Providing Explanations for Recommendations in Reciprocal Environments

Akiva Kleinerman
Bar-Ilan University
Ramat-Gan, Israel

Ariel Rosenfeld
Weizmann Institute of Science
Rehovot, Israel

Sarit Kraus
Bar-Ilan University
Ramat-Gan, Israel

ABSTRACT

Automated platforms which support users in finding a mutually beneficial match, such as online dating and job recruitment sites, are becoming increasingly popular. These platforms often include recommender systems that assist users in finding a suitable match. While recommender systems which provide *explanations* for their recommendations have shown many benefits, explanation methods have yet to be adapted and tested in recommending suitable matches. In this paper, we introduce and extensively evaluate the use of "reciprocal explanations" – explanations which provide reasoning as to why both parties are expected to benefit from the match. Through an extensive empirical evaluation, in both simulated and real-world dating platforms with 287 human participants, we find that when the acceptance of a recommendation involves a significant cost (e.g., monetary or emotional), reciprocal explanations outperform standard explanation methods, which consider the recommendation receiver alone. However, contrary to what one may expect, when the cost of accepting a recommendation is negligible, reciprocal explanations are shown to be less effective than the traditional explanation methods.

CCS CONCEPTS

• **Information systems → Recommender systems**;

KEYWORDS

Reciprocal Recommender Systems; Explanations; Online-dating Application

1 INTRODUCTION

Automated platforms for assisting people in finding a suitable match, such as online-dating and job recruitment web-services, are rapidly gaining popularity. However, finding a suitable match in these platforms can be a difficult and time-consuming task for users, especially since both sides of a potential match have to agree to form a match. Specifically, a user who seeks to find a desirable counterpart (e.g., a spouse or a partner) needs to account for both her own preferences as well as her potential counterpart's preferences in order to best utilize her time and effort. We refer to these platforms as *Reciprocal Environments* (REs). To assist users in finding a suitable match, REs often offer recommender systems, commonly known as *Reciprocal Recommender Systems* (RRSs) [19, 30].

Previous work on RRSs found that considering the preferences of *both sides* of a potential match, i.e., the recommendation receiver and the recommended user, is better suited for REs than the traditional approach which considers the recommendation receiver alone [19, 29, 30]. For example, say Alice and Bob are users in a online-dating platform. The traditional approach would generate Bob as a recommended match to Alice if it estimated that Alice would be interested in Bob. However, considering both Alice and Bob's preferences in order to generate a recommendation was shown to outperform this approach. In tandem, the question of how an RRS should *explain* its recommendations to the recommendation receiver arises. Specifically, while the traditional explanation methods which consider the preferences of the recommendation receiver alone have been demonstrated to increase the user's *acceptance rate* of the system's recommendations, the user's subjective *satisfaction* from the system and the user's *trust* in the system *for non-REs* (e.g., [2, 5, 12]), it remains unclear whether this approach is also suited for REs. To the best of our knowledge, previous work has not addressed this question in either simulation or the real world.

Continuing with our previous example, a traditional *explanation method* would explain to Alice why she would be interested in Bob (e.g., "He is tall and an artist"). However, additional information as to why Bob is expected to be interested in Alice (e.g., "He is likely to be interested in you because you are a doctor and like to hike") can be leveraged by an explanation method. To utilize this potentially useful information, in this paper, we introduce and extensively evaluate a novel explanation method based on the preferences of both the recommendation receiver and the recommended user, denoted *reciprocal explanations*.

We focus on the online-dating domain, which is perhaps today's most popular RE online[1]. Through three experimental setups, both in simulated and real world online-dating platforms, with 287 human participants, we show that the proposed reciprocal explanations approach can significantly outperform the traditional explanation method (i.e., which considers the recommendation receiver alone) *when a cost is associated with the acceptance of a recommendation*. Specifically, when accepting a recommendation is associated with a cost (e.g., time spent in sending a personalized message to the recommended user, the emotional cost of being rejected, etc), providing a reciprocal explanation brings about a higher acceptance rate and trust in the system. Interestingly, contrary to what one may expect, when the cost associated with accepting a

RecSys '18, October 2–7, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5901-6/18/10…$15.00
https://doi.org/10.1145/3240323.3240362

[1] According to a recent survey, 74% of single people in the United States between the ages of 18 and 65 have signed up with one of the various online-dating sites, *https://www.statisticbrain.com/online-dating-statistics/*.

recommendation is negligible (e.g., indicating interest in the recommended user by giving a "like", no strong emotional involvement, etc), we find that the traditional methods outperform the reciprocal explanations approach.

2 RELATED WORK AND BACKGROUND

Previous studies have designed and investigated different methods for generating *recommendations* in REs (e.g., [19, 29–31]). These studies have found that methods that contemplate the presumed preferences of both sides of the recommendation outperform methods that consider one side alone. In practice, many popular online-dating sites and other REs include recommender systems that take into account the preferences of both sides, such as the popular Match [2] and OkCupid[3] platforms. These and other RRSs often provide explanations for the generated recommendations.

Explainable Artificial Intelligence (XAI) is an emerging field which aims to make automated systems understandable to humans in order to enhance their effectiveness [8]. This field of research was highly prioritized in the recent National Artificial Intelligence Research and Development Strategic Plan [18, p. 28]. The need for explanations is also acknowledged by regulatory bodies. For example, the European Union passed a General Data Protection Regulation[4] in May 2016 including a "right to explanation", by which a user can ask for an explanation of an algorithmic decision made about him [6]. In recent years, providing an explanation has become a standard in many online platforms such as Google and Amazon.

