The influence of perceived credibility on preferences for recommender systems as sources of advice

Kyung Hyan Yoo  
Texas A&M University

Ulrike Gretzel  
Texas A&M University, ugretzel@uow.edu.au

Publication Details
The influence of perceived credibility on preferences for recommender systems as sources of advice

Abstract
Recommender systems promise to support travelers in complex decision-making processes; however, whether a recommendation is seen as credible advice and actually taken into account not only depends on travelers' perceptions of the recommendation but also of the system as the advice giver. A scale to measure recommender system credibility was developed and tested. The results confirm that credibility has two dimensions: expertise and trustworthiness. Further, significant gender differences in credibility perceptions were found. The findings also indicate that respondents prefer humans as recommendation sources and that this preference is influenced by perceptions of lack of credibility of recommender systems as well as gender-specific preferences. Implications for future research and for recommender system design are discussed.

Keywords
influence, perceived, credibility, preferences, for, recommender, systems, sources, advice

Disciplines
Business | Social and Behavioral Sciences

Publication Details
THE INFLUENCE OF PERCEIVED CREDIBILITY ON PREFERENCES FOR RECOMMENDER SYSTEMS AS SOURCES OF ADVICE

KYUNG HYAN YOO and ULRIKE GRETZEL

Laboratory for Intelligent Systems in Tourism, Department of Recreation, Park & Tourism Sciences, Texas A&M University, USA

Recommender systems promise to support travelers in complex decision-making processes; however, whether a recommendation is seen as credible advice and actually taken into account not only depends on travelers’ perceptions of the recommendation but also of the system as the advice giver. A scale to measure recommender system credibility was developed and tested. The results confirm that credibility has two dimensions: expertise and trustworthiness. Further, significant gender differences in credibility perceptions were found. The findings also indicate that respondents prefer humans as recommendation sources and that this preference is influenced by perceptions of lack of credibility of recommender systems as well as gender-specific preferences. Implications for future research and for recommender system design are discussed.

Key words: Recommender systems; Credibility; Expertise; Trustworthiness; Recommendation source preference; Gender differences

Introduction

The Internet has without doubt become an important travel information source. According to the Travel Industry Association of America (2005), more than half (52%) of US travelers used the Internet for travel planning in 2005. Recent empirical studies even indicate that Internet searches are overtaking personal sources such as friends and relatives as the preferred means for obtaining travel information (eMarketer, 2005). However, it is often difficult for consumers to find information in digital environments and too much information can easily lead to confusion and information overload (Henry, 2005). In order to effectively manage the amount of information to be processed during online searches, consumers seek and accept recommendations (Smith, Menon, & Sivakumar, 2005). Online recommendations can be obtained from three broad categories of sources: 1) other consumers, 2) human experts, and 3) expert systems such as recommender systems (Sénécal & Nantel, 2003).

Recommender systems have been developed for many websites (e.g., Amazon.com and Netflix.com) and are expected to play an increasingly im-

Address correspondence to Ulrike Gretzel, Laboratory for Intelligent Systems in Tourism, Department of Recreation, Park & Tourism Sciences, Texas A&M University, 2261 TAMU, College Station, TX 77843-2261, USA. Tel: +1-979-862-4043; Fax: +1-979-845-0446; E-mail: ugretzel@tamu.edu
portant role in helping consumers find what they need and want (Barwise, Hammond, & Elberse, 2002; B.-D. Kim & Kim, 2001). In the context of tourism, recommender system applications are anticipated to have a great impact on travel information distribution and consumers’ travel planning behavior (Werthner & Ricci, 2004). Yet, although recommender systems make recommendations based on often sophisticated data mining and analysis techniques, it cannot be automatically implied that the advice provided by the system will be accepted by the consumers. When consumers receive recommendations, they selectively choose whether the piece of advice is taken into account. It is important for recommender system research and design to examine factors that can influence the likelihood of recommendations to be accepted and integrated into decision-making processes. What makes message recipients perceive a recommendation as persuasive has been studied in numerous disciplinary fields, most prominently however in the persuasion literature (O’Keefe, 2002). Persuasion has recently been studied in the context of technology (Fogg, 2003) and also specifically in the realm of recommender systems (Dijkstra, Liebrand, & Timminga, 1998; Gretzel & Fesenmaier, 2006; Häubl & Murray, 2003). Gretzel and Fesenmaier (2006), for instance, argue that recommender system users’ preferences can be affected by the interaction with the system and found that structural characteristics of the preference elicitation process influence users’ construction of preferences for vacation destinations.

