Proceedings of the Eighth AAAI Conference on Human Computation and Crowdsourcing (HCOMP-20)

Understanding the Effects of Explanation Types and User Motivations on Recommender System Use

Qing Li,[1] Sharon Lynn Chu,[2] Nanjie Rao,[2] Mahsan Nourani[2]
[1]Santa Fe College, Gainesville, FL 32606, USA
[2]University of Florida, Gainesville, FL 32611, USA

Abstract

It is becoming increasingly common for intelligent systems, such as recommender systems, to provide explanations for their generated recommendations to the users. However, we still do not have a good understanding of what types of explanations work and what factors affect the effectiveness of different types of explanations. Our work focuses on explanations for movie recommender systems. This paper presents a mixed study where we hypothesize that the type of explanation, as well as user motivation for watching movies, will affect how users respond to recommendation system explanations. Our study compares three types of explanations: i) neighbor-ratings, ii) profile-based, and iii) event-based, as well as three types of user movie-watching motivations: i) hedonic (fun and relaxation), ii) eudaimonic (inspiration and meaningfulness), and iii) educational (learning new content). We discuss the implications of the study results for the design of explanations for movie recommender systems, and future novel research directions that the study results uncover.

Introduction

Nowadays, many popular movie recommender systems, such as Amazon Prime Video and Netflix, make an attempt to provide explains for their recommendations to the users. This paper presents a mixed-design online study that investigates how effective three different types of explanation interfaces are for movie recommender systems, in the context of various user motivations for watching movies.

Explanations provided with intelligent systems are typically answers to the 'why-questions' that the user may have about recommended items (Miller 2019; Mittelstadt, Russell, and Wachter 2019). As recommendation algorithms are often considered as 'black boxes', explanations provide users with some level of understanding of how the system works. Miller (Miller 2019) describes explanations formally as "ways of exchanging information about a phenomenon, in this case the functionality of a model or the rationale and criteria for a decision, to different stakeholders".

The provision of explanations serve a variety of purposes for both expert (e.g., developers) and non-expert users (e.g., consumers). For the former, they may help in debugging functions (Kulesza et al. 2015) or for legalistic purposes (Doshi-Velez et al. 2017). For the latter, beyond simply enabling users to learn about the system (Cleger, Fernández-Luna, and Huete 2014), explanation may allow users to better accept the system (Herlocker, Konstan, and Riedl 2000), predict future system outcomes, or regulate the behavior of the intelligent agent (Miller 2019). Particularly interesting to us, explanations can persuade novice users to take specific actions (Lombrozo 2006; Tintarev and Masthoff 2007), such as deciding to watch a movie that was recommended to them.

While there are many different computational approaches to the development of explanations for recommender systems and other AI systems, the problem of designing effective explanations is fundamentally a "human-agent interaction problem" (Miller 2019). Thus, it is crucial that we understand how to present explanations to the users from a human-centric approach. As a representational problem, explanation interfaces require designers to make decisions about both 'what needs to be explained' and 'how it should be explained' (Mohseni, Zarei, and Ragan 2018).

Prior research on explanation interfaces have looked at different types of explanation formats and visual designs, comparing methods such as tag clouds (Gedikli, Ge, and Jannach 2011; Vig, Sen, and Riedl 2009) and histograms, barcharts, and star ratings (Herlocker, Konstan, and Riedl 2000). Such work have typically viewed content and visualization of the explanation format to be tightly coupled (Gedikli, Jannach, and Ge 2014), and we do so as well in our work. Hence, different explanation formats emphasize different aspects of the system's recommendation. For example, neighbor ratings explanations present "the ratings of similar users to the target user for the recommended item" (Daher, Brun, and Boyer 2017) and emphasize others' preferences of the recommended movie. Keyword-based explanations show the keywords or tags that have been extracted directly from the recommended item and emphasize the content of the recommendation. Our work compares explanations showcasing neighbor ratings of movies to explanations showcasing user profile characteristics, and those showing user context characteristics, such as events in a user's life.

No explanation is able to present the entire universe of the system model or the entire causal chain taken by the system. In fact, it has been shown that "the cognitive burden of complete explanations is too great" (Miller 2019). In this paper, we propose that one way of constraining the explanations is to ensure that they are relevant to the user's real-world context. Contextually-relevant explanations are normal in human conversations, and are usually aligned with users' expectations, interests, and conversational purpose. (Mittelstadt, Russell, and Wachter 2019). For movie recommender systems, a key context feature is why the user wants to watch a movie and seeks to use the recommender system. However, little work in the literature has explored the effects of real-world context on users' response to recommender systems. Our work, thus, also aims to fill in this gap by factoring in user movie-watching motivations in our investigation of explanation types. We address common motivations for movie recommender systems, including people wanting to watch movies for fun and relaxation, to gain inspiration and meaningfulness, or to acquire new knowledge.