A wide variety of methods for generating *explanations* for a given recommendation were proposed and evaluated in the literature. Two practices are commonly applied in this realm: First, existing explanation methods focus on the recommendation receiver alone. To the best of our knowledge, none of the existing methods were developed or deployed in an RE. One exception to the above is Guy et al. [9], who presented a recommender system for an RE which is transparent (i.e., provides accurate reasoning as to how the recommendation was generated). Unfortunately, the authors did not compare the effects of their approach with other explanation methods nor did they consider the unique characteristics of REs. Secondly, existing explanation methods are often tailored for specific applications or are heavily dependent on the underlying algorithm for generating the recommendation and therefore cannot be easily adapted or evaluated in different domains. In this work, we relieve these two practices by designing and extensively evaluating two novel general-purpose explanation methods for REs.

Many studies have demonstrated the potential benefits of providing explanations to automated recommendations. For example, Herlocker et al. [12] found that adding explanations to recommendations can significantly improve the *acceptance rate* of the provided recommendation and the *satisfaction* of the users thereof. Sinha et al. [26] further found that transparent recommendations can also increase the user's *trust* in the system. These results were replicated under various domains and explanation methods (e.g., [2, 5, 25]). The results of these works and others have combined to suggest two widely acknowledged guidelines for developing explanation methods: (1) Explanations which include *specific features* of the recommended item/user are highly effective, even if these features are not the actual reason the recommendation was generated [5, 12, 20]; and (2) It is important to limit the length of the explanation in order to avoid an information overload which can make explanations counterproductive [5, 20]. We follow these guidelines in our designed reciprocal explanation methods.

Recommendation Methods for Online-dating

In this work we focus on the domain of online-dating. An RRS in online-dating may provide a user x with a list of recommendations for suitable matches where each recommendation consists of a single user y. Note that unlike the original formulation of economical matching markets [4], an RRS in online-dating, as well as in many other REs, may recommend any user y to more or less than a single user x.

In this study we focus on generating explanations. As such, we use two state-of-the-art *recommendation* methods that have been developed and tested in online-dating: RECON and Two-sided collaborative filtering.

RECON [19] is an effective content-based algorithm which was empirically shown to be superior to baseline algorithms in online-dating sites. In the RECON algorithm, each user x in the system is defined by two components:

(1) A predefined list of personal attributes which the user fills out in her profile, denoted as follows:
$$A_x = \{v_a\}$$
where v_a is the user's associated value with attribute a.

(2) The preference of user x regarding every attribute a of potential counterparts, denoted $p_{x,a}$, which is represented by the user's message history in the environment:
$$p_{x,a} = \{(v_a, n) : n = \#\text{messages sent by } x \text{ to users with } v_a\}$$
That is, $p_{x,a}$ contains a list of pairs, each consisting of a possible (discretized) value for a and the number of messages sent by x to users characterized by v_a.

Example 2.1. Bob is a male user who has sent messages to 10 different female users. For simplicity, let us assume each user is only characterized by two attributes: smoking habits and body type. Bob sent messages to female users with smoking habits as follows: 1 smokes regularly, 3 smoke occasionally and 6 never smoke. Regarding their body type: 4 were slim, 4 average and 2 athletic. Bob's preferences would be presented as follows:
$$p_{Bob,smoke} = \{(1, regularly), (3, occasionally), (6, never)\}$$
$$p_{Bob,body-type} = \{(4, slim), (4, average), (2, athletic)\}$$

The RECON algorithm derives the predicted preferences of each pair of users x and y using a heuristic function that reflects how much their respective preferences and attributes are aligned.

The second recommendation algorithm we use is the Two-sided collaborative filtering [30] which was found to outperform RECON. The algorithm uses a collaborative filtering approach where the similarity between users is derived from their message history. Namely, two users will be considered similar if a large portion of

[2] *http://www.match.com/help/faq/8/164*
[3] *http://www.okcupid.com*
[4] http://ec.europa.eu/justice/data-protection/

their messages were sent to the same recipients. Given a recommendation receiver x and a potential recommended user y, the method first calculates the presumed interest of x in user y by measuring the similarity of x to users who sent messages to y. Later, the interest of y in x is calculated symmetrically. Finally both measures are aggregated into a single measure, which models the mutual interest of the match.

3 GENERATING RECIPROCAL EXPLANATIONS

Let us assume an RRS has decided to recommend user y to user x based on one of the algorithms discussed above. The recommendation may be provided with or without an accompanying explanation. If the explanation only addresses the potential interest of user x in user y (and not vice versa) we refer to it as a *one-sided explanation* and denote it as $e_{x,y}$. Similarly, if the explanation addresses the potential interest of user x in user y *and vice versa*, we refer to it as a *reciprocal explanation*. Naturally, a reciprocal explanation may be decomposed into a pair of one-sided explanations, $e_{x,y}$ and $e_{y,x}$.

The generic framework for providing recommendations with *reciprocal explanations* is provided in Algorithm 1.

