Effects of Argumentative Explanation Types on the Perception of Review-Based Recommendations

Diana C. Hernandez-Bocanegra
University of Duisburg-Essen
Duisburg, Germany
diana.hernandez-bocanegra@uni-due.de

Tim Donkers
University of Duisburg-Essen
Duisburg, Germany
tim.donkers@uni-due.de

Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
juergen.ziegler@uni-due.de

ABSTRACT

Recommender systems have achieved considerable maturity and accuracy in recent years. However, the rationale behind recommendations mostly remains opaque. Providing textual explanations based on user reviews may increase users' perception of transparency and, by that, overall system satisfaction. However, little is known about how these explanations can be effectively and efficiently presented to the user. In the following paper, we present an empirical study conducted in the domain of hotels to investigate the effect of different textual explanation types on, among others, perceived system transparency and trustworthiness, as well as the overall assessment of explanation quality. The explanations presented to participants follow an argument-based design, which we propose to provide a rationale to support a recommendation in a structured way. Our results show that people prefer explanations that include an aggregation using percentages of other users' opinions, over explanations that only include a brief summary of opinions. The results additionally indicate that user characteristics such as social awareness may influence the perception of explanation quality.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Human-centered computing → User studies**.

KEYWORDS

Recommender systems, user study, explanations

ACM Reference Format:
Diana C. Hernandez-Bocanegra, Tim Donkers, and Jürgen Ziegler. 2020. Effects of Argumentative Explanation Types on the Perception of Review-Based Recommendations. In *Adjunct Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20 Adjunct), July 14–17, 2020, Genoa, Italy*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3386392.3399302

UMAP '20 Adjunct, July 14–17, 2020, Genoa, Italy

ACM ISBN 978-1-4503-7950-2/20/07…$15.00
https://doi.org/10.1145/3386392.3399302

1 INTRODUCTION

Providing explanations of the rationale behind a recommendation can bring several benefits to recommender systems (RS). In particular, explanations may serve the following aims [27]: transparency (the system explains how it works), effectiveness (user can make good decisions), efficiency (user can make decisions faster), and trust in the system. Explanations based on collaborative filtering inform that a recommendation is based on preferences of similar users or items that the user liked in the past, e.g. Amazon's "Customers who bought … also bought...", while content-based explanations present users with item features that can be relevant to them, e.g. [15, 29]. On the other hand, exploiting user reviews has drawn research interest recently, in particular to facilitate the generation of textual explanations, as proposed by [34] and [18], where a brief assessment of relevant aspects based on opinions from reviews is provided as explanation. However, avid users in need of specific details may be more satisfied when more robust arguments or a higher level of justification is provided. Here, important questions are still unresolved: Do users prefer concise explanations over those that include more specific details? Do they prefer an aggregated view of other users' opinions, over reading individual reviews written by similar users? Specifically, we regarded three different types of review-based explanation:

- Explanations with aggregated results: An accumulated view using bullet points and percentages of positive and negative opinions, as proposed by [11].
- Explanations with only textual summary: Summarization of opinions without bullet points nor percentages. It resembles a system generated review, as proposed by [8], and [3].
- Explanations using a helpful review: Indicate that the recommendation was based on the reviews that might be helpful to the user, as proposed by [6], and show just one of them as an example.

In this respect, both summaries and helpful reviews have proven to be an effective means of assisting users in making purchasing decisions, while helping them cope with the overwhelming amount of information available [3, 12, 16, 20, 23]. However, little is known about the suitability of one information style over another when offered as part of an explanation. Additionally, we also aimed to evaluate the effect of two levels of justification:

- High: Specific details about the main aspects (e.g. cleanliness) and finer-grained aspects (e.g. cleanliness of bathroom) are provided.
- Low: Only brief information about main aspects is provided.

In addition, and taking into account that differences in user characteristics also contribute to differences in the general perception of RS [17, 31], we set out to focus on one of the main objectives of RS, which is to help users make better decisions. Particularly, individual differences between decision-making styles are determined significantly by preferences and abilities to process available information [9]. Accordingly, decision making styles are defined by [14] as a "habit-based propensity" to exhaustively search for information and to systematically evaluate possible alternatives (rational style), or to use a quick process based on hunches and feelings (intuitive style), in order to make decisions. Two main aspects provide basis to describe the differences in decision styles: information use (amount of information used during the process) and focus (alternatives addressed) [9]. In this respect, "good enough" information might be sufficient for some people, whereas others prefer to obtain and address all relevant information in order to minimize risks.

