Examining Context-Aware Explanations for Recommender Systems

MARISSA RADENSKY*, RAYMOND FOK*, and JOYCE ZHOU*

Recommender systems have become a common way for users to efficiently locate items of interest on the Web. Occasionally, a system-generated explanation is provided to offer transparency into why the system may have recommended a particular item. These explanations can serve to improve user trust, efficiency, or persuasiveness of the system recommendation, among other aims. Though the benefit of personalized recommendations has been well accepted, the effect of **personalized explanations** on these aims is less evident. In this paper, we study user preferences for various explanation types (*e.g.*, item-based, social-based) in a scientific paper recommendation system. In particular, we study how these preferences may vary when within a familiar or unfamiliar research area. From interviews with nine graduate students, we find that overall, social-based explanations were preferred by participants in both contexts, though preferences for certain styles varied greatly between participants based on factors such as research experience and goal of using the recommender system. We present observations on how user preferences for different explanations varied between familiar and unfamiliar contexts, affecting aspects such as their desire for explanation interactivity, trust in the underlying system, and view on the system's overall utility. Based on these observations, we outline a number of design guidelines for better development of personalized explanations within a paper recommendation system.

CCS Concepts: • **Information systems** → *Online analytical processing*; • **Human-centered computing** → **Empirical studies in interaction design**.

Additional Key Words and Phrases: Explainable artificial intelligence, explainable intelligent user interfaces, personalized explanations

ACM Reference Format:
Marissa Radensky, Raymond Fok, and Joyce Zhou. 2021. Examining Context-Aware Explanations for Recommender Systems. In *CHI '21: ACM CHI Conference on Human Factors in Computing Systems, May 08–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The use of explanations in recommender systems has been fairly well explored and established as beneficial to users' satisfaction in understanding why they receive a particular recommendation. However, there has been limited investigation into what users in different contexts want from different styles of explanations. While prior research in generating personalized explanations generally considers user characteristics such as personality traits and user history [1, 7], we focus our study on how contextual factors like the user's goal in using the system and the user's level of expertise in the research field they are investigating may affect what elements of an explanation the user finds most useful.

To investigate these factors, we interviewed nine students with varying level of experience in research to get their thoughts on a currently-existing paper search engine (Google Scholar) and five types of paper recommendation explanations (content-based, item-based, user-based, social-based, and metrics-based) when used to find research

*All authors contributed equally to this research.

Manuscript submitted to ACM

papers in both familiar and unfamiliar fields. We coded these nine interviews and collected 8 major themes that arose throughout these interviews concerning explanation design for familiar and unfamiliar fields. Within each of these themes, we have summarized a couple design guidelines to take into account for explanation design in future.

2 RELATED WORK

2.1 Styles of Personalized Explanations

Several personalized explanation styles, such as demographic-based and item average rating, have been used and studied in recommender systems [6–8]. We employ the breakdown of personalized and non-personalized explanation styles utilized by Kouki et al. [7]. They describe five styles: content-based, item-based, user-based, social-based, and popularity-based (non-personalized), which are shown in Figure 1.

2.2 Objectives of Personalized Explanations

Personalized explanations for recommender systems have tended to address the issue of matching the type of explanation to the type of user. In other words, the explanation is selected or created to fit the user's innate characteristics, such as their personality traits, their user history, and their beliefs about the system prior to using it [1, 3, 5, 7]. While user characteristics are slow to change if at all, user contexts are often dynamic. For example, in this paper, we investigate how the context of looking for recommendations under a familiar topic versus the context of looking for recommendations under an unfamiliar topic affect the kind of personalized explanation desired by users. Users of any recommendation system are likely to find themselves in both of these contexts at some point. Other work has examined and identified contextual differences in explanation preferences, particularly that novice users versus expert users of a given recommender system desire different explanation styles [5, 7, 10]. Our work diverges from these studies in that we are studying novices versus experts in the specific topic under which they are looking for recommendations rather than novices versus experts in the overarching domain of the recommender system.

2.3 Evaluation of Personalized Explanations

Personalized explanations for recommender systems have been evaluated in multifarious dimensions such as trust, efficiency, user satisfaction, effectiveness, persuasiveness, scrutability, and transparency [4, 9]. Although our study is exploratory in nature and therefore does not explicitly evaluate personalized explanation styles, we highlight patterns in the kinds of personalized explanations desired by participants in terms of perceived effectiveness in guiding users towards relevant recommendations. In our studies, we define effectiveness of an explanation along two dimensions: ease of use and usefulness.