Algorithm 1 Reciprocal Explanations

Require: User x, *GenerateRecommendations*: a Recommendation method, returns a list of recommended matches, *Explain*: an explanation method

1: $Output \leftarrow \emptyset$
2: $R \leftarrow GenerateRecommendations(x)$
3: **for all** $r \in R$ **do**
4: $\quad e_{x,r} \leftarrow Explain(x, r)$
5: $\quad e_{r,x} \leftarrow Explain(r, x)$
6: $\quad Output = Output \cup (r, e_{x,r}, e_{r,x})$
7: **return** $Output$

Providing a recommendation with a *one-sided explanation* is naturally derived from Algorithm 1 by omitting Row 5 and amending Row 6 accordingly.

To realize Algorithm 1, one needs to define both the recommendation method and the *Explain* method. Specifically, one would need to choose the underlying methods to be used in order to provide either a one-sided or reciprocal explanations.

4 EMPIRICAL INVESTIGATION

In order to evaluate and compare the one-sided and reciprocal explanations methods, we performed three experiments: two in a simulated online-dating environment developed specifically for this study and one in an operational online-dating platform. Each environment has its own benefits: Results from the operational online-dating platform naturally reflect the real-world impact of both explanation methods, whereas in the simulated environment one receives detailed and explicit feedback from the users, which otherwise would be impractical to gather in an active online-dating platform. We discuss these experiments below.

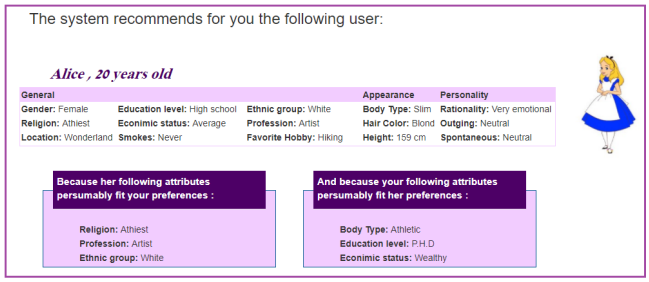

Figure 1: A recommendation with a reciprocal explanation in MM.

4.1 The MATCHMAKER Simulated Environment

We created a realistic simulated online-dating platform, which we call MATCHMAKER (MM for short). Using MM, users can view profiles of other users, interact with each other by sending messages and receive recommendations from the system for suitable matches. With the collaboration of experts in online-dating who did not co-author this paper, we designed MM's features to reflect those of popular online-dating platforms. Figure 1 presents a snapshot of a recommendation in the MM platform.

MM is a web-based platform and can be accessed at *www.biu-ai.com/Dating*.

In order to develop an RRS for MM, it is necessary to obtain the attributes and preferences of both the participants in the experiment and the potential recommended users. In order to create profiles in MM which would be as realistic as possible, we used the public attributes of profiles from real online-dating sites, such as *www.date4dos.co.il*. However, note that the data does not consist of the users' message history or preferences, hence the designed RRS would be very limited. To overcome this challenge we preformed the following data collection: We recruited 121 participants, 63 males and 58 females ranging in age between 18 and 35 (average 23.3), all of whom are self-reportedly single and heterosexual. All participants were university students recruited by posting ads in relevant classes. First, the participants entered MM and filled out a personal attributes questionnaire common in on-line dating platforms (e.g., age, occupation). Later, the participants viewed the profiles obtained from the real online-dating sites as discussed above and sent fictitious messages to the profiles that they perceived as suitable matches[5]. Participants were instructed to view at least thirty profiles and to send messages to at least ten relevant profiles in order to generate sufficient data for deriving their preferences. An average of 50.72 profiles (s.d.= 30.99) were viewed and 11.92 messages (s.d.=3.96) were sent by each participant. The data of three of the participants was removed due to their failure to comply with our instructions.

Following the above data collection procedure, we obtained 118 participant profiles and preferences. We anonymized the participants' profiles and preferences and used them as the initial profiles in MM for later investigation.

[5]Participants were aware that the profiles were simulated although based upon real data and that the messages were not actually sent to recipients. They were guided to send simulated messages to profiles they viewed as relevant matches for them.

4.2 Choosing the *Explain* Method

Before we turn our attention to the main point of this paper – the evaluation of one-sided and reciprocal explanations in REs – we performed a preliminary investigation in order to find the best suited *Explain* method for online-dating, the domain on which we focus throughout this paper.

For our investigation, we use an *Explain* method which returns a list of k attributes of a user which can presumably best explain why the recommendation is suitable. This approach was shown to be very effective in prior work [5, 28]. In order to avoid an information overload, we limited the number of attributes included in the explanation to three, as suggested in [21].

We investigate two *Explain* methods which correspond with the suggested format above: 1) *Transparent* (Algorithm 2); and 2) *Correlation-based* (Algorithm 3).

The transparent *Explain* method, which aims to reflect the actual reasoning for the recommendations provided by the RECON algorithm, works as follows: to explain to user x a recommendation of user y, the method returns the top-k attributes of y which are the most prominent among users who received a message from user x.

Algorithm 2 Transparent Explanation Method

Require: two users x and y, number of attributes for explanation k.