Additionally, we were interested in a second factor that may influence the way users perceive explanations: the extent to which they are able to adopt the perspective of others when making decisions. The rationale for this interest stems from the tendency of individuals to adjust their own opinions using those of others, while choosing between various alternatives [26], which may even be beneficial [32]. Particularly, individuals with greater perspective-taking skills tend to understand the views of others better [2, 5], skills that are also characterized by [10] as "social awareness", which represents the propensity of individuals to empathize and take into account the opinions of others.

Accordingly, we aimed to answer the following research questions, in relation to our variables of interest (i.e. quality of explanation, transparency, effectiveness, efficiency and trust), and taking the hotels domain as a case in point:

RQ1: Does the *type* of explanation influence the perception of the variables of interest?

RQ2: Does the *level* of justification influence the perception of the variables of interest?

RQ3: Do individual differences in decision making styles and social awareness influence the perception of the variables of interest, when different types of explanation or levels of justification are provided?

Consequently, we conducted a study, in which users were asked to examine and read the explanations of a fixed set of hotel recommendations, and to report their perception of the quality of such explanations, as well as their perceived transparency, effectiveness, efficiency and trust of the system. The hand-made explanations provided were based on designed templates that follow principles of argumentation theory, as elaborated in detail in section 3.

2 RELATED WORK

Textual explanations in RS seek to provide reasons behind a recommendation, while assisting users making a decision. In this respect, in recent years there has been a growing interest in exploiting user reviews, given their richness in explanatory and argumentative information. [34] proposed a matrix factorization model to align explicit features and the latent representations of items and user preferences obtained from reviews, which allows to generate textual explanations based on templates (e.g. "You might be interested in [feature], on which this product performs well"). An extension of this work was presented by [6], who argued that reviews should have different weights when calculating predictions, and that, therefore, the most useful for the user should have a higher priority, and be used to generate explanations; however, no explanations are actually generated, but only selected reviews are provided. On the other hand, [8] proposed an natural language generation (NLG) procedure for creating reviews (as a real user would) and providing them as explanations without using templates, whereas [7] proposed a denoising mechanism to extract relevant sentences with explainable purposes, to generate natural language textual explanations (e.g. "The bottle is very light and the smell is very strong"). Additionally, [21] had proposed a series of interface variations, that provide users with display pros and cons scores using bars, as well as a report of feature performance in comparison with other alternatives; however, their visualizations do not provide details on the fine-grained aspects, nor possible reasons for conflicting opinions.

The above approaches result in explanations that may be perceived by users as being too general, and lacking solid arguments to justify the recommendation offered. On the other hand, [4] proposed a framework to generate arguments in the context of tasks like selecting a house to buy. [33] compares explanations with brief sentences and an argumentative structure - two facts and a claim -, for recommendations of hiking routes, energy and mobile phone plans; however, no counter-arguments are provided. [18] proposed a method based on [1] for generating explanations with convincing arguments in a mobile shopping recommender using templates: strong argument (e.g. "Mainly because you currently like X."), supporting argument (e.g. "Also, slightly because of your current interest in X.", and negative argument (e.g. "However, it has the following features you don't like: X, Y (...)."). The rather concrete and brief sentences proposed by [1] and [18] are oriented to provide interactive explanations in the mobile domain, where users might face both space and time limitations. However, we aimed to investigate the effect that more detailed explanations may have on users' perception of recommender systems, while keeping an argumentative nature. To this end, we propose an explanations design with an argumentative structure, that is inspired on the scheme proposed by [13], a variation of original Toulmin's model [28], that seeks to represent the kind of arguments usually provided in user-generated web discourse.

3 EXPLANATION DESIGN

We designed a series of templates that represent the combination of the two factors: *type* of textual explanation and *level* of justification. These templates were used to create the explanations we presented to participants in our empirical study. Table 1 shows the designed templates. Furthermore, the proposed design reflects an argumentative structure, inspired by the scheme proposed by Habernal et al. [13], and includes: a conclusion that informs how good the choice is for the user, evidence that supports such a claim, and possible reasons behind contradictory opinions.

Additionally, we considered a number of template variations in order to explain items with higher prediction ratings (*very good* or an *adequate* option), or lower prediction ratings (*not so good* option), depending on whether positive opinions are much much greater (very good) or greater (adequate) than negative ones, or if they are more negative than positive (not so good). These variations are represented mainly by differences in the rebuttal and the backing section of the explanation, as well as the presence of refutation statements, as depicted in the scheme of figure 1.