3 STUDY DESIGN

We conducted a 1-hour semi-structured interview with 9 participants to explore and identify patterns in the personalized explanation styles desired by recommender system users in the context of recommendations under a familiar topic versus in the context of recommendations under an unfamiliar topic. We chose computer science research papers as the domain for our recommender system. Though much of our results and analysis apply largely to recommender systems in general, we also analyzed how some of our results were specific to the domain. In the following section, we describe our participant recruitment process, interview study protocol, and codebook creation process.

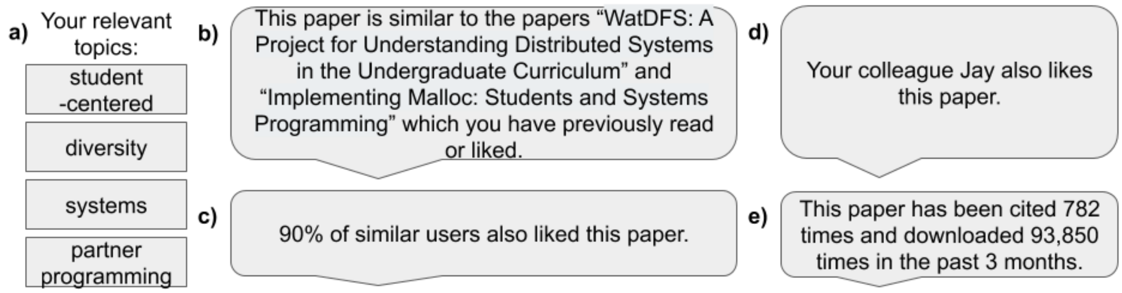

Fig. 1. The five explanation styles presented to participants, in this case for the topic of computer science education: a) content-based, which connects the recommendation to characteristics the user seems to like, b) item-based, which connects the recommendation to other recommendations the user seems to have liked, c) user-based, which connects the recommendation to other users similar to the user, d) social-based, which connects the recommendation to a someone the user knows, and e) popularity-based, which connects the recommendation to metrics indicating its general popularity.

3.1 Participants

Given the chosen domain, all of our participants were computer science students: 1 undergraduate (P1), 1 fifth-year BS/MS student (P2), 4 first-year PhD students (P3, P4, P5, P7), and 3 fifth-year PhD students (P6, P8, P9). All but the undergraduate had some computer science research experience.

3.2 Interview Protocol

The interviews were conducted as recorded video calls over Zoom. After a few preliminary questions gathering demographic information and information about the participants' experiences looking for research papers, each interview proceeded through two phases. The first was in the context of the participant's familiar topic, and the second was in the context of the participant's unfamiliar topic. Prior to the interviews, we asked each participant to provide a familiar topic as well as an unfamiliar topic in computer science in which they were interested. The participants' familiar topics spanned several computer science areas such as accessibility, web development, and database systems. Likewise, the participants' unfamiliar topics included diverse subjects such as security, embedded systems, and machine learning. For each phase, the participant was first asked to share their screen, navigate to the website Google Scholar, and enter a query in their (un)familiar topic. Occasionally, the topic used here was not the topic that the participant chose before the interview. Because we only needed the participant to see recommendations under some (un)familiar topic, we were not concerned about which (un)familiar topic they chose. While Google Scholar is a search engine and not a recommender system, we asked the participant to imagine that they were looking at a research paper recommender system feed. We then asked what questions the participant had about their feed. We also probed what the system could tell or show the participant about the recommendations in order to help them determine what results were relevant to them.

To further elicit thoughts and desires about what explanations users of recommender systems would prefer in the given context, we shared our screen with the participant and used Google Slides to present them with mock-ups of five recommendation explanation styles, which are presented and detailed in Figure 1. All but the popularity-based style are personalized. The content-based and item-based explanation styles contain information specific to the topic at hand, so

Table 1. Participants' favorite explanation styles in familiar and unfamiliar research contexts, based on their Likert scale ratings of explanation usefulness. Cells with multiple explanation styles indicate those participants had multiple favorite explanation styles in that context. In a familiar research context, most participants preferred social-based explanations. In an unfamiliar research context, item-based, user-based, and social-based explanations were similarly preferred.

Participant	Familiar	Unfamiliar
P1	metric/social	item
P2	content/social	user
P3	item/metric	item
P4	item/social	social
P5	social	social
P6	social	item
P7	content/social	item/social
P8	social	user/social/metric
P9	item/social	user/social

we created unique slides for each participant and each topic. For each explanation style, we provided a quick summary of how the explanation style worked and asked the participant if they had any questions about how it worked. We then asked them to rate the explanation style on a Likert scale for ease of use as well as usefulness for determining which recommendations were most relevant to them. Lastly, we asked the participant why they gave those ratings and what they might want to change about the explanation style. After viewing all the explanation styles, we asked the participant for any final thoughts about how the explanation styles compared to one another.