Following, we first present an overview of how explanation types have been distinguished and compared in prior work, and various kinds of user movie-watching motivations discussed in the literature, before describing our study.

Background and Related Work

Explanation Types

Explanations for recommender systems and other intelligent systems have been distinguished from many different perspectives in the literature.

Lipton (2018) delineates two main aims of explanations: first, *transparency* - to make transparent to a user how a system model works; and second, *post-hoc interpretability* - to extract information from the trained model to help users understand the model decisions. Our focus in this paper is on post-hoc interpretability, which provides information that can help practitioners and lay users make sense of the recommendations. Explanations with the goal of transparency tend to be more useful to expert users. Explanations for post-hoc interpretability can take a variety of forms. For recommender systems, examples of explanation types include neighbor ratings and similarity-based explanations.

In *neighbor ratings explanations*, profiles similar to the user's own profile information are found (i.e., neighbors) and the ratings to the recommended item given by these neighbors are shown (Herlocker, Konstan, and Riedl 2000).

Similarity-based explanations are grounded in the notion that in real life, people usually ask for recommendations from people they know or people who are like them (Guy et al. 2009). Such explanations thus portray similarities between the user's profile and profiles of similar users. For a comprehensive survey of explanation types for recommender systems, refer to Nunes and Jannach (2017).

Inputs for the explanations, including those mentioned above, are typically based on evaluations (ratings) by other users, characteristics of the users, or the content of the items. Few explanation types are based on the context of users. One example of this type of *context-based explanations* can be fashioned from the work of Chu et al. (2020), who presented a movie recommendation approach based on events that occur in the user's daily life. In their approach, movies are recommended based on how well they match with characteristics of significant events reported by the user. Explanations in this case can emphasize event characteristics used by the recommender system. Sato et al. (2018) present another context-style explanation that takes into account context in the sense of "when users would consume recommended items". One example of their explanations is "This restaurant is recommended to you because the restaurant is suitable for dates with your girlfriend/boyfriend".

In this paper, we address neighbor ratings, similarity-based, and context-based explanations. The first two are two contrasting types of commonly used explanations. The emphasis in neighbor ratings explanations is on others' direct evaluation of the recommended item, while similarity-based explanations emphasize aspects of user profiles that guided the system's recommendation process. And in this paper, we emphasize user profiles, so we call it profile-based explanation. Context-based explanations emphasize characteristics of the user's personal context, have been less studied in prior research, and their effects are not well understood. And in this paper, we use the user's personal event as their context, so we call it event-based explanation.

Similar to this study, various previous work have compared different types of explanations. For instance, Herlocker et al (2000) compared 21 explanation interfaces, each emphasizing different aspects (e.g., users' past performance, table of neighbor ratings, and recommended by movie critics.), for a movie recommender system. In their study, participants rated how likely it is for them to watch the recommended movie based on the explanation shown. The explanations with the best average ratings were *histogram with grouping*, *past performance*, and *neighbor ratings histogram*. Both *histogram with grouping* and *neighbor ratings histogram* were neighbor-ratings style explanations, but the former showed the star ratings clustered as '1 and 2', '3', and '4 and 5' while the latter was a standard barchart with one bar representing one star rating. The *past performance* explanation emphasized the recommendation performance of the system. The explanation took, for example, the following form "MovieLens has predicted correctly for you 80% of the time in the past".

Gedikli et al.'s study (2014) compared ten explanation types (seven taken from Herlocker et al.'s (2000) study with three new ones - *pie chart*, *tag cloud*, and *personalized tag cloud*). They found that an explanation that simply shows the overall percentage of movies with ratings of 4+ ('rated4+') and an explanation showing the overall average star rating for the recommended movie (e.g., 8.2 stars/10) ('average') were the most efficient (measured by time taken to rate). The three explanation types that performed the best in terms of effectiveness were those showing tag clouds, the percentage confidence of the recommender system, and a table of the number of neighbors who gave each star rating level (*neighbors counts*). Further, the *neighbors counts* explanation was the highest in terms of persuasiveness. Transparency was the highest for explanations showing a personalized tag cloud ('pertagcloud'), a pie chart, and a histogram of neighbor ratings. The 'pertagcloud', 'average', and 'rated4+' explanations scored highest for user satisfaction.

Finally, Symeonidis et al. (2009) compared explanations emphasizing user profiles, those emphasizing item features, and those showing a combination of the two types of explanations. Their study results showed that the hybrid explanation type was perceived as the most accurate. We did not find any study that investigated the comparisons of explanation types that we address in our work.

Context and User Motivations

Explanations are said to be more effective in helping users make decisions when they match user's real-world context (Mittelstadt, Russell, and Wachter 2019). Yet, few studies investigate the effects of context on explanations. Some studies have shown that factors relevant to the context can affect users' response to explanations. For example, Millercamp et al. (2019) investigated the effects of users' personal characteristics, specifically their levels of need for cognition, on their perception of feature-based explanations for recommender systems in various domains. Their results showed that depending on the domain, the need for cognition moderates user perception of the explanations.