Explanations with aggregated results: Summarizes opinions found in reviews using bullet points and percentages of positive and negative opinions. It corresponds to the "Aggregation" condition of the empirical study.

Explanations with only textual summary: Summarizes opinions using just text (no bullet points nor percentages). The condition "Summary" refers to this type of explanation.

Explanations using a helpful review: This type of explanations indicate that recommendation was based on information provided in helpful reviews, and offers one of such reviews as an example. The condition "Review" refers to this type of explanation.

In turn, every type of explanation is provided in one of two variations:

Low level of justification. To address the main aspects of interest to users (e.g. overall cleanliness), without further elaboration or details.

High level of justification. To address the main aspects of interest to users (e.g. overall cleanliness) by providing fine grained details with several sentences about more specific aspects (e.g. cleanliness of bathroom).

4 EMPIRICAL STUDY

We intended to compare the users' subjective assessment of different types of explanation and different levels of justification. In particular, we hypothesized the following:

H1: People will be more satisfied with explanations that involve a higher level of justification.

H2: People will be more satisfied with aggregated explanations as opposed to mere summaries.

H3: People will be more satisfied with explanations that involve helpful reviews as opposed to mere summaries.

H4: More rational users would prefer a higher level of justification and explanations that involve helpful reviews or an aggregation of opinions, as opposed to summaries.

To test the above, we recruited 152 participants (87 female, mean age 39.84 and range between 18 and 75) through Amazon Mechanical Turk. We restricted the execution of the task to workers located in the U.S, with a HIT (Human Intelligence Task) approval rate greater than 95%. Although 334 workers completed the task, only 152 workers passed the quality check (i.e. at least 6 of the 7 validation questions were answered correctly, more than 20s were spent on the recommendation step and more than 30s on the evaluation questionnaire), so only the data for these participants were used for the following analysis. This sample size allows us to achieve a statistical power of 82.5% with the performed MANCOVA analysis (α =0.05). Participants were rewarded with $0.8 (time to complete task in minutes: M=8.56, SD= 1.86)

The study follows a 3x2 between-subjects design, and each participant was assigned randomly to one of six conditions that represent the combination of the two factors: *type* of explanation and *level* of justification. Participants were presented with a prototype that provided them with a fixed list of 5 hotels that represented the recommendations for a hypothetical hotel search. Each recommendation included an explanation of why the item was recommended. After the participants explored the information for all the hotels, they were asked to rate their perception of the recommender and its explanations. No real system was used to generate recommendations or explanations, as the main objective here was to test users' perception of explanation design.

Conditions: We regarded three different types of explanation: with aggregated results ("aggregation"), with only textual summary ("summary") and explanations using a helpful review ("review"). We also evaluated the effect of two levels of justification: "high" and "low". Section 3 provides further details on every type and level.

Procedure: After some questions on demographics, users answered the questionnaire on user characteristics. Instructions to participants indicated that a list of 5 hotels would be displayed, representing the results of a hypothetical search for hotels already performed. Here, participants were instructed to click the button "View Details" of each hotel and read the information provided, including the explanation of why the item was recommended. We then presented a cover story, which sought to establish a common starting point in terms of travel motivation (a business trip), and the presumed aspects of greatest interest to the user (cleanliness and location). The cover story also stated that different recommended hotels within the same price range would be shown. The users were then presented with a list of recommended hotels and their explanations. An example of the functionality provided to the users is shown in figure 2. The list of hotels, hotel names, photos, prices and ratings were the same for all users. Only the explanations provided varied according to the condition to which each participant was assigned. Next, users answered the evaluation questionnaire. In addition, we included an open-ended question, so that participants could indicate in their own words their general opinion about the explanations provided. We included 4 validation questions to check attentiveness within the questionnaires, and 3 validation questions related to the content of both textual and visual elements presented throughout the task.

Questionnaires:

User characteristics: We used the Rational and Intuitive Decision Styles Scale [14], and the scale of the social awareness competency, proposed by [10]. We used a 1-5 Likert-scale to evaluate all the items (1:Strongly disagree, 5: Strongly agree).

Evaluation: We used items from: [25] to measure the perception of transparency, [17] of effectiveness, [19] of efficiency, and [19] of trust. Finally, we also adapted 3 items from [17] to address explanation quality. We used a 1-5 Likert-scale to evaluate all the items (1:Strongly disagree, 5: Strongly agree).