1: $temp \leftarrow \emptyset$
2: obtain P_x from user x
3: obtain A_y from user y
4: **for all** attributes a $\in A$ **do**
5: obtain the value v_a of attribute a in A_y
6: obtain $P_{x,a}$ from P_x.
7: find $(v, n) \in P_{x,a}$ s.t. $v = v_a$
8: $temp = temp \cup (v_a, n)$
9: **sort** $temp$ by the values n
10: $e_{x,y}$ = top-k attribute values of $temp$
11: **return** $e_{x,y}$

The correlation-based *Explain* method is inspired by the commonly used Correlation Feature Selection method from the field of Machine Learning [10]. In our context, we would like to measure the correlation between the presence of attribute value v_a in a user's profile and the likelihood that x will choose to send him/her a message. To that end, for each user x, we need to identify which users x has viewed in the past and whether he chose to send them a message. Also, we need to identify which of the viewed users is characterized by each attribute value v_a.

Formally, for each user x, we first identify the set of users $I = \{i\}$ that user x has viewed in the past, and define

$$M_x(i) = \begin{cases} 1, & x \text{ sent a message to } i \in I \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$S_{x,v_a}(i) = \begin{cases} 1, & \text{User } i \in I \text{ is characterized by } v_a \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Using M_x and S_{x,v_a} we define the correlation-based method described in Algorithm 3.

Algorithm 3 Correlation-based Explanation Method

Require: two users x and y , number of attributes for explanation k.

1: $temp \leftarrow \emptyset$
2: obtain M_x
3: **for all** attributes a $\in A$ **do**
4: obtain the value v_a of attribute a in A_y
5: obtain S_{x,v_a}
6: $w_{v_a} = PEARSON(M_x, S_{x,v_a})$
7: $temp = temp \cup (v_a, w_{v_a})$
8: **sort** $temp$ by the values w_{v_a}
9: $e_{x,y}$ = top-k attribute values of $temp$
10: **return** $e_{x,y}$

The PEARSON function, used in line 6 of Algorithm 3, is the well known Pearson correlation coefficient for measuring correlation between two variables [1]. Note that the correlation-based explanation method was specifically designed in order to provide more intuitive explanations. As such, our *working hypothesis* was that the correlation-based explanation method would have a greater positive effect than the transparent explanation.

To illustrate the difference between the explanation methods, we revisit Example 2.1. Assume an RRS has decided to recommend Alice, who never smokes and is slim, to Bob. Recall that Bob sent 6 messages to users who never smoke and 4 to slim users. For $k = 1$, the transparent explanation method would provide "never smoke" as an explanation because Bob sent more messages to users who never smoke than to users who are slim. Now say Bob viewed a total of 25 users, of whom 18 never smoke and 4 were slim. In other words, Bob sent messages to only a third of the users he viewed who never smoke, and to all users he viewed who are slim. Thus, the correlation-based method would find a stronger correlation between the presence of "slim body" and Bob's messaging behavior, hence "slim body" would be provided as an explanation.

In order to compare the two *Explain* methods, we used the MM simulated system discussed above. We asked 59 of the 118 participants who took part in the data collection phase (Section 4.1) to reenter the MM platform, where each participant then received a list of five personal recommendations generated by the RECON algorithm along with either transparent explanations (30 participants) or correlation-based explanations (29 participants). Participants were randomly assigned to one of the two conditions. Participants were asked to rate the *relevance* of each recommendation separately, on a five point Likert scale from 1 (extremely irrelevant) to 5 (extremely relevant). Next, participants answered a short questionnaire (available in the appendix in Section 7), debriefing them on their user experience. The questionnaire included questions which are commonly used for measuring four prominent factors in user experience: user *satisfaction* from the recommendations, perceived *competence* of the system, perceived *transparency* of the system, and *trust* in the system [2, 16, 22]. In addition, the users were asked specifically about the *explanation usefulness*, namely the extent to which the users considered the explanations to be helpful. All questions were answered on a five point Likert scale.

Note that we chose the RECON algorithm for the recommendations in *MM* since the collaborative filtering method described in [30] can only recommend users who have previously received messages. As described above, the recommended users in our experimental setup were created specifically for the recommendations, and were not viewed by any users prior to the recommendations.

Results. All collected data was found to be approximately normally distributed according to the Anderson-Darling normality test [23]. All reported results were compared using an unpaired t-test. The results show that participants in the correlation-based condition were significantly more satisfied than those in the transparent explanation condition (mean= 3.58, s.d.= 0.82 vs. mean= 3.14, s.d.= 0.65, $p \leq 0.02$). Similarly, the perceived transparency was reported to be significantly higher in the correlation-based condition (mean= 3.97, s.d.= 0.93 vs. mean= 3.41, s.d.= 0.65, $p \leq 0.04$), as was the perceived usefulness of the explanations (mean= 3.8, s.d.= 0.81 vs. mean= 3.17, s.d.= 0.8, $p \leq 0.02$). We did not find a significant difference in the way participants rated the relevance of the provided recommendations nor did we find a significant difference in the reported trust in the system or the perceived competence of the system.

Based on the above results, from this point onwards we adopt the correlation-based method as the *Explain* method for our investigation.

4.3 Evaluation in a Simulated Online-dating Environment

One of the main challenges in designing a realistic online-dating environment is the challenge of incorporating and modeling the costs and potential gains associated with accepting recommendations in the platform. Specifically, previous research has shown that different costs, especially an emotional cost such as fear of rejection, play prominent factors in determining the behavior of users in online dating platforms [13, 30]. Since the costs and potential gains involved with the acceptance of a recommendation (i.e., sending a message to the recommended user) may vary significantly between users, we consider two models: First, a model in which no explicit cost is introduced. Specifically, users are asked to rate the relevance of the recommended profiles without encountering any explicit cost or gain, as in the preliminary investigation described in section 4.2. We then consider a model in which explicit costs and potential gains are associated with accepting recommendations and users are incentivized to maximize their performance. The first model will assist us in understanding the effects of the explanation method when the cost is negligible, and the second when the cost is significant.