Tomsett et al. (2018) proposed a model for explanability that considers the role that a user plays in an overall intelligent system, arguing that different agents have different goals of use. For example, 'creators' ("agents that create the machine learning system") have different needs from 'decision-subjects' ("agents who are affected by decision(s) made by the executor(s)"). However, the researchers did not implement and evaluate their proposed model.

In this paper, we consider the users' motivation for engagement as a context factor of focus for recommendation system use. Motivation refers to the rationale behind the way people behave, think, or feel at a specific time (Dweck 2017). In the specific domain of movie recommender systems that we address, three motivation types can be distinguished: watching a movie for 1) *hedonic*, 2) *eudaimonic*, and 3) *educational* motivations.

Hedonic motivations entail users watching movies to seek pleasure, fun, or comfort. With *hedonic* experiences, users' goals are typically for passing time, entertainment, obtaining information, escape, relaxation, and status enhancement (Conway and Rubin 1991). In media studies, this is the most commonly studied type of motivation. *Eudaimonic* motivations conversely are less studied and are associated with watching movies for meaningfulness, appreciation, or to have thought-provoking responses (Oliver and Bartsch 2010). One might watch "farcical comedies, thrilling action movies, and romantic love stories" for their entertainment offerings, but some deeply gratifying forms of entertainment, e.g., "tragic drama, moving cinema, heartbreaking opera, or poignant novels and poems" are not "enjoyable" in the colloquial sense of the term (Oliver and Raney 2011). Last but not least, users may also watch movies seeking new knowledge or with a desire to be exposed to new content, i.e., with *educational* motivations.

When watching a movie, users may not be explicitly conscious or aware of their motivations, but their choices are still underlined by a certain motivation. Further, we do not claim that the three motivations are necessarily mutually exclusive, meaning one can enjoy meaningful entertainment while also learning some new knowledge, but media studies literature have documented ways to measure one's primary gratification factor (e.g., (Oliver and Raney 2011)).

To our knowledge, no prior study has factored in and investigated the effects of these different types of user motivations in the context of explanations for recommender systems.

Research Questions and Study Design

Given the gaps in the literature on explanation types and user motivations, we sought to answer the following questions: *Are there significant differences in terms of efficiency, effectiveness, persuasiveness, transparency and user satisfaction among...*

RQ1: *...neighbor-ratings, profile-based, and event-based explanations?*

RQ2: *...neighbor-ratings, profile-based, and event-based explanations, given differing user movie-watching motivations (hedonic, eudaimonic, educational)?*

To answer the above research questions, we conducted an online user study with 78 participants, with 50 males and 28 females, recruited through the Amazon Mechanical Turk platform in exchange for compensation. The study used a mixed design, with two independent variables: i) type of explanations as a within-subjects factor and ii) user motivations as a between-subjects factor.

'Explanation type' had three levels: *neighbor-ratings*, *profile-based*, and *event-based* explanation. 'User motivation' had three levels as well: *hedonic*, *eudaimonic*, and *educational*. This study design resulted in a participant engaging with all explanation types but only in the context of one user motivation. Each participant were recommended and rated five movies per each explanation type. Thus, a participant went through 15 movies in total. The motivation context was randomly assigned to participants and the order of presentation of the explanation types was randomized for each participant.

In total, we obtained 27 participants for the *educational* user motivation context, 25 for the *hedonic* motivation context, and 26 for *eudaimonic* motivation context.

System Description

We developed a web application, with which the participants interacted and which implemented all the necessary steps to conduct the study. We explain the different aspects of the application below.

Generating Movie Recommendations

Since the focus of this work was on explanations, we did not implement a full recommender system. The generation of movie recommendations was handled as follows:

For the neighbor-ratings explanation condition, we assembled a database of movies by selecting 20 movies from the IMDb ‘top 250’ movie list (sorted by ascending popularity) based on the star ratings distribution of the movies. Movies recommended to participants for the neighbor ratings explanation type condition were randomly selected by the system from that database. For the profile-based and event-based explanation types conditions, the system asked for user input on the following: i) sex (male/female), ii) occupation type (STEM or non-STEM), iii) hobby preference (arts & craft/sports), iv) description of one significant event that has happened in the user’s life recently (open text), and v) category in which the significant event can be classified into (e.g., Travel, Health and Fitness, Pets, Society, etc.) We note that all user input required was a forced choice among different options, except for the description of the significant life event. This was done to simplify the movie selection process by the system for the purposes of this study.

Our system used the IBM Watson natural language API to analyze the movie plots of 255 movies from the Movielens dataset (Harper and Konstan 2015), and generated the top 3 categories that each movie can be tagged with. For the event-based explanation recommendation, the system used the event category selected by the user for their significant life event and found a movie with a top 1 category that matches the closest with the user’s selection.