3.3 Creating a Codebook

To analyze the interviews, we downloaded and cleaned the Zoom transcripts for each of the interviews. The three authors then engaged in open coding of the interviews as described by [2]. More specifically, each author first took notes on 3 of the interviews in order to begin looking for themes in the interviews. The authors then discussed potential themes and narrowed them down into a codebook of 8 themes: *human trust, system trust, scope, effort of setup/maintenance, effort of use, interactivity, recommendation influence,* and *additional explanation uses*. These themes are defined and explored in the Results section of this paper. Each author then coded 3 of the interviews for these themes. Finally, the authors analyzed the quotes associated with each theme in order to determine specific findings related to each theme.

4 RESULTS

The following section presents a discussion of observations corresponding to each of the eight themes that arose from the interview study. The discussion also includes a number of design considerations that may be useful for future development of context-aware explanations.

4.1 Human Trust

Human trust refers to instances in which an explanation's reputability is affected by whether or not it is associated with a human known to the participant. This theme was found mostly in participant feedback to user and social-based explanations.

Social-based explanations concisely encode immense amounts of information via prior interpersonal relationships, allowing users to efficiently determine relevancy. Participants overall felt more inclined to trust papers when recommended by someone they knew. Figure 2 shows that participants rated social-based explanations as the easiest to use and the most useful for determining paper relevancy. The following quote from P2 gives some insight as to why social explanations were so highly rated amongst all participants:

"I'd be very much more inclined to read a paper that was recommended, like, if I knew the other person who recommended it. It's sort of a personal recommendation, I guess. And it would be straightforward."

Social explanations were seen as a reflection of in-person paper recommendations from other researchers, and as such, carried a sense of accountability that ensured the quality of the recommendation. The conciseness of the explanation was also appreciated, since with social explanations, users would only have to read the name of the recommending researcher. Furthermore, the mention of a known researcher was essentially–as one participant put it–a "dense encoding" of a number of features including the researcher's background, research history and reputation, and relationship with the user. Several participants noted that they would regard recommendations that had an explanation citing support from a reputable professor over recommendations from someone they were less familiar with or with less experience in that research area. They would also use these social recommendations as a method to efficiently triage papers of interest, and determining a reading order, since if they were familiar with the recommender, they could easily determine which person was doing research most related to their current thread of interest.

Though, social-based explanations may require further clarification of recommender context. For instance, some participants found it hard to understand what it meant for a user to "like" a paper. Some questioned, "How much do they like the paper? Did they like the paper because it was extremely relevant, or because they skimmed it and wanted to bookmark it for later, or because they kind of liked it?" Some participants also expressed concern over the potential difficulty of understanding the nuances of a colleague's preferences, and how they might align with their own preferences. Furthermore, if the colleague's research interests spanned multiple areas, some of the recommendations might not be directly relevant. Therefore, while participants generally liked the idea of social-based explanations, they were occasionally hesitant in blindly trusting these explanations, requiring further clarification of the recommenders' context.

Within a familiar research area, some participants felt that they occasionally wanted to be "rebellious" and not necessarily explore popular papers. Given their expertise, they were less interested in finding popular papers that provided an overview, and wanted papers that aligned closely with their specific research direction that they might have missed, which they felt would be difficult given the much smaller population of interested users. As a result, additional detail in the explanation describing what the similar users were specifically researching when they liked or cited that specific paper would be necessary to determine whether the paper was actually relevant to the expert user. On the other hand, within an unfamiliar research area, participants mentioned that they would be more inclined to trust other people's interests (and transitively, popular papers that were liked by many other users), since those papers may be a "good place to start" and present an overview of the field.