5 RESULTS

User characteristics scores. In regard to decision making styles, we calculated the rational (M = 4.31, SD= 0.52) and the intuitive (M = 2.72, SD= 0.83) scores for each individual as the average of the values reported for the five items on both rational and intuitive decision-making style subscales. Likewise, we calculated the social awareness score (M = 3.99, SD= 0.49) for each individual based on

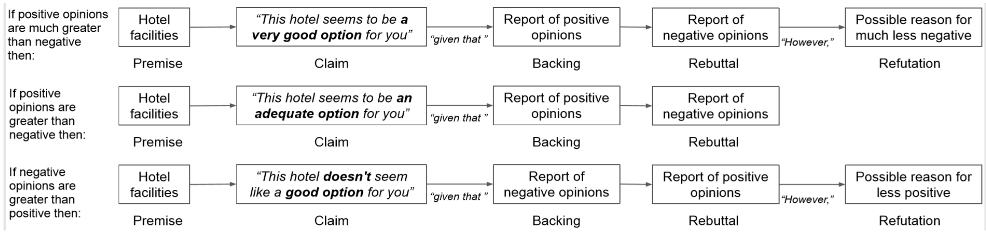

Figure 1: Argument scheme used to create explanation templates, to provide reasons for recommending items with higher prediction ratings (very good or an adequate option), or lower prediction ratings (not so good option).

Aggregation	Summary	Review
Explanation Beginning (Both levels): [It is located in ..., and provides ... in all rooms]$_{\text{Premise}}$. [This hotel seems to be a very good option for you]$_{\text{Claim}}$, given that:	[It is located in ..., and provides ... in all rooms]$_{\text{Premise}}$. [This hotel seems to be an adequate option for you]$_{\text{Claim}}$, given that	[Based on the reviews that contain useful information and might be relevant to you.]$_{\text{Backing}}$, we believe that [this hotel is an adequate option for you]$_{\text{Claim}}$. This is an example of one of these reviews:
Low level: [*n*% of visitors reported positive comments about ... and *n*% about ...]$_{\text{Backing}}$. [Some visitors mentioned negative comments about ... (*n*%)]$_{\text{Rebuttal}}$, however [such claims are seemingly related to particular incidents, rather than a usual situation, or perhaps to very high expectations that were not met.]$_{\text{Refutation}}$	[usually ... is not a problem here, and the ... is ...]$_{\text{Backing}}$. [Although some reviews include negative comments about ...]$_{\text{Rebuttal}}$, [such claims seem to be more related to incidents rather than a usual situation, or perhaps to very high expectations that were not met.]$_{\text{Refutation}}$	["I've visited the River Hotel for a business trip. Coffee and tea in the room, clean, good location, near to ... Overall, a very good option, I would definitely come back!!!"]$_{\text{Backing}}$
High level: ... [visitors reported positive comments about: - The ... (*n*% of reviews), in particular about the state of ... (*n*% of reviews) - The … (*n*% of reviews), especially about ... (*n*% of reviews).]$_{\text{Backing}}$. [Some visitors also mentioned negative comments about these aspects]$_{\text{Rebuttal}}$. However, [due to the lower number of similar comments, this opinions correspond seamingly to: - Incidents rather than a usual situation, related to the state of ... (*n*% of reviews). - Unfulfilled very high expectations related to ... (*n*% of reviews) or ... (*n*% of reviews).]$_{\text{Refutation}}$	[usually ... is not a problem here, in particular the state of ..., and the ... is quite good in general. The ... very convenient for your purposes, since it is ..., and it is also ...]$_{\text{Backing}}$. [Although some reviews include negative comments about ..., in particular in relation to ...]$_{\text{Rebuttal}}$, [such claims seem more related to incidents rather than a usual situation, or perhaps to very high expectations that were not met.]$_{\text{Refutation}}$	["I stayed at the Sofia Hotel in June. The location is convenient to... And very convenient when you need to work and not being disturbed by kids or drunk teenagers! My room was clean but more care for windows wouldn't hurt. Also, I think left the towels ..., I expected them to be changed, but that didn't happen until ..., but overall a minor issue, given the overall quality of the room. Parking is free, but you may not need it, as ... Overall, you get what it is advertised. I'd come back"]$_{\text{Backing}}$

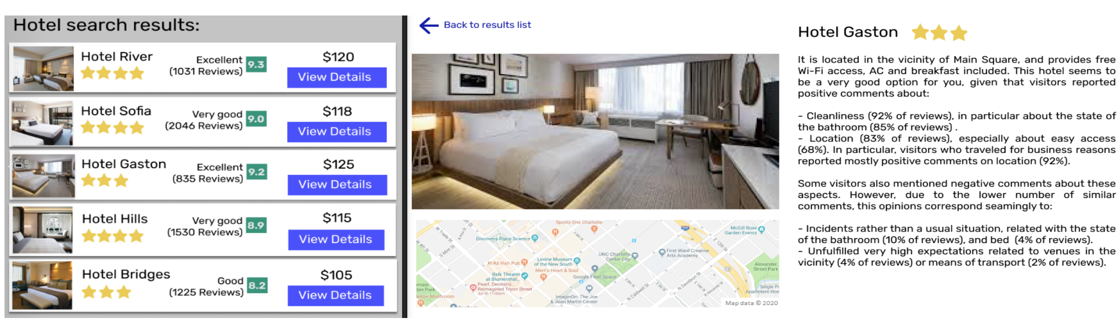

Figure 2: Prototype screens displayed in empirical study. List of recommended hotels (left) and hotel details of the 3rd hotel of the list (right), depicting an explanation of the aggregation type with a high level of justification. Location screenshot Map data ©2020

the values reported for the items of this scale. Figure 3a depicts the different scores distributions.

Evaluation scores. We calculated evaluation scores for every variable of interest (explanation quality and the explanations aims: transparency, effectiveness, efficiency, and trust), as the average of the individual values reported for the items corresponding to each variable. Table 2 show the descriptive evaluation results by type and level, respectively.

Analysis of covariance Since our dependent variables are correlated (see Table 2), we performed a MANCOVA analysis to evaluate the simultaneous effect of type of explanations and level of justification on all variables that represent user's perception, and to what extent the individual decision-making styles or social awareness might influence such perception. Here, evaluation scores were used as the dependent variables, *level* and *type* as fixed factors (independent variables), and user characteristics scores as covariates. Smaller ANCOVA analyses were also performed, to test the interactions between independent variables and covariates, and their effect on each of the dependent variables. The results are summarized below.

Multivariate effects:

Significant multivariate effects were found for the variables: type $F(5, 140) = 4.68$, $p < .001$ and social awareness $F(5, 139) = 2.41$, $p < .05$. No significant overall effects were found for the level of justification, nor for the rational or intuitive decision-making style.

Univariate effects:

We performed a set of 5 ANCOVA analyses, to test interaction and main effects of the variables that reported a significant overall effect (type and social awareness) on each of the 5 dependent variables (explanation aims). Tests were conducted using Bonferroni adjusted alpha levels of .01 per test (.05/5)

Explanation quality: The type of explanation influences significantly the perception of explanation quality, $F(2, 146) = 5.37$, $p<.01$. A post-hoc test using Tukey HSD reveals a significant difference between aggregation and summary conditions ($p < .01$), such that the average explanation quality was significantly higher for aggregation (M = 3.98, SD = 0.65) than for summary (M = 3.56, SD = 0.75). No significant interaction was found between social awareness and type after the Bonferroni correction, $F(2, 146) = 5.37$, $p=.019$; however, we observed that the relationship between social awareness and explanation quality has a positive tendency for the aggregation and summary types, (aggregation having a steeper slope), whereas for review the relationship tends to be negative (Figure 3b).

Transparency: We observed that the type of explanation influences significantly the perception of transparency ("the system explains why the items were recommended"), $F(2, 146) = 5.49$, $p < .01$. A post-hoc test using Tukey HSD reveals a significant difference between aggregation and review conditions ($p < .05$), such that the average perception of transparency was significantly higher for aggregation (M =4.05, SD =0.69) than for review (M = 3.68, SD = 0.63). However, no significant influence of type was found in relation to whether users actually understood why the system recommended the items. There was also no significant interaction between type and social awareness, although a significant effect of social awareness on transparency was found, $F(1, 146) = 7.15$, $p < .01$. Here we observed a positive trend in the relationship between social awareness and transparency, as depicted in figure 3c.

Effectiveness: No main effects of type were found, neither significant interaction between social awareness and type.

Efficiency: No main effects of type were found, neither significant interaction between social awareness and type.

Trust: A significant effect of social awareness on trust was found, $F(1, 146) = 11.92$, $p<0.001$. Here we observed a positive trend in the relationship between social awareness and trust, as depicted in figure 3c. We found no major effects of type, nor significant interaction between type and social awareness.

6 DISCUSSION

We observed that the type of explanation seems to significantly influence the quality perception of explanations. Explanations that include an aggregated view with percentages of positive and negative opinions are perceived as more satisfying over explanations that only provide a mere summary of opinions, which confirms our hypothesis H2. This suggests that percentages may serve as easy anchors to convey more compelling information, while summaries may be perceived as too imprecise to convince. In fact, judgments and decision making can be influenced by changes in attitude, which in turn can result from the effortless use of cues such as numerical anchors, when people lack motivation or ability [24, 30]. In addition, although the difference in perception of quality between explanations with summaries and helpful reviews is not significant to confirm our H3 hypothesis, there seems to be a tendency to prefer reviews over summaries. This may reflect that some people trust a single opinion more than summaries that may hide details of special interest to them. Furthermore, there is not enough evidence to confirm our H1 hypothesis that users would prefer a higher level of justification in explanations, nor that reporting additional details of fine-grained aspects may influence the general perception of the recommender system. On the other hand, and contrary to our H4 hypothesis, we found no influence of rationality on this perception. First, it is difficult to make assumptions with respect to this variable since our sample is very skewed: to the right for the rational style and to the left for the intuitive, as depicted in Figure 3a. Additionally, this may be related to our observation that rationality and intuition are not diametrically opposed constructs: although most participants consider themselves to be someone who thoroughly evaluate available information, many of them also have a tendency to use their intuition when making decisions. In this regard, [14] have indicated that people with a greater tendency to process information in a rational manner (i.e. a prevalent rational cognitive style, according to [22]) are less likely to be intuitive decision-makers, whereas subjects with a greater tendency to process information in a more intuitive manner (i.e. a predominantly intuitive cognitive style) may be either rational or irrational decision-makers [14]. On the other hand, and although the interaction between social awareness and type of explanation is not statistically significant, we observed a tendency to prefer aggregated explanations for subjects with higher social awareness scores. A similar effect is observed for reviews, although smaller; here, summaries may sound too detached from actual opinions people express, therefore, the effect is negative for more social aware people.

In terms of transparency, the results suggest that, even when some types of explanations seem to serve better than others to

Table 1: Mean values and standard deviations of perception on explanation aims, per *level of justification* and *type of explanation* (n=152); values reported with a 5-Likert scale; high values of means represent a positive perception of recommender and explanations. Pearson correlation matrix, p<0.001 for all correlation coefficients.

Level:	Low		High		*Type:* Aggregation		Summary		Review		*Corr:* Variable			
Variable	M	SD	M	SD	M	SD	M	SD	M	SD	1	2	3	4
1. Explanation Quality	3.79	0.70	3.83	0.73	3.98	0.65	3.56	0.75	3.88	0.68				
2. Transparency	3.93	0.69	3.92	0.69	4.05	0.69	3.99	0.69	3.68	0.63	0.41			
3. Effectiveness	3.87	0.72	3.85	0.70	3.95	0.70	3.69	0.73	3.93	0.68	0.82	0.48		
4. Efficiency	3.96	0.78	3.90	0.79	4.07	0.65	3.70	0.93	4.00	0.71	0.58	0.45	0.70	
5. Trust	3.84	0.58	3.71	0.68	3.85	0.66	3.65	0.68	3.80	0.57	0.75	0.50	0.80	0.73