For example, if the user selected “Food” as the event category, the system will randomly select a movie previously tagged with ‘food’ as one of its top category. A similar method was used for the profile-based explanation recommendation. The system used the options selected by the user for sex, occupation type and hobby preference, and checked the top 3 categories of all the movies in the database for a closest match. For example, if a user chooses “sports” as hobby preference and “STEM” as occupation type, the system will choose a movie tagged with both ‘sports’ and ‘science’ in its top 3 categories.

Generating Recommendation Explanations

The explanations were dynamically generated based on the movies to be recommended to each specific participant. All the explanation types used a consistent barchart visual layout, but each presented different dimensions based on the inputs used to recommend the movies for each condition. For the neighbor ratings explanation type, the explanation generated displayed the percentage of other users who rated the recommended movie at each number of stars (1 & 2 stars, 3 stars, and 4 & 5 stars) based on information obtained from the IMDb movie database. Figure 1 shows an example screenshot of neighbor rating explanation from the system for this condition.

For the profile-based explanation type, the system generated explanations that showed the percentage of users who has similar profile dimensions as the specific participant based on their sex (female or male), occupation type (STEM or Non-STEM), and hobby preference (arts & craft or sports). Figure 2 shows an example of a profile-based explanation. Finally, for the event-based explanation type, the system explanations showed the percentage of match to the three most relevant categories that the system matched to the event type that the user selected prior. Figure 3 shows an event-based explanation.

Figure 1: Screenshot of study website with a neighbor-ratings explanation

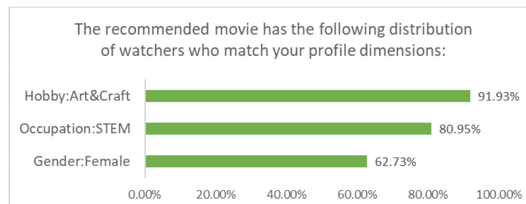

Figure 2: Example of profile-based explanation used

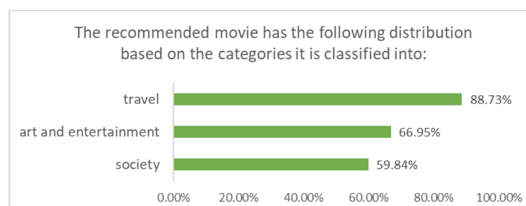

Figure 3: Example of event-based explanation used

Factoring in User Motivations

A key goal of our study was to see the effects of users adopting a specific type of motivational context for engaging with the recommender system. We primed participants to adopt specific motivational movie-watching context as follows: following guidelines from the literature as to what defines the *hedonic*, *eudaimonic* and *educational* user motivation (see Section 2), we created three scenarios, one reflecting each user motivation. The three scenarios are shown in Table 1. The scenario assigned to a participant was shown to them before they begin rating explanations and movie recommendations, and was persistently displayed during the rating sessions as well. An example of the scenario, as displayed to the participant during the rating sessions, can be seen at the top of Figure 1 enclosed in a blue box.

Motivation	Scenario
Hedonic	*It's a Saturday night. You want to watch a movie to relax and have fun.*
Eudaimonic	*Things are not going too well in your life right now, and you decide to watch a movie to hopefully lift your spirits a bit.*
Educational	*You are curious about something, and you want to watch a movie to learn about it.*

Table 1: Scenarios created for each user motivation

Study Protocol

Our study procedures were modeled after the study by Gedikli et al. (2014), who adapted their protocol from Bilgic and Mooney's work (2004). The steps of the study from a participant's perspective are summarized in Figure 4, and are described fully below:

1. Participants elect to participate in the Mechanical Turk (MTurk) HIT associated with our study task.
2. They read an information page describing the study. If they decided to proceed at this point, they were automatically considered to have given informed consent to participate in the study.
3. Participants who opted to continue the study were transferred from the MTurk platform to our website, and were shown a page describing our system, which was introduced as the MovieRec recommender system.
4. They were asked to provided their demographic information, one personal event, indicating as well the category type of the event, and fill in the entertainment motivations scale (see Section 6).
5. They were shown a scenario prompt (see Section 4).
6. *Study Phase I:* With the scenario prompt being persistently present on the interface, participants were shown a series of explanations for their recommended movies and rated them based on their willingness to watch the movie given the explanation. The order of explanations presented was randomly determined to minimize possible order effects.
7. *Study Phase II:* With the scenario prompt still being persistently present on the interface, participants were shown a series of movie recommendations to them. In this phase, movie details—for example, the movie title, the release year, genre, and plot description—were available and shown to the participants. Based on the movie details, participants re-rated their willingness to watch the recommended movies. Similar to explanations, the order of presentation of the movies was also randomly determined for this phase.
8. At the end of the study, the participants were presented with examples of each explanation types that they have seen before and filled a post-questionnaire about the explanation type. The explanation types were presented in a random order for each participant. They also completed an open-ended question, asking them about their opinions on each explanation type: *What do you think of this type of explanation for movie recommendations?*

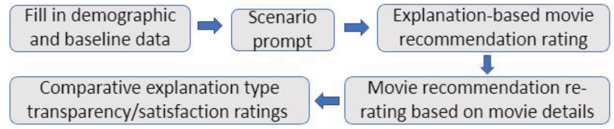

Figure 4: Summary of study procedures

Construct	Item Statement
Transparency	1. The explanation helped me understand what the system was trying to do. 2. The explanation makes the recommendation process clear to me.
Satisfaction	1. The explanation is good for a movie recommend system. 2. In general, I would rather use the same system with explanations rather than not having them. 3. The explanation is easy to understand.