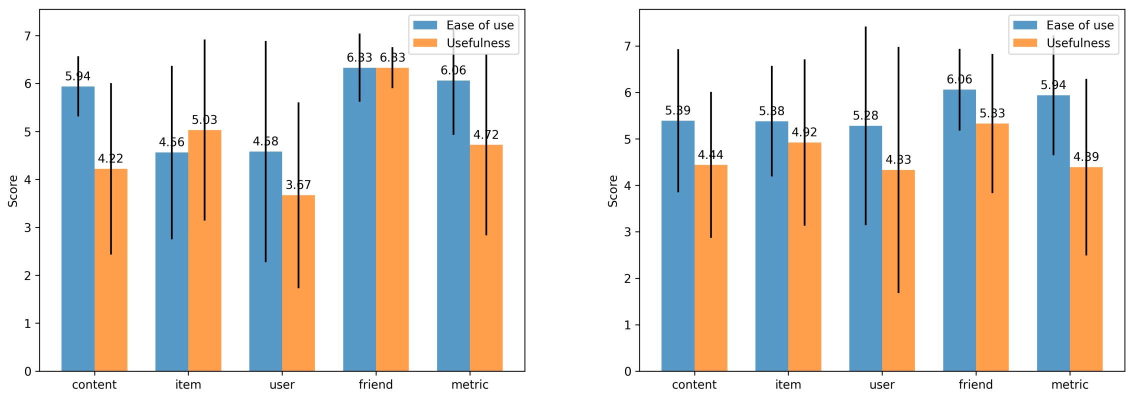

Fig. 2. Participant scores for explanation *ease of use* and *usefulness*, for each explanation style. Familiar research context is on the left, and unfamiliar is on the right. Explanation preferences overall varied greatly between participants, though within a familiar research context, participants rated social-based explanations highest across both dimensions. This effect was less pronounced within an unfamiliar research context.

4.2 System Trust

System trust refers to instances in which an explanation's reputability is affected by whether or not the participant understands what the system is doing. Many participants noted that explanations that revolved around "similarity" lacked sufficient transparency and overall negatively impacted system trust.

Participants were hesitant to trust the recommender system's determination of "similarity". Many participants expressed that the word "similar", while concise, raised issues around system opaqueness and user trust. For example, when shown item-based explanations, participants wanted the system to provide additional detail as to why a paper was deemed similar (*e.g.*, was it recommended because the authors were the same or related, or because the content of the paper overlapped with the previously read papers). With user-based explanations, participants similarly wanted more detail as to why a user was similar (e.g., was it an alignment of research interests, experience level, etc.). For some participants, the value of a system's recommendation based on general similarity was outweighed by the mental effort that they would have to apply to understand the system's recommendation and its explanation; for example, P3 raised the following issue: "*So I don't know how you would define similar users. And if I were looking at it, I don't know what that means. And I'm gonna have to figure out what you mean. And by the time I have gone through that it's going to be long enough that I could have just read the paper and decided on my own.*"

Some participants raised similar concerns within the user and social-based explanations with the word "like", specifically how it was simply not expressive enough to capture the gradient of user reactions to a paper. For instance, P4 asked, "*Does Dan super-like this paper or does he just like it because he wants to save it in his collection?*" Other participants mentioned that it was unlikely that current user feedback mechanisms provided by current paper recommendation systems (*e.g.*, starring a paper in Google Scholar) would be sufficient to pinpoint user opinion along this "liking" gradient.

Explanations relying on concepts with potential for ambiguity, such as "similarity", should include interactions to support user requests for further transparency of the system around that concept. Some participants mentioned that increased system transparency via detail would allow them to explicitly evaluate how the system was modeling them as a user, and use that mental model to guide future interactions with the system. For instance, during P7's interview, one of the topics shown within the content-based explanation was "avatar", which they felt was irrelevant to their current thread of exploration. They comment, "*I think a breakdown of what the system thinks like the relevant topics are will be useful. And in addition to that, add this percentage score so that I understand what the system thinks. If the system thinks the similar topic is "avatar" and it's 90%, then I know that I don't understand it, right?*" Here, P7 expresses how they might use additional detail (the model confidence score) to deduce that either they are misunderstanding the system, and therefore would need to adjust how they interact with the system in the future, or the system was incorrectly modeling their paper preferences. Our results suggest that within both familiar and unfamiliar research contexts, explanations that revolve around modeling a user's preferences need to provide sufficient transparency and detail to enable users to create a mental model of the recommender system, perhaps even more so with scientific papers than in other contexts such as video or shopping recommendations.

Within an unfamiliar research area, participants felt somewhat more comfortable trusting the system's determination of similarity. For instance, P6 noted: "*Being similar to other novice users in an unfamiliar research area is more useful/trustworthy than being similar to other experts in their familiar research area.*" Several other participants also echoed that they were more likely to trust a system recommendation based on opaque similarity to other users or paper when they were less knowledgeable of a particular area, compared to a research area that they were more familiar with and knew the multitude of nuances that could separate researchers within that single area.

4.3 Scope

Scope refers to instances where the participants' differing levels of familiarity with different research areas, general research literature, and their own paper exploration methodology affected how well they understood explanations with field-specific terminology.