4.3.1 Negligible Cost. We asked the remaining 59 participants, out of the 118 participants who participated in the data collection (but did not participate in the evaluation of the *Explain* method discussed above), to take part in this experiment. Each participant was randomly assigned to one of two conditions: 1) one-sided explanations (30 participants); and 2) reciprocal explanations (29 participants). The participants reentered the MM environment and received five recommendations with an explanation corresponding to their condition. Similar to the experimental design discussed in Section 4.1, participants were asked to rate the *relevance* of each recommendation separately, on a five point Likert scale from 1 (extremely irrelevant) to 5 (extremely relevant), followed by the user experience questionnaire (see Appendix). In this setting, our working hypothesis was that the reciprocal explanation would have a significantly different effect on the participants in comparison with the one-sided explanations. However, since the recommendations in this setting did not involve cost, our hypothesis was non-directional, meaning we did not expect that the influence of reciprocal explanations would necessarily be positive or negative.

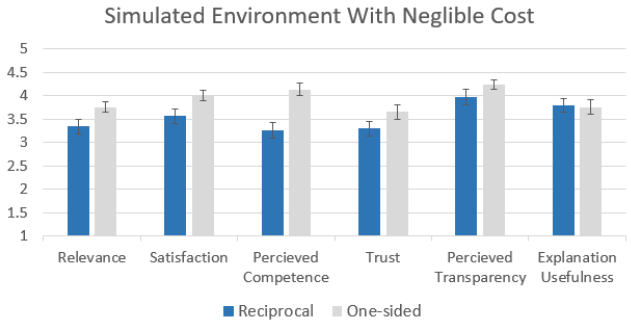

Figure 2: Reciprocal vs. one-sided explanations in MM with negligible cost. Error bars represent the standard error.

Results: All data was found to be distributed normally according to the Anderson-Darling normality test. In contrast to what one may expect, the one-sided explanation outperformed the reciprocal explanation in almost all tested measures. Specifically, using a *two-tailed unpaired t-test*, we found that the reported relevance (one-sided: mean= 3.76, s.d.= 0.61 vs. reciprocal: mean=3.34, s.d.= 0.84 $p \leq 0.04$), satisfaction (mean= 4 s.d.= 0.85 vs. mean= 3.57, s.d.=0.86 , $p \leq 0.05$) and perceived competence (mean= 4.13 s.d.= 0.83 vs. mean=3.27, s.d.=0.9 , $p \leq 0.01$) were all found to be significantly superior for the one-sided explanations condition. No statistically significant difference was found between the the conditions for the remaining measures.

Due the relatively small sample sizes it is extremely difficult to assess the differences between subgroups of the conditions and participants. For example, it is difficult to derive insights as to the possible difference in how females benefit from reciprocal explanation compared to the one-sided explanation condition. The experiment in the active online-dating application, described in Section 4.4, includes a significantly larger sample size and thus enables us to statistically analyze such subgroups.

4.3.2 Explicit Cost. For this experiment, we recruited 67 new participants who had not yet participated in this study (35 male and 32 female) ranging in age from 18 to 35 (average= 24.8 s.d.=4.74). Participants were then randomly assigned to one of the two conditions: one-sided explanations or reciprocal explanations. As was the case in the negligible-cost setting, participants created profiles, browsed profiles and sent messages to users they viewed as potential matches (as described above in Section 4.1). However, in the recommendation phase, the participants were given an incentive to maximize an artificial score which was effected by costs and gains as follows: Upon receiving a recommendation, each participant had

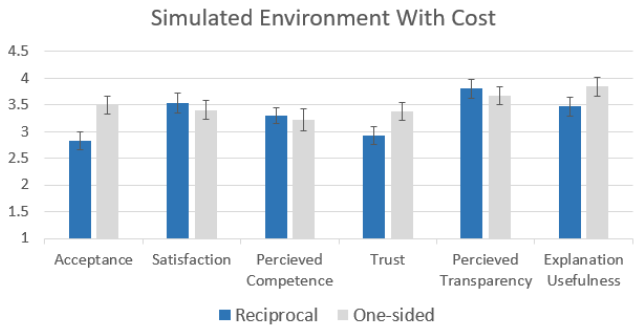

Figure 3: Reciprocal vs. one-sided explanations in MM with explicit cost. Error bars represent the standard error.

two options – either send a message to the recommended user or not. If the participant did not send a message, he or she did not gain or lose any points. If the participant did send a message, the recommended user returned a positive or negative reply according to a probability derived from the recommended user's preferences. Specifically, we used the interest of the recommended user in the participant, as estimated by the RECON algorithm. Participants were informed that the probability is based on the preferences of the recommended user. If the recommended user replied positively, the participant gained points proportional to how RECON estimated that the recommended user fit the user's preferences (between three and four points). If the recommended user replied negatively, the participant lost three points. This scoring scheme was chosen in order to encourage users to send messages to other users in whom they are interested while considering the probability of being rejected. Participants were paid proportional to their score. Complete technical details about this scoring and payment methodology are available on the *MM* website. Each participant then received 5 recommendations accompanied by an explanation according to their assigned condition. In this setup, we define the *acceptance rate* as the number of recommended users to which the participant chose to send messages. Later the participants filled out the user experience questionnaire as done in the previous setups. In this setting, our working hypothesis was that the reciprocal explanation condition would outperform the one-sided explanation condition, since we believed that reciprocal explanations would decrease the participants' uncertainty and reduce concerns regarding the cost.