Table 2: Items used to measure transparency and user satisfaction

Measures and Data Collection

The key dependent variables in our study were efficiency, effectiveness, and persuasiveness of explanations types.

To measure **efficiency**, following Gedikli et al. (2014), we tracked the amount of time taken by the participant to rate a recommended movie based on the given explanation. Efficiency can be defined as the extent to which the explanation "helps the user to decide more quickly or when it helps to reduce the cognitive effort required in the decision process" (Gedikli, Jannach, and Ge 2014).

Effectiveness was measured by calculating the difference between participant ratings for a recommended movie based *only on the given explanation* (that we will refer to as 'explanation-based ratings'), and participant ratings given for a recommended movie based on *holistic details*, such as movie title and plot description ('information-based ratings'). Participant ratings were done on a single 7-point likert scale item about the participant's willingness to watch the recommended movie: *How likely will you watch this movie?*. This approach to measuring explanation effectiveness again followed Gedikli et al.'s (2014) approach. They define effectiveness as "the ability of an explanation facility to help users make better decisions".

Persuasiveness can be defined as "the ability of an explanation type to convince the user to accept or disregard certain items" (Gedikli, Jannach, and Ge 2014). This construct was measured by the 7-point scale participant ratings for a recommended movie based only on the given explanation in terms of their willingness to watch the movie.

As secondary dependent variables of interest, we also as-

Study Phase	Data Collected
Pre-study questionnaire	Demographics; Inputs required for recommendations; Baseline movie-watching motivation
Explanation-based ratings	Single item rating willingness to watch recommended movie; Time needed to rate
Information-based ratings	Single item rating willingness to watch recommended movie
Post-study questionnaire	Transparency of explanation types; User satisfaction of explanation types; Open-ended comments on explanation types

Table 3: Summary of data collected in the study

sessed the perceived transparency and user satisfaction with the different explanation types. **Transparency** can be defined as the extent to which an explanation allows a user to understand how the system works. To measure transparency, we used two items from Pu et al.'s (2011) *ResQue* questionnaire. The two items required participants to rate on a 7-point likert scale from 'strongly agree' to 'strongly disagree', and are shown in Table 2.

User satisfaction of the explanations was adopted from items used by Chang et al. (2016) to assess satisfaction. Participants were asked to rate 3 items, shown in Table 2, on a 7-point likert scale of 'strongly agree' to 'strongly disagree'. Besides measures for the dependent variables and information collected to be used as input for the movie recommendation system, to understand the perspective that each participant is approaching the study with, we measured the participants' usual motivation for watching movies at the beginning of the study using the scale developed by Oliver and Raney (2011) to evaluate people's entertainment motivations. Table 3 provides a summary of all the data that was collected in our study.

Data Analysis

Quantitative Analysis

All items belonging to scales were averaged out to obtain a single score for each scale/construct. Single items were treated independently. Effectiveness scores were calculated by subtracting the participants' information-based ratings from their explanation-based ratings. The amount of time taken by participants to submit explanation-based ratings was extracted from system logs. The average baseline user motivation scores were calculated for participants in each condition. For participants assigned to the *hedonic* scenario, the average baseline *hedonic* motivation score was 5.41. For those in the *eudaimonic* scenario, the average baseline *eudaimonic* motivation score was 5.8. Furthermore, for participants in the *educational* scenario, the average baseline *educational* motivation score was 5.16. These relatively high baseline average scores indicate that participants should not have had a large cognitive dissonance when asked to engage in their assigned motivational scenarios.

Descriptive statistics were ran on all variables, and outliers were excluded. A mixed ANOVA was conducted with 'explanation type' as repeated measures factor and 'scenario' as between-subjects factor on time taken for rating movies based on explanations (efficiency), the effectiveness scores, participants' explanation-based ratings (persuasiveness), transparency scores, and user satisfaction scores. All tests were done at a 95% confidence level. Post-hoc pairwise comparisons using the Games-Howell's test were ran for mixed ANOVA results that turned out significant results.