Explanation text should match the expertise of users. Especially in unfamiliar research areas, participants found the topics listed in content-based explanations to be too specific for them to understand. More generally, participants were often entirely unsure what a term meant or how to orient themselves to a domain-specific paper or trend. P7 said that "*if the relevant topic is very short, and I'm not familiar with this area in general, then I don't know what it's talking about.*" P1 said that a content-based explanation in an unfamiliar field was especially challenging to use because "*now it requires me to start looking up terms that I don't recognize.*" This challenge isn't unique to domain terminology. Participants also found that it was harder to figure out what papers were about in an item-based explanation when looking at an unfamiliar area. For an item-based explanation, P9 described this challenge as not "*having those same points of reference, like 'Oh, I haven't yet found the papers I love.'*"

Participants also mentioned that they were less familiar with the significance of information such as authorship, conference name, or citation count when reading papers in fields they were unfamiliar with. This was even more true for participants who had less experience with the research world as a whole. While reading through the search results, P2 (who has relatively little research experience) said that "*I'm not too familiar with the research world, so I don't know exactly what it means to be cited.*" On the metrics-based explanation in an unfamiliar topic, P2 also said they "*don't know* *how many people are active researchers and actively downloading things*" so the exact citation or downloads count isn't helpful. P9 also commented that "*if I don't know this space, the authors don't help.*"

Explanations should be informative beyond a user's search query or what is already obvious from paper metadata. In both familiar and unfamiliar research areas, participants often observed that the system's content-based explanation topics were too broad to be helpful. P4 aptly summarized this up: "*If everything I see on the page is deep learning and if my search term is clearly deep learning that's not gonna help me at all.*" P8 found the content-based explanation as we demonstrated it unhelpful for familiar fields because "*I know what the related topic is and I already know some of the related work to that.*" However, whether a topic or detail is obvious based on paper metadata depends on how clearly authors write their abstracts or titles, as well as the user's own understanding of field norms. In an unfamiliar field, P4 said that "*I can't immediately tell what topic something is from its title or from its abstract, so this [content-based explanation] could be helpful for me since I don't have expertise.*" Thus, the specificity of terms that explanations use should adapt to how much detail the users are able to handle: too broad and users already have the information, too specific and users don't know what terminology means.

Participants also wanted a better understanding of how a term related to a research paper and what the significance of different terms was. It's not always enough to know that a paper mentions some concept: explanations also need to convey how the paper interacts with that concept. For the content-based explanation, P2 said that they're "*sort of vaguely familiar with some of these topics, but don't think I'm getting at it in the same way that they are.*" In an unfamiliar field, P9 said they might recognize specific terminology but "*don't quite know how they're used.*"

Explanations should support users' paper discovery methodology. Participants discussed how the level of detail that explanations had fit into the methods of paper discovery they already have learned, both for familiar and unfamiliar fields. In a familiar field, P8 said that they start with "*have something in mind that I really want to search along this direction*", already with specific topics and related works in mind. On the other hand, when looking for papers in unfamiliar fields, P9 starts off by "*trying to build up context" and figure out "what are the words that even define this space or what spaces are there to be defined.*" P9 wanted the content-based explanation to be helpful especially in unfamiliar fields "*to help me narrow down and start learning the vocabulary that I need to make those searches helpful.*" Likewise, P6 said that "*when I'm doing a broad search like this, I don't really know what I'm looking for.*" An explanation that gave broader context to a paper's relevance would support users in unfamiliar fields in the discovery process.

4.4 Effort of Setup/Maintenance

Effort of setup/maintenance refers to instances where participants were concerned about the expected amount of effort required from them for the system to generate helpful explanations. These observations were frequent when discussing both the familiar and unfamiliar research areas.

The system should be able to infer users' preferences instead of relying on their manual input. For both familiar and unfamiliar research areas, participants expressed concern about how much effort they might have to manually put into organizing the profile that the recommendation system has of them. Ideally, they expected the system to be able to infer their topics of interest. For content-based explanations, P6 said that the explanation would be great "*as long as I don't have to manually type in those topics and it can just, you know, infer them from my search history and* *whatnot in my research area.*" Likewise, P2 supposed that the 'relevant topics' would be "*based not just off of [the current] paper, but [also] previous viewing history.*" This sentiment was echoed for the item-based and social-based explanations as well. Participants didn't want to spend too much effort curating their own collection of 'liked' papers. P4 said that it was "*a lot of work*" to "*indicate to whatever system that I'm currently working on that I actually read the paper.*" P9 said that "*I imagine I have to provide ratings*" to be able to judge future papers as recommended or not. P4 also criticized a weakness of the social-based explanation, where it may be difficult for a system to find papers to recommend "*because people are lazy and they don't put papers in their previous 'read' or 'like' bins.*"