Results: In contrast to the results of the previous experiment, the results here show a significant benefit to the reciprocal explanations method compared to one-sided explanations. Specifically, the acceptance of the reciprocal explanation condition was reported to be significantly higher than the one-sided condition (one-sided: mean=2.83 s.d.=0.88 vs. reciprocal: mean=3.49 s.d.=1.02, $p \leq 0.01$). Also, participants' trust in the system was found to be higher under the reciprocal explanation condition (one-sided: mean=2.93 s.d.=1.14 vs. reciprocal: mean=3.38 s.d.=1.01 , $p \leq 0.05$). No statistically significant difference was found between the conditions for the remaining measures.

The results are presented in Figure 3.

4.4 Evaluation in an Active Online-dating Application

After completing both experiments in the MM environment, we contacted *Doovdevan*, an Israeli online-dating application, and received permission to conduct a similar experiment within their application, using active users as participants.

Doovdevan is a web and mobile application customized for android and iOS operating systems. Similar to other online-dating applications, users of this platform can create profiles, search for possible matches and interact with other users via messages. Doovdevan currently consists of about 32, 000 users and is growing rapidly. We chose to perform our experiment on Doovdevan since it is relatively new and none of the users had received recommendations from the system prior to the experiment. This was important since previous recommendations can affect the trust of the users in the system and subsequently effect their attitude towards new recommendations [2, 17].

The recommendation algorithm that was implemented in the Doovdevan application was the TWO-SIDED COLLABORATIVE FILTERING method described above in Section 2.1.

We randomly selected a group of 161 active users on the site (i.e., users who logged on to the platform at least once in the week prior to the experiment), 78 males and 83 females, ranging in age from 18 to 69 (mean= 36.1, s.d.= 13.01), and randomly assigned them to one of the two examined conditions: one-sided explanations or reciprocal explanations. Due to privacy concerns, we were not permitted to reveal the recommended user's preferences to the recommendation receiver. Therefore, the reciprocal explanation included two (asymmetrical) parts: First, an explanation of the presumed interest of the recommendation receiver in the recommended user, including specific attributes of the recommended user, as done in the simulated MM environment. Second, a statement that the system believes that the recommendation receiver fits the recommended user's preferences, thus he/she is likely to reply positively.

The recommendations were sent to users' inboxes, and the user received a notification on her smartphone. The recommendation has a unique tagging in the application that distinguishes it from other incoming messages. The recommendation includes a brief description of the recommended user: low-resolution photograph, name, age, location, marital status. The user may click on the recommendation and thereby receive a higher quality photograph of the recommended user and an explanation (Figure 4). At this stage the user may send a message to the recommended user.

As in the previous experiment, each participant received five recommendations. However, unlike previous experiments, with Doovdevan only one recommendation was sent per day, based on the advice from the site owner who suggested that users would find it odd to receive multiple recommendations in a single day after not receiving a single recommendation thus far. Unlike the MM environment, in Doovdevan we could not explicitly ask participants for their experience. Therefore, we measure the *acceptance rate* of the provided recommendations as the number of recommendations that resulted in the recommendation receiver sending a message to the recommended user divided by the number of recommendations the recommendation receiver had viewed (clicked on). Although the recommendations in this real-world setting did not involve

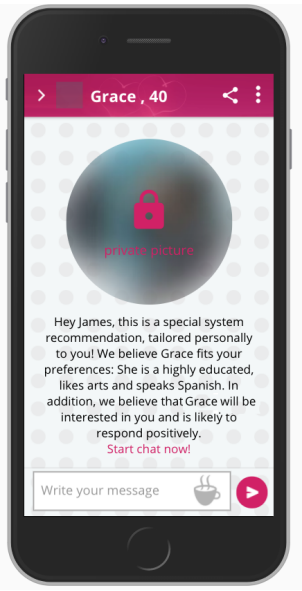

Figure 4: Screen shot of a reciprocal explanation for a recommendation in the active online-dating platform.

any monetary cost, we expect that the *emotional cost*, which is an established prominent factor in decision-making in online-dating environments [14, 15], will have an effect similar to the explicit cost in the simulated environment (Section 4.3.2). Therefore, we hypothesized that the reciprocal explanation condition will have a higher acceptance rate, similar to the results of the simulated environment.

Results. All data was found to be distributed normally according to the Anderson-Darling normality test. We compared both conditions using a t-test. The results show that users who received reciprocal explanations presented significantly higher acceptance rates compared to users who received one-sided explanations ($p < 0.05$). Specifically, on average, users who received reciprocal explanations sent messages to 53% of the recommended users they viewed while the same was true for only 36% of the recommended users under the one-sided explanations condition.

Interestingly, we find that reciprocal explanations outperform one-sided explanations for *women* while they do not show a statistically significant difference for men. Specifically, for women we find an average acceptance rate of 39% under the reciprocal explanation condition while only 25% under the one-sided explanations condition. For men, we find that the reciprocal explanation method achieves an average acceptance rate of 64% compared to 55% under the one-sided explanation method, but the difference is not statistically significant.