Qualitative Analysis

The open-question asking for participants' opinions on the explanation types were coded following a qualitative coding process. First, irrelevant or inappropriate responses were excluded. Examples of such exclusions are when the participant provided the exact same answer for all explanation types, the response had nothing to do with explanations, or when they left a blank response. There were 53 valid responses that were considered for analysis. An *a priori* coding scheme was set up to code for 'affect towards the explanation type' and 'reasons for affect'. Codes for affect were *positive affect*, *negative affect*, and *neutral*. Codes for reasons for affect related to the dimensions that we addressed in the quantitative analysis: *efficiency*, *persuasiveness*, *transparency*, and *satisfaction*. All the valid responses were tagged with an 'affect' code and as many 'reasons for affect' codes as necessary. For instance, one response was: *"This recommendations system clearly mentions the star ratings for the movie, so easily decide how interesting [the] movie [is.]"*. This response was tagged with both the *transparency* and *efficiency* codes for 'reason for affect'.

The coding was done by two coders independently. After the independent rounds of coding, the two coders compared and discussed the responses and their codes. The intercoder agreement was 71.07% at the beginning of the discussion. An agreement level of at least 70% is typically deemed acceptable (Lombard, Snyder-Duch, and Bracken 2002). At the end of the discussion, the two coders resolved codings with disagreements and a final mutually-agreed coding for the responses was achieved. The frequencies of each code for each of the three explanation types were calculated. Examples of responses that tagged with each code in our coding scheme are shown in Table 4.

Results

Results are reported below in terms of the differences among explanation types and among scenarios (user motivations).

No significant main or interaction effects were found on the time taken to provide movie ratings based on the explanations provided (**efficiency**), the difference scores between information-based movie ratings and explanation-based movie ratings (**effectiveness**), and user satisfaction ratings (**satisfaction**). And no significant main effect of scenario and no interaction effects were found.

There was, however, a significant main effect of explanation type ($F(2, 128) = 4.34$, $p < 0.05$), and an interaction effect between explanation type and scenario ($F(4, 128) = 7.76$, $p < 0.001$) on the explanation-based movie ratings, which was used as a measure of **persuasiveness**. There was

Focus	Code	Example Response
Affect	Positive	*"I like this explanation since it gives a better understanding of what others watched the movie."*
	Negative	*"It's not detailed enough. It tells me almost nothing about the movie itself."*
	Neutral	*"This is a good system, but it doesn't reveal the plot of the movie, which is a flaw."*
Reason for Affect	Transparency	*"That is good recommendation process clear to me."*
	Efficiency	*"This system clearly mentions the movie ratings, so easily decide how interesting movie."*
	Persuasiveness	*" I feel I want to see the movie."*
	Satisfaction	*"I like this type of movie recommendation system."*

Table 4: Coding scheme used in qualitative analysis

no significant main effect for scenario. The pairwise comparison showed that only the neighbor ratings explanation type (M = 5.78) was significantly different ($p < 0.05$) from the profile-based explanation type (M = 5.30). Figure 5 illustrates the significant main and interaction effects.

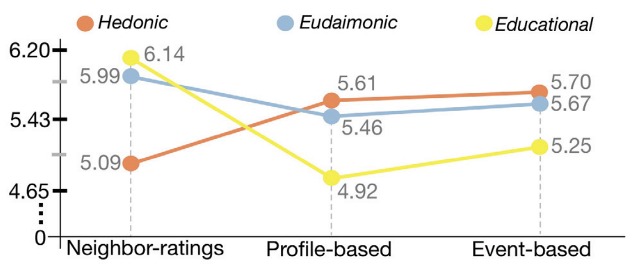

Figure 5: Main and interaction effects of persuasiveness

For **transparency** ratings, we found a significant main effect of explanation type ($F(2, 126) = 8.65$, $p < 0.001$). Pairwise comparisons showed that the neighbor ratings explanation (M = 5.11) was significantly different ($p < 0.05$) from both the profile-based explanation (M = 4.58), and the event-based explanation (M = 4.75).

Qualitative Results

Figure 6 shows the percentage distribution of the codes for 'affect towards explanation type' for each of the three types of explanations. The neighbor-rating explanation type had the largest proportion of *positive affect* code (62.26%), and the lowest proportion of *negative affect* code (13.21%). The distribution of *positive*, *negative*, and *neutral* affect codes for the profile-based and event-based explanations types were rather similar. This suggests that participants' affect towards these two explanation types were ambivalent.

Figures 7 and 8 show the percentage distribution of the codes for 'reason for affect' categorized under the *positive affect* and the *negative affect* code for each of the explanation type.

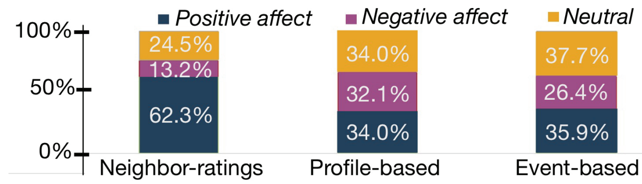

Figure 6: Distribution of affect codes

We note again here that in our analysis, one response could be coded with more than one 'reason for affect' code if necessary. Responses indicating positive affect towards three explanation types mainly mentioned reasons of transparency and satisfaction. For example, one participant wrote: *"I like it because it clearly shows the star rating percentage."* for the neighbor-ratings explanation.

Figure 7: Code distribution for reasons for positive affect

Figure 8: Code distribution for reasons for negative affect

Discussion

This paper presented a study comparing three types of explanations in the context of different movie-watching user motivations.

Our goal was to understand whether people respond differently to explanations that portray fundamentally distinct approaches to system recommendation, and whether their responses to the explanation types change given different motivations for use of the recommender system.

Study results showed that participants only found the neighbor-ratings, profile-based, and event-based explanation types to be different in terms of persuasiveness and transparency (RQ1). Gedikli et al. (2014) found differences in efficiency, effectiveness and satisfaction in their study among the explanations they compared, but their explanation types varied greatly in terms of visual representations (e.g., tag cloud vs pie chart vs table of neighbor ratings). The explanation types we compared all used a barchart-style design, but emphasized instead different input dimensions to the recommender system.

In term of persuasiveness, the main effect of explanation type suggests that irrespective of user motivations, the three explanation types are able to persuade users to watch the recommended movie at different levels. Explanations of type neighbor-ratings are significantly more persuasive than profile-based and event-based explanations. The persuasiveness of profile-based and event-based explanations seems to be similar. The results of Herlocker et al.'s (2000) study were that the 'histogram with grouping' and 'neighbor ratings histogram' explanations were the most persuasive. The 'histogram with grouping' explanation is equivalent to the neighbor-ratings explanation that we used in our study. Thus, our results validate that neighbor-ratings explanations showing grouped rating distributions are still the best-performing, even when compared with explanations that emphasize other types of recommendation inputs.

However, a significant interaction effect was also found for persuasiveness, indicating that the level of persuasiveness within each explanation type varies based on user motivations (RQ2). Users wanting to watch movies for hedonic and educational purposes find neighbor-ratings explanations to be more persuasive than users wanting to watch movies for eudaimonic reasons. But both profile-based and event-based explanations are more persuasive for users with eudaimonic and hedonic motivations than those with an educational motivation. These results are interesting because they provide evidence that user motivations cannot be ignored in the design of recommender system explanations.

As to possible reasons for the interaction effect, we posit that given that neighbor-ratings explanations emphasize what other users think of the recommended movie, it plays into the human psyche of groupthink. Groupthink refers to the "tendency for cohesive groups to become so concerned about group solidarity that they fail to critically and realistically evaluate their decisions and antecedent assumptions" (Park 1990), and can have a significant effect on one's decision-making. In the case of the neighbor-ratings explanation type, if users want to pass the time or relax (hedonic motivations) or to acquire new knowledge (educational motivation), they may decide to go with what others who have already engaged with the product seem to like. But if users are more interested to find a movie to watch to understand themselves and to find meaning (eudaimonic motivations), they may be more keen to engage with movies that are recommended because they match with aspects of the users' themselves (profile-based explanation) or experiences that they have had (event-based explanation).

In terms of transparency, a main effect of explanation type was seen, with the neighbor-ratings explanation type being perceived as more transparent than both the profile-based and event-based explanation types. In other words, explanations emphasizing that the recommender system uses others' ratings to recommend movies are easier to understand than explanations stating that the system uses the user's profile fields or life events. This result was further supported by the qualitative finding that the most prevalent reason for participants having positive affect for the neighbor-rating explanation was transparency. For example, one participant wrote: *"It's fairly clear how the recommendation is being made, since the 4/5 star ratings are so high of a percentage. It implies that this movie is almost universally liked."*.

Two possible reasons can be advanced for why the neighbor-ratings explanation type is more transparent than the others. First, variations of this explanation type are widely used in many recommender platforms, like Netflix, IMDb, and VUDU. Thus, users are familiar with the fundamental idea of this type of explanation. Conversely, although many recommender services such as Facebook uses profile-based parameters to display advertisements for example, explanations are not necessarily presented for these system outputs. As for the event-based explanation type, it seems to be, at least at this point of time, mostly in the research space rather than used in practice. A second reason for the transparency results could be that the dimensions listed in the profile-based and event-based explanation types were not specific or explicit enough. From our qualitative findings, some participants mentioned that the categories shown in these two explanations were too broad. (e.g., *"It's not really clear to me what is causing the recommendation to be made. For one thing, the categories are fairly broadly defined, so they don't really say much about what the movie might be about."*). Our system only extracted the highest-level categories from IBM Watson's classification analysis, such as society, health, and food. Extracting and presenting deeper or more precise categories in the explanations may give users a finer-grained understanding of the system.

Conclusion and Limitations

This paper has made several contributions: first, we showed that users respond differently in terms of persuasiveness and transparency according to what the explanations emphasize in a recommender system.

And second, we presented and provided evidence that user motivations for using recommender systems affect how users respond to system explanations, and thus need to be taken into account. Future research will need to explore how such motivations can best be integrated into recommender systems and their explanations.

Our research is certainly limited in depth and scope. First, we relied on participants being primed to think about a specific movie-watching motivation while engaging in the study. Considering motivations that participants intrinsically possess at the current moment of use may lead to different results. And second, we evaluated only specific forms of the explanation types. It is possible that the same explanation type presented in a visually different form (e.g., tag cloud instead of barchart) may result in different user responses.

Acknowledgment

Thanks to Dr. Eric Ragan, who initially offered some thoughts on the project.

References

Bilgic, M.; Mooney, R.; and Rich, E. 2004. Explanation for recommender systems: satisfaction vs. promotion. *Computer Sciences Austin, University of Texas. Undergraduate Honors* 27.

Chang, S.; Harper, F. M.; and Terveen, L. G. 2016. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, 175–182.

Chu, S. L.; Brown, S.; Park, H.; and Spornhauer, B. 2020. Towards personalized movie selection for wellness: Investigating event-inspired movies. *International Journal of Human–Computer Interaction* 0(0):1–13.

Cleger, S.; Fernández-Luna, J. M.; and Huete, J. F. 2014. Learning from explanations in recommender systems. *Information Sciences* 287:90–108.

Conway, J. C., and Rubin, A. M. 1991. Psychological predictors of television viewing motivation. *Communication Research* 18(4):443–463.

Daher, J.; Brun, A.; and Boyer, A. 2017. A review on explanations in recommender systems.

Doshi-Velez, F.; Kortz, M.; Budish, R.; Bavitz, C.; Gershman, S.; O'Brien, D.; Schieber, S.; Waldo, J.; Weinberger, D.; and Wood, A. 2017. Accountability of ai under the law: The role of explanation. *arXiv preprint arXiv:1711.01134*.

Dweck, C. S. 2017. From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological review* 124(6):689.

Gedikli, F.; Ge, M.; and Jannach, D. 2011. Understanding recommendations by reading the clouds. In *International Conference on Electronic Commerce and Web Technologies*, 196–208. Springer.

Gedikli, F.; Jannach, D.; and Ge, M. 2014. How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72(4):367–382.

Guy, I.; Zwerdling, N.; Carmel, D.; Ronen, I.; Uziel, E.; Yogev, S.; and Ofek-Koifman, S. 2009. Personalized recommendation of social software items based on social relations. In *Proceedings of the third ACM conference on Recommender systems*, 53–60.

Harper, F. M., and Konstan, J. A. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5(4):1–19.

Herlocker, J. L.; Konstan, J. A.; and Riedl, J. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241–250.

Kulesza, T.; Burnett, M.; Wong, W.-K.; and Stumpf, S. 2015. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th international conference on intelligent user interfaces*, 126–137.

Lipton, Z. C. 2018. The mythos of model interpretability. *Queue* 16(3):31–57.

Lombard, M.; Snyder-Duch, J.; and Bracken, C. C. 2002. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research* 28(4):587–604.

Lombrozo, T. 2006. The structure and function of explanations. *Trends in cognitive sciences* 10(10):464–470.

Millecamp, M.; Htun, N. N.; Conati, C.; and Verbert, K. 2019. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 397–407.

Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267:1–38.

Mittelstadt, B.; Russell, C.; and Wachter, S. 2019. Explaining explanations in ai. In *Proceedings of the conference on fairness, accountability, and transparency*, 279–288.

Mohseni, S.; Zarei, N.; and Ragan, E. D. 2018. A survey of evaluation methods and measures for interpretable machine learning. *arXiv preprint arXiv:1811.11839*.

Nunes, I., and Jannach, D. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction* 27(3-5):393–444.

Oliver, M. B., and Bartsch, A. 2010. Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research* 36(1):53–81.

Oliver, M. B., and Raney, A. A. 2011. Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication* 61(5):984–1004.

Park, W.-W. 1990. A review of research on groupthink. *Journal of Behavioral Decision Making* 3(4):229–245.

Pu, P.; Chen, L.; and Hu, R. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, 157–164.

Sato, M.; Ahsan, B.; Nagatani, K.; Sonoda, T.; Zhang, Q.; and Ohkuma, T. 2018. Explaining recommendations using contexts. In *23rd International Conference on Intelligent User Interfaces*, 659–664.

Symeonidis, P.; Nanopoulos, A.; and Manolopoulos, Y. 2009. Moviexplain: a recommender system with explanations. In *Proceedings of the third ACM conference on Recommender systems*, 317–320.

Tintarev, N., and Masthoff, J. 2007. A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*, 801–810. IEEE.

Tomsett, R.; Braines, D.; Harborne, D.; Preece, A.; and Chakraborty, S. 2018. Interpretable to whom? a role-based model for analyzing interpretable machine learning systems. *arXiv preprint arXiv:1806.07552*.

Vig, J.; Sen, S.; and Riedl, J. 2009. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th international conference on Intelligent user interfaces*, 47–56.