This is also linked to our observations about system trust: participants mentioned that they may click on a lot of papers to explore if they're relevant, but don't trust the system to be able to label those papers accurately (it may erroneously decide that a user is interested in a topic that they were just exploring in and didn't actually like). They would have to spend more effort correcting the system when it gets things wrong. There is a lot of nuance in what "liking" a paper means, which participants neither trust the system to get right nor want to spend the effort to keep organized themselves. P9 described how it was difficult even for a human to keep themselves organized when reading papers in a field they're familiar with: "*I have a hard time keeping track of things that I've liked, just because I've opened it and read it doesn't mean it's relevant to my work.*" P8 described how their exploration style for unfamiliar topics may confuse the system: "*I just randomly click into something, and those things may not be very good quality. So yeah, I don't want the system to recommend me similar stuff.*"

Explanations could take advantage of information that is already accessible to the system. Specifically in regards to social-based explanations, participants observed that the social network that the recommendation system uses could be integrated into other social networks or contacts lists. This saves users from having to input this information again and the system from having to learn it by itself. P7 said that "*if this is in Google and I use a lot of other Google stuff*", it makes it "*relatively easy to manage a friend list*". However, this approach might run into ethical or privacy issues. P9 noted that "*obviously the world knows enough data about [them]*" to be able to make accurate paper recommendations: there's enough tracking and social data latent from other recommender systems and across the Internet as a whole that we could take advantage of. As always, this is still a risk for invasion of privacy.

4.5 Effort of Use

Effort of use refers to instances where participants wanted explanations to be easy to use.

Explanations should be easy to read at a glance. In general, participants liked content-based explanations for both familiar and unfamiliar topics because they were succinct and easy to skim through. On the other hand, item-based explanations were much more verbose and took more time for participants to comprehend. P6 liked that the content-based explanation was "*really succinct, so that would help [them] skimming*". P9 simply stated that "*glancing at [the topics] is pretty easy*". As a complementary example, participants found the item-based explanation too long to be easy to use. P6 wanted the explanation to "*keep it as few words as possible... I don't need a verbose Clippy thing*". P9 echoed this thought, explaining that "*it's so much text and you have to slow down*". Longer explanations are more difficult to read and interrupt the flow of skimming through recommendations. On the whole, explanations should be glance-able and focused on a precise reason for relevance.

If an explanation requires users to recall outside information, it should be information that is easy to recall. Participants found that certain explanations were more or less easy to use because of how accessible the information they presented was. For instance, the item-based explanation lists paper titles, which participants said were difficult to use because it required them to recall what the paper was about. P9 said that a recurring "*challenge for [them] is remembering paper names*", and the item-based explanation requires them to "*remember what the papers are*". P1 summarized this issue as "*if I do read a lot of papers, then this would require me to go dig for what that paper was about*". On the flip side, participants said the social-based explanation was easy to use because it was easy for them to recall what different peoples' research preferences and interests were. P3 explained that "*the words 'your colleague' encode an immense amount of information*", since they already "*have a wealth of information about who they are, what kind of research they do*". Participants found social information easier to recall than paper titles.

Explanations should provide an advantage over pre-existing paper classification methods. This measure depends hugely on how familiar users are with the field in which they're looking for papers. Participants wanted explanations to be easier to use than the other methods they already employ to determine paper relevance, such as paper title, authorship, or skimming the abstract. In a familiar field, P9 that they "*don't know how much more information [the content-based explanation] gives me that [they] don't get just from the little excerpts, suggested related searches, and stuff*", while P4 said the topics "*seem to be somewhat more descriptive than the title which oftentimes doesn't have all the information*". However, the more unfamiliar with a research area participants is, the more likely they are to be receptive to assistance that explanations offer because they don't understand the norms of the field. P6 said they were "*much more open to get help from a recommendation system if it's not [their] field... because it's so hard to get to the right nuggets of information*".

4.6 Interactivity

Interactivity refers to instances in which the participant expressed interest in interacting with the explanation. Participants suggested several ways of interacting with our example explanation styles, which can be categorized into two main motivations: to understand the explanation itself better, and as a method of refining their search query or recommendation feed.