We further analyze the explanations' effect on users who sent more or less messages than the median number of messages sent by users in the system. We found that for the group who sent fewer messages than the median, the reciprocal explanation significantly outperformed the one-sided explanation, averaging a 47% acceptance rate compared to 25% under the one-sided explanations

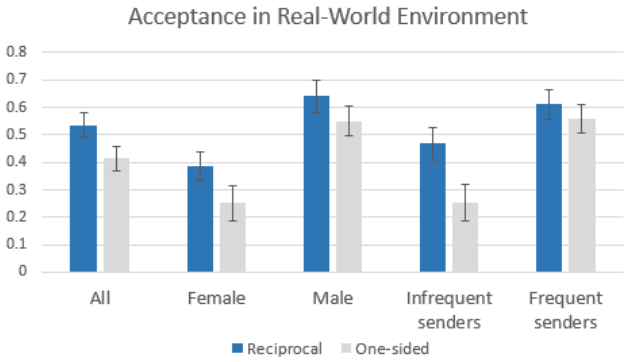

Figure 5: Reciprocal vs. one-sided explanations in a real online-dating environment. Error bars represent the standard error.

condition. For the complementary group, the reciprocal explanation averaged approximately 61% compared to 56% in the one-sided explanation, without a significant difference between the two. The results are presented in Figure 5.

We also examined the number of log-ins of the participants in the week following the recommendation as an additional potential impact of the explanation method. The results show that the participants under the reciprocal explanations condition logged-in significantly more often than those under the one-sided explanations, with an average of 56 log-ins compared to 23 log-ins under the one-sided explanations condition ($p <= 0.05$). These results possibly indicate that the users who received reciprocal explanations were more satisfied with the system.

5 DISCUSSION

The results from both the synthetic and real-world investigations suggest that the choice of explanation method depends on the users' cost for following the recommendations. Specifically, in environments where the cost of accepting a recommendation is high, the reciprocal explanations favorably compare to one-sided explanations. We suggest that this is because the additional information in the reciprocal explanation makes the user feel more confident in the outcome of accepting the recommendation, and subsequently this increases her willingness to take the risk.

The results are consistent with previous research which found that many users in online-dating platforms have an emotional cost for sending a message, mainly due to the fear of rejection [13, 14]. Specifically, when the fear of rejection was removed, as in our first simulation, the one-sided explanation method was found to be superior. In addition, our findings align with recent research which found that the cost associated with the advice has a significant effect on the acceptance of the recommendation [27].

Still, one may wonder *why* one-sided explanations were found to be superior to reciprocal explanations when negligible cost is introduced. We suggest two possible explanations:

(1) *Information overload.* Reciprocal explanations contain additional information which, if not deemed relevant by the recommendation receiver, may cause the recommendation as a whole to be less effective [12, 20].

(2) Users often perceive their own attractiveness in a different manner than others [3]. Therefore, it is possible that the users will have a negative reaction to an explanation that describes reasons for their attractiveness which do not match their own perception.

In a short informal interview subsequent to the experiment in the simulated environment, some participants expressed discomfort with the component of the explanation that focused on the the other's side preferences. This strengthens the last suggested reason for the results.

We further find that not all users respond to explanations in the same way, possibly suggesting that a "one-size-fits-all" explanation method is not likely to be found. Specifically, the cost associated with accepting a recommendation may vary between users. Previous work in the online dating domain has revealed that men tend to focus more on their own preferences compared to women who take into account their own attractiveness to the other side of the match [15, 30]. We find support for these insights in our study as well. We further find that users who are more "choosy" in their messaging behavior tend to benefit more from reciprocal explanations compared to other users. These differences between males and females or frequent and infrequent senders possibly indicate an underlying factor of emotional cost for sending messages, which is more likely to be prominent in infrequent message senders and females [14].

Our main contribution in this work is the introduction of reciprocal explanations and the evaluation of their effectiveness. We acknowledge that the explanation methods used in this study are relatively simple and we are currently working on more sophisticated methods. Specifically, in this work we used a generalized explanation method, which did not differentiate between users' presumed cost of rejection. We intend to extend this research and build a fully-personalized user model [24], which will model the user's considerations in a RRS based on her historical interactions, and provide reciprocal or one-sided explanations accordingly.

It is important to note that since we focused on online-dating, the above results are not immediately generalized to other reciprocal environments, such as job recruitment or roommate matching. Specifically, REs can vary widely in their inherent emotional cost of rejection which will presumably influence the effectiveness of reciprocal explanations. Therefore, we intend to explore additional REs in future work and include an investigation of how to personalize the explanation method to each specific user. We also intend to investigate *coalitional* reciprocal environments, where a user seeks to form or join a group of partners with whom to form a coalition. For example, a system which recommends potential research collaborators for scholars. In these environments, users often have preferences for a group of partners and therefore the explanations should be adapted accordingly.

6 CONCLUSION

In this paper we present a first-of-its-kind study which explores explanations for recommendations in REs. We introduce the use of reciprocal explanations, which includes reasoning for the presumed interest of both sides of the recommendation in the match. We extensively evaluated the proposed approach, compared it to the traditional one-sided explanation method in both simulated and real-world online-dating platforms, and found that the explanation method should depend on the users' cost (e.g. emotional) for accepting recommendations. Specifically, in environments where accepting the recommendations has a high cost, reciprocal explanations should be adopted, while if the cost is negligible, one-sided explanations should be adopted.