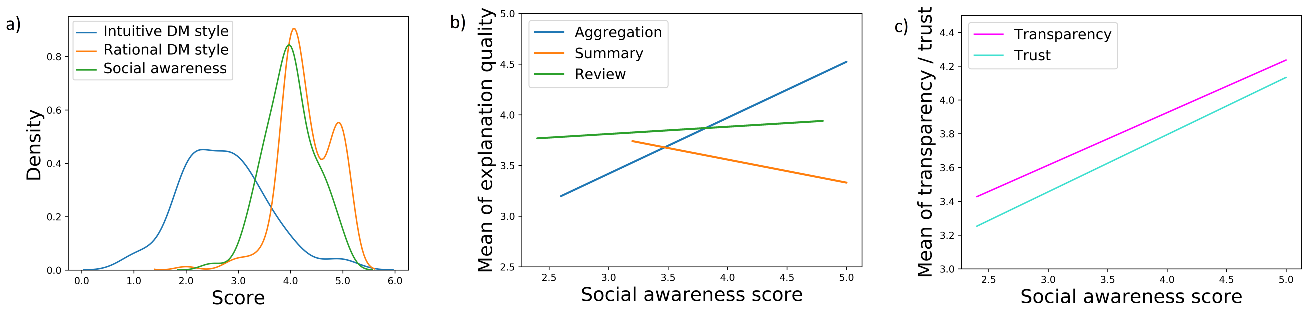

Figure 3: a) Kernel density estimate of user characteristics scores: rational and intuitive decision making styles and social awareness. b) Interaction plot for explanation quality (fitted means of individual scores) between type and social awareness. c) Effect of social awareness on transparency and trust (fitted means of individual scores). All scores on a 5-Likert scale.

explain the recommended items (in particular aggregations are perceived as more transparent than explanations based on helpful reviews), the users' understanding of the reasons behind the recommendations is not statistically different between types, i.e. a possible dichotomy between "the system explains why" and "I understood why". In this regard, some users mentioned, for example, that despite explanations were good, more details about how the algorithm actually works could further improve their understanding of reasons behind recommendations.

Finally, our results suggest that social awareness may play a role in the perception of both transparency and trust by users, that is, people with a higher disposition to listen and take into account others' opinions, tend to perceive the system as more transparent and to trust more in the recommender when the proposed explanations are provided, independent of their type or justification level.

7 CONCLUSION

In this paper, we have proposed the design of argumentative textual explanations, as well as examined and discussed the differences between types of explanations and levels of justification, their influence on users' perception of different characteristics of the system, and the influence that individual differences (namely decision-making style and social awareness) may have on such perception. We conclude that providing arguments based on aggregated results seems to be a meaningful way of presenting explanations. We cannot state though whether high or low levels of justification are *per se* better, or that differences between users' decision-making style influence significantly the perception of the proposed explanations. However, when taking into account another user characteristic, i.e. social awareness, differences in perception between users can be better understood, which can lead to better explanation designs and interaction possibilities. We believe that our findings lead to practical implications, e.g. that effective explanations should provide an initial aggregated overview of the main findings, and then allow the user to examine them in as much detail as preferred (e.g. by reading a list of the most useful reviews).

It is important, however, to recognize the limitations that the implementation of the proposed approach may have. For example, template-based explanations may be perceived as too repetitive for users, while implementations based on natural language generation (NLG) may be better received as seemingly more flexible. Therefore, as future work, we plan to extend our approach to the generation of explanations that are not template-based, leveraging NLG techniques, but still reflecting an argumentative structure. In addition, our evaluation has limitations, such as the use of a prototype instead of a system with real recommendations, as well as the use of Amazon Mechanical Turk, where despite our quality control implemented, it is difficult to encourage users to genuinely make a decision, which could guarantee higher quality in the execution of the task. Therefore, an evaluation on a real set and using a more effective motivation strategy will be part of the future work.

ACKNOWLEDGMENTS

This work was funded by the German Research Foundation (DFG) under grant No. GRK 2167, Research Training Group "User-Centred Social Media".

REFERENCES

[1] Roland Bader, Wolfgang Woerndl, Andreas Karitnig, and Gerhard Leitner. 2012. Designing an explanation interface for proactive recommendations in automotive scenarios. In *Proceedings of the 19th International Conference on User Modeling, Adaptation, and Personalization (UMAP'11)*. 92–104.

[2] Jacob A. Burack, Tara Flanagan, Terry Peled, Hazel M. Sutton, Catherine Zygmuntowicz, and Jody T. Manly. 2006. Social Perspective-Taking Skills in Maltreated Children and Adolescents. *Developmental Psychology* 42, 2 (2006), 207–217.

[3] Giuseppe Carenini, Jackie Chi Kit Cheung, and Adam Pauls. 2013. Multi document summarization of evaluative text. In *Computational Intelligence*, Vol. 29. 545–574.

[4] Giuseppe Carenini and Johanna D. Moore. 2006. Generating and evaluating evaluative arguments. In *Artif. Intell.*, Vol. 170. 925–952.

[5] Michael Chandler. 1973. Egocentrism and Antisocial Behavior: The Assessment and Training of Social Perspective-Taking Skills. *Developmental Psychology* 9, 3 (1973), 326–332.

[6] Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web. International World Wide Web Conferences Steering Committee*. 1583–1592.