Participants wanted interactive explanations to offer detail about the terms they use and how they were generated - an "explanation for the explanation", as it were. This also ties back to the issue of system trust: participants want to know why the explanation says what it does. With content-based explanations, participants wondered what the topics listed meant and what made them relevant to the paper. P9 wanted the content-based explanation to "*[pull] up the sentence or sentence or two in the paper where it's used... [and understand] where these words came from*", similar to how Google Scholar currently displays context for search terms. When participants were unfamiliar with the topic, they want detail about what terminology means and why the explanation uses it. P2 said they wanted a better understanding of "*what it means to be a paper related to that topic*". P5 said "*if it could link to a Wikipedia page, that'd be great*".

Interactive explanations themselves can be a way of improving the recommendation system: refining search input, editing a user's recommendation profile, or as a method of paper discovery. P6 said that "*it would be more helpful if in this case I could add keywords*" to the content-based explanation, to edit the paper search directly. P7 echoed the same thought, asking for functionality that can "*refine the feed that [they] get out of doing another search*". P8 imagined interactivity that could both clarify what a topic consisted of and help them explore an unfamiliar area, where they can "*click into one topic and then show related papers in that topic*". Even a social-based explanation could be used for recommendation refinement and paper discovery. P8 noted that user profiles could be displayed publicly and the social-based explanation can provide "*easy navigation into which other paper in this area that this person also likes*".

4.7 Recommendation Influence

Recommendation influence refers to instances in which the participant expressed desires related to the explanation informing them of a recommendation's impact on other recommendations and beyond. Participants raised questions about if, when, where, why, and how a recommendation is influential in order to determine its relevance to them. However, only the questions of if and how spotlighted significant differences between the contexts of unfamiliar and familiar topics. Thus, we address only those questions.

Communicating if a recommendation is influential is important, especially for unfamiliar-topic recommendations, but insufficient as a recommendation explanation. In the familiar and unfamiliar topic context alike, multiple participants noted that an explanation indicating the popularity of the recommendation did not provide adequate information about the recommendation's relevance to the user. P3 described the popularity-based explanation as "*important as ancillary information, utterly terrible on its own.*" P2 placed even less weight on popularity, asserting that "*just because a whole bunch of people are reading [the recommendation] doesn't mean I want to read it.*" A couple participants indicated that they used popularity-based explanations to determine how to prioritize recommendations, but not to rule out recommendations as irrelevant. P5 commented, "*I'll read papers that have zero citations a lot as long as they're published in reputable conferences, you know. But I'll use the number of citations as a way to like order the way that I read them.*" Nonetheless, multiple participants including P5 acknowledged that when looking for recommendations specifically under an unfamiliar topic, participants often targeted the most popular recommendations, which would help them obtain a foundational understanding of that topic. P6 said of the popularity-based explanation, "*So it not only tells me it's cited, is it a seminal paper or is it a summary paper, which would be cited more highly typically, and then it gives me this information, like is it hot, is it like this area that a lot of people are looking at right now that's exciting or fresh in this field, and I'm not going to know the freshness or the excitement level of things because this is not my field. So this information would actually be helpful.*" Thus, our results provide evidence that popularity-based explanations can be helpful for recommendations, particularly for unfamiliar-topic recommendations, but personalized explanations should accompany them to provide more effective explanations.

Highlighting connections among familiar-topic recommendations is useful for personalized explanations about how familiar-topic recommendations are influential. When looking for papers under a familiar topic, multiple participants expressed the desire to find the most relevant papers that have built upon or influenced another paper. Presenting the most relevant papers to a given recommended paper can streamline the user's search for the best recommendations. While viewing the item-based explanation style, P8 noted, "*The recommendation system sometimes may be very helpful to kind of like give you a ranked list of like which is the most relevant [citation]... because yeah so usually like a paper has 40/50 citations, and it's impossible to go through all of them.*" Going through their Google Scholar results, P2 explained, "*A lot of stuff gets published, and I'd be wanting to know if there's any updates to this, especially since* *it came out in 2001.*" Under the familiar-topic context, our results suggest that the system's recommendation explanations should be personalized to include information about how recommendations are related to other recommendations the user has already liked.

4.8 Additional Explanation Uses

We asked participants about what they desired in explanations with respect to finding recommendations relevant to them. *Additional explanation uses* refers to instances in which the participant describes using the explanation for a purpose other than determining its relevance to the participant. We found evidence that providing social-based explanations allows users to familiarize themselves with their colleagues, but as this was the case whether or not the recommendation's topic was familiar, we will not address the point further.