Detailed information about the MM platform and the collected data are available on the MM website: *www.biu-ai.com/Dating*.

7 APPENDIX: QUESTIONNAIRE FOR EVALUATION OF USER EXPERIENCE

Our questionnaire included 5 Likert scale questions, with a scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). These questions measured five prominent factors of user experience in recommender systems. We based the questions on previous questionnaires, such as [2, 16]. The questions are presented in Table 1. The second question, which is 'negatively worded', was *reversed-scored* [11]. In order to test discriminant validity (meaning that the different questions actually evaluate different measures) we calculated the Pearson correlation coefficients [7] which show that the answers to the question are not strongly correlated. The full correlation table is presented in Table 2. In future work we intend to use more elaborate measurement scales and perform SEM analysis.

Measure	Question
Satisfaction	1) I like the profiles that the system recommended to me.
System competence	2) The system is useless for me.
Trust	3) I trust the system to recommend all profiles that are of interest to me.
Transparency	4) I understand why the system recommended the profiles it did.
Explanation Usefulness	5) The explanations that were provided along with the recommendation were helpful.

Table 1: User Experience Questionnaire

	Question 2	Question 3	Question 4	Question 5
Question 1	0.427	0.448	0.426	0.291
Question 2		0.443	0.271	0.283
Question 3			0.449	0.292
Question 4				0.442

Table 2: Cross-scale Pearson Correlation Coefficients

REFERENCES

[1] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 1–4.

[2] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455–496.

[3] Tal Eyal and Nicholas Epley. 2010. How to seem telepathic: Enabling mind reading by matching construal. *Psychological Science* 21, 5 (2010), 700–705.

[4] David Gale and Lloyd S Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15.

[5] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382.

[6] Bryce Goodman and Seth Flaxman. 2016. European Union regulations on algorithmic decision-making and a" right to explanation". *Workshop on Human Interpretability in Machine Learning at the International Conference on Machine Learning* (2016).

[7] Robert Goodman. 2001. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry* 40, 11 (2001), 1337–1345.

[8] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* (2017).

[9] Ido Guy, Inbal Ronen, and Eric Wilcox. 2009. Do you know?: recommending people to invite into your social network. In *Proceedings of the 14th international conference on Intelligent user interfaces*. ACM, 77–86.

[10] Mark Andrew Hall. 1999. *Correlation-based feature selection for machine learning*. Ph.D. Dissertation. University of Waikato Hamilton.

[11] James Hartley. 2014. Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology* 14, 1 (2014), 83–86.

[12] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.

[13] Gunter J Hitsch, Ali Hortaçsu, and Dan Ariely. 2010. Matching and sorting in online dating. *American Economic Review* 100, 1 (2010), 130–63.

[14] Günter J Hitsch, Ali Hortaçsu, and Dan Ariely. 2010. What makes you click?âĂŤMate preferences in online dating. *Quantitative marketing and Economics* 8, 4 (2010), 393–427.

[15] Akiva Kleinerman, Ariel Rosenfeld, Francesco Ricci, and Sarit Kraus. 2018. Optimally Balancing Receiver and Recommended Users' Importance in Reciprocal Recommender Systems. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM.

[16] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.

[17] Sherrie YX Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly* (2006), 941–960.

[18] National Science and Technology Council. 2016. The National Artificial Intelligence Research And Development Strategic Plan. (2016).

[19] Luiz Pizzato, Tomek Rej, Thomas Chung, Irena Koprinska, and Judy Kay. 2010. RECON: a reciprocal recommender for online dating. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 207–214.

[20] Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces*. ACM, 93–100.

[21] Pearl Pu and Li Chen. 2007. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems* 20, 6 (2007), 542–556.

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

[23] Nornadiah Mohd Razali, Yap Bee Wah, et al. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics* 2, 1 (2011), 21–33.

[24] Ariel Rosenfeld and Sarit Kraus. 2018. Predicting Human Decision-Making: From Prediction to Action. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 12, 1 (2018), 1–150.

[25] Amit Sharma and Dan Cosley. 2013. Do social explanations work?: studying and modeling the effects of social explanations in recommender systems. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 1133–1144.

[26] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*. ACM, 830–831.

[27] Steven C Sutherland, Casper Harteveld, and Michael E Young. 2016. Effects of the Advisor and Environment on Requesting and Complying With Automated Advice. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, 4 (2016), 27.

[28] Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos. 2009. MoviExplain: a recommender system with explanations. In *Proceedings of the third ACM conference on Recommender systems*. ACM, 317–320.

[29] Kun Tu, Bruno Ribeiro, David Jensen, Don Towsley, Benyuan Liu, Hua Jiang, and Xiaodong Wang. 2014. Online dating recommendations: matching markets and learning preferences. In *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 787–792.

[30] Peng Xia, Benyuan Liu, Yizhou Sun, and Cindy Chen. 2015. Reciprocal recommendation system for online dating. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*. ACM, 234–241.

[31] Hongtao Yu, Chaoran Liu, and Fuzhi ZHANG. 2011. Reciprocal recommendation algorithm for the field of recruitment. *JOURNAL OF INFORMATION AND COMPUTATIONAL SCIENCE* 8, 16 (2011), 4061–4068.