[7] Hanxiong Chen, Xu Chen, Shaoyun Shi, and Yongfeng Zhang. 2019. Generate Natural Language Explanations for Recommendation. In *Proceedings of SIGIR 2019 Workshop on ExplainAble Recommendation and Search (EARS19)*.

[8] Felipe Costa, Sixun Ouyang, Peter Dolog, and Aonghus Lawlor. 2018. Automatic Generation of Natural Language Explanations. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*. 57:1–57:2.

[9] Michael J. Driver, Kenneth E. Brousseau, and Phil L. Hunsaker. 1990. The dynamic decision maker. (1990).

[10] Collaborative for Academic Social and Emotional Learning. 2013. 2013 CASEL guide: Effective social and emotional learning programs - Preschool and elementary school edition. (2013).

[11] Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng, and Bita Nejat. 2014. Abstractive Summarization of Product Reviews Using Discourse Structure. In *Empirical Methods in Natural Language Processing*, Vol. 53. 1602–1613.

[12] Anindya Ghose and Panagiotis G. Ipeirotis. 2011. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *Journal of Revenue and Pricing Management* 23 (2011), 1498–1512.

[13] Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. In *Computational Linguistics 43*, Vol. 1. 125–179.

[14] Katherine Hamilton, Shin-I Shih, and Susan Mohammed. 2016. The Development and Validation of the Rational and Intuitive Decision Styles Scale. *Journal of Personality Assessment* 98, 5 (2016), 523–535.

[15] Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. 2015. Trirank: Review aware explainable recommendation by modeling aspects. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1661–1670.

[16] Ya-Han Hu, Yen-Liang Chen, and Hui-Ling Chou. 2017. Opinion mining from online hotel reviews: A text summarization approach. In *Information Processing and Management*, Vol. 53. 436–449.

[17] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the User Experience of Recommender Systems. In *User Modeling and User-Adapted Interaction*. 441–504.

[18] Béatrice Lamche, Ugur Adigüzel, and Wolfgang Wörndl. 2012. Interactive explanations in mobile shopping recommender systems. In *Proceedings of the 4th International Workshop on Personalization Approaches in Learning Environments (PALE'14), held in conjunction with the 22nd International Conference on User Modeling, Adaptation, and Personalization (UMAP'14)*. 92–104.

[19] D. Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. In *Information Systems Research*, Vol. 13.

[20] Susan M. Mudambi and David Schuff. 2010. What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly* (2010), 185–200.

[21] Khalil Ibrahim Muhammad, Aonghus Lawlor, and Barry Smyth. 2016. A Live-User Study of Opinionated Explanations for Recommender Systems. In *Intelligent User Interfaces (IUI 16)*, Vol. 2. 256–260.

[22] Rosemary Pacini and Seymour Epstein. 1999. The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology* 76 (1999), 972–987.

[23] Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. In *Foundations and Trends in Information Retrieval*, Vol. 2. 1–135.

[24] Richard E. Petty and John T. Cacioppo. 1986. *Communication and persuasion: Central and peripheral routes to attitude change*. Springer-Verlag, New York.

[25] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems - RecSys 11*. 157–164.

[26] Janet A. Sniezek and Timothy Buckley. 1995. Cueing and Cognitive Conflict in Judge Advisor Decision Making. *Organizational Behavior and Human Decision Processes* 62, 2 (1995), 159–174.

[27] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction* 22 (2012), 399–439.

[28] Stephen E. Toulmin. 1958. The Uses of Argument. (1958).

[29] Jesse Vig, Shilad Sen, and John Riedl. 2009. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th international conference on Intelligent User Interfaces*. ACM, 47–56.

[30] Duane T. Wegener, Richard E. Petty, Kevin L. Blankenship, and Brian Detweiler-Bedell. 2010. Elaboration and numerical anchoring: Implications of attitude theories for consumer judgment and decision making. *Consumer Psychology* 20 (2010), 5–16.

[31] Bo Xiao and Izak Benbasat. 2007. ECommerce product recommendation agents: use, characteristics, and impact. *MIS Quarterly* 31, 1 (2007), 137–209.

[32] Ilan Yaniv and Maxim Milyavsky. 2007. Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes* 103 (2007), 104–120.

[33] Markus Zanker and Martin Schoberegger. 2014. An empirical study on the persuasiveness of fact-based explanations for recommender systems. In *Joint Workshop on Interfaces and Human Decision Making in Recommender Systems*. 33–36.

[34] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. 2014. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th international ACM SIGIR conference on Research and development in information retrieval*. 83–92.