Fitzsimons and Lehmann (2004) stress the importance of reducing the potential for reactance (i.e., resistance to persuasion) in advice-seeking relationships. Trust seems to play an important role in reducing reactance and, thus, in increasing the potential for advice to be taken into account. Several studies in the field of information systems have investigated the issue of trust and its effect on consumer behavior (Gefen, 2000; Jarvenpaa, Tractinsky & Vitale, 2000; K. Kim & Prabhakar, 2000; Stewart, 1999). Trust has also been treated as an important issue in recent recommender system research (Bauernfeind & Zins, 2006; Bickmore & Cassell, 2001; Wang & Benbasat, 2005). However, most online trust studies are largely focused on security issues, ignoring the social psychological process of trust. In human-to-human advice-seeking relationships, source credibility has been identified as a highly influential factor (Gilly, Graham, Wilfinbarger, & Yale, 1998; Harmon & Coney, 1982; Lascu, Bearden, & Rose, 1995; O’Keefe, 2002). Importantly, source credibility has been found to matter when computers give advice or provide instructions to users (Fogg, 2003; Sénécal & Nantel, 2003, 2004). Consequently, this article argues that trust in recommender systems is more an issue of credibility rather than security and that a specific credibility measurement instrument needs to be developed to capture important trust issues in human–recommender system interactions as they will likely determine a user’s propensity to accept a recommendation made by the system. However, there is no commonly agreed upon credibility measurement scale. Thus, the goals of our study are 1) to develop a survey instrument to measure the perceived credibility of recommender systems, and 2) to apply the developed measurement scales to investigate the influence of perceived source credibility on recommendation source preference.

**Theoretical Foundations**

**Role of Recommender Systems**

Häubl and Trifts (2000) defined recommender systems as software tools that make recommendations based on learned information about a user’s preference function. Similarly, Xiao and Benbasat (2007) defined recommender systems as “software agents that elicit the interests or preferences of individual users for products, either explicitly or implicitly, and make recommendations accordingly” (p. 137). Schafer, Konstan, and Riedl (2001) provide a typology of recommender systems stressing that these systems can take on many forms with different levels of sophistication, different inputs used to derive recommendations, and various ways in which they present recommendations. Patton (1999) acknowledged that recommender systems promise to make shopping on the Internet better, not just by finding lower prices but by matching products to the needs and tastes of individual consumers. Through this provision of personalized and consequently more relevant advice, these systems offer critical support in an online
shopping context (Cheung, Kwok, Law, & Tsui, 2003). Electronic recommender systems have been considered as one way for marketers to enhance their e-services by reducing cognitive effort for individuals when making decisions in online contexts (Kleinmutz & Schkade, 1993; Rust, 2001); thus, they are expected to become the fastest growing domain of Internet applications (Spiekermann & Paraschiv, 2002).

In tourism, the role of recommender systems is even more essential considering the extensive amounts of tourism-related information available online and the experiential nature of tourism products. Indeed, tourism is especially affected by the online information explosion and searching for travel-related information is one of the most common online activities (Pew Internet & American Life Project, 2006). Tourism recommender systems can play a vital role in travel information search processes by providing decision-support in the form of inspiration (i.e., the expansion of one's consideration set) as well as by narrowing down available alternatives (Fesenmaier, Werthner, & Wöber, 2006). Online travel agencies as well as travel providers have started to develop recommender systems to support travel planning (Higgins, 2007), with Homeandabroad.com, the Travelocity Experiencefinder (http://labs.travelocity.com/experiencefinder), VIBE (Jannach, Zanker, Jessenitschnig, & Seidler, 2007), and Hungrysuitcase.com being recent examples. With the rising importance of providing travelers with online travel planning support, the topic of travel recommender systems has received growing research attention. The current literature on recommender systems in tourism addresses not only the importance of recommender systems in travel planning processes but also the need to develop systems that are more human centric and tailored to the specific needs of travelers in different travel planning stages. Gretzel, Hwang, and Fesenmaier (2006) proposed that understanding the complexities of tourists’ information search and decision-making-related behaviors is fundamental for successful recommender system design. Ricci, Blaas, Mirzadeh, Venturini, and Werthner (2002) have developed and tested prototype systems for different travel planning contexts (before a trip and en route) such as DieToRecs and NutKing and stressed the importance of adding visual cues to destination descriptions. Franke (2002) and Kramer, Motschinger, ten Hagen, and Gretzel (2006) draw attention to the social nature of travel that is currently not reflected in recommender systems. Jannach et al. (2007) propose a conversational agent to mimic conversations one would have with a traditional travel agent. Further, Gretzel and Fesenmaier (2002) suggest including principles of narrative design into recommender systems, arguing that tourism experiences are bundles of tourism product elements that can be better understood as cohesive wholes if presented through narratives. The importance of hedonic aspects in designing travel recommender systