteren Introduction-Abschnitten gelungen gemacht hast.

Hypothesen: Der Vollständigkeit halber könnte man eventuell die jeweiligen H0 zu den Hypothesen noch anführen.

Methodenteil: Dir ist es gelungen, deine Methode sehr anschaulich und verständlich zu beschreiben.

**Diskussion:**

Der erste Teil deiner Diskussion ist sehr ausführlich, du gibst einige Denkanstöße und mögliche Erklärungen, wie die Ergebnisse deiner Studie zustande gekommen sein könnten. Die Limitationen könnten vielleicht noch etwas ausführlicher beleuchtet werden, vielleicht anhand eines Beispiels um die aussage „The scale was used, of which the items only partially fit“ noch etwas verständlicher zu untermauern.

Im letzten Abschnitt „Future Work“ S. 18 fände ich im ersten Satz den Begriff „participants“ anstelle von „users“ (was die Gesamtheit aller user\*innen weltweit impliziert) noch passender.

**Formale Aspekte:**

Du erfüllst meines Erachtens nach die Richtlinien nach APA größtenteils genau (zB Einrücken der Absätze im Theorieteil, Formalia der Überschriften/Unterüberschriften usw).

**Einige kleine Anmerkungen:**

- Seitenzahlen kommen nach APA in den oberen rechten Seitenrand
- Überschriften der Hauptkapitel (Introduction, Method, Results, Discussion, References) werden zentriert
- Tabellen sollten einen kurzen Titel haben, welcher nach APA kursiv geschrieben werden soll. Die „Note“ zu den Tabellen sollte unter die Tabelle eingefügt werden, nicht über die Tabelle (Table 1, S. 9; Table 2, S. 12)
- Bei Table 3 und Table 4 könnte die Erläuterung „a p smaller than 0.05 would be significant and show that the experimental group rated the questions higher“ entweder in einen kurzen Fließtext oder ebenfalls in die Notes geschrieben werden, bin mir nicht sicher ob diese Erläuterung in den Titel der Tabelle gehört.
- Du könntest noch einen Table of Contents nach deinem Titelblatt einfügen, ebenso eine „List of Tables“ und „List of Figures“.

Anhang: Ich nehme an du gibst noch den fertigen, verwendeten Fragebogen/RS in den Anhang. Fand es sehr beeindruckend, dass du ein ganzes Book-RS selbst simuliert hast, das darfst du deiner Leserschaft nicht vorenthalten.

Deine Arbeit ist sehr gelungen, spannend, aufschlussreich und interessant.

Ort, Datum: Wien, 26.06.2023

Unterschrift Gutachter/in: Linda Basso