Explaining how a familiar-topic recommended paper can help users to write their own papers adds usefulness to recommendation explanations. A couple participants mentioned using recommendation explanations to help them produce their own papers under a familiar topic. P5 stated that "*another useful thing*" about the content-based explanation "*is that I feel like it would help me structure my related work section.*" Meanwhile, P8 considered that "*how you see other people's views on [a] paper*" through the social-based explanation "*can also help you to shape your own paper.*" This result may not have been connected to the context of unfamiliar topics because users are unlikely to write a paper in an unfamiliar topic. The result indicates that explanations for familiar-topic recommendations may assist users in recognizing how the paper can influence their own work. Though this idea may seem unique to the domain of research paper feeds, it may also apply to other domains in which users can also be creators of the recommendations (e.g., music recommender system, recipe recommender system).

Designing explanations to encourage discovery and exploration benefits users when they are looking for unfamiliar-topic recommendations. When looking for papers under an unfamiliar topic, several participants desired to use explanations to discover and explore recommendations, rather than simply to find what recommendations were most relevant to them. Many participants expressed the desired to learn more about keywords and subtopics of an unfamiliar area. In reaction to the content-based explanation, some participants such as P9 mentioned that the explanation was "*helping me learn the language.*" P5 wanted a direct means to the definitions, stating, "*If [topics] could link to a Wikipedia page, that'd be great.*" Another manner in which multiple participants wanted to explore recommendations was through connections between topics with which they were familiar and topics with which they were unfamiliar. P8 remarked, "*If this is like an unfamiliar topic and I'm kind of like searching for related work... I may want to know like which paper, also in my area, cites those works as well. So that's kind of like… help me to build a bridge between the topics.*" These results suggest that explanations for unfamiliar-topic recommendations should provide easy access to topic definitions and should highlight connections between these recommendations and familiar topics.

5 CONCLUSION

We have presented observations from an exploratory study of explanation preferences within a paper recommender system, and highlighted several design considerations for designing better context-aware explanations. Our results suggest that user context such as expertise in a research area, may influence the types of explanations that would be most useful in determining recommendation relevancy. Future work may explore how these insights generalize to explanations of recommender systems in other domains, and analyze other dimensions of user context. Furthermore, we found that while preferences varied between participants, most participants rated social-based explanations as the most useful, since it enabled them to leverage information from prior interpersonal relationships such as the referenced colleague's research history and reputation in the decision-making process of relevancy.

ACKNOWLEDGMENTS

We thank James Fogarty and Jina Suh for their unwavering support and helpful feedback on all stages of this project. We also thank all of our interview participants for their time and participation in this project.

REFERENCES

[1] Shuo Chang, F. Maxwell Harper, and Loren Gilbert Terveen. 2016. Crowd-Based Personalized Natural Language Explanations for Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 175–182. https://doi.org/10.1145/2959100.2959153

[2] Juliet Corbin and Anselm Strauss. 1990. Grounded Theory Research: Procedures, Canons and Evaluative Criteria. *Zeitschrift für Soziologie* 19, 6 (1990), 418 – 427. https://doi.org/10.1515/zfsoz-1990-0602

[3] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. 2019. Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 275–285. https://doi.org/10.1145/3301275.3302310

[4] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How Should I Explain? A Comparison of Different Explanation Types for Recommender Systems. *Int. J. Hum.-Comput. Stud.* 72, 4 (April 2014), 367–382. https://doi.org/10.1016/j.ijhcs.2013.12.007

[5] Bart Knijnenburg, Niels Reijmer, and Martijn Willemsen. 2011. Each to his own: How different users call for different interaction methods in recommender systems. *RecSys'11 - Proceedings of the 5th ACM Conference on Recommender Systems*, 141–148. https://doi.org/10.1145/2043932.2043960

[6] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2017. User Preferences for Hybrid Explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (Como, Italy) *(RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 84–88. https://doi.org/10.1145/3109859.3109915

[7] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2019. Personalized Explanations for Hybrid Recommender Systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 379–390. https://doi.org/10.1145/3301275.3302306

[8] Martin Svrcek, Michal Kompan, and Mária Bieliková. 2019. Towards understandable personalized recommendations: Hybrid explanations. *Comput. Sci. Inf. Syst.* 16 (2019), 179–203. https://doi.org/10.2298/CSIS171217012S

[9] Nava Tintarev and Judith Masthoff. 2012. Evaluating the Effectiveness of Explanations for Recommender Systems. *User Modeling and User-Adapted Interaction* 22, 4–5 (Oct. 2012), 399–439. https://doi.org/10.1007/s11257-011-9117-5

[10] Gesa Wiegand, Matthias Schmidmaier, Thomas Weber, Yuanting Liu, and Heinrich Hussmann. 2019. I Drive - You Trust: Explaining Driving Behavior Of Autonomous Cars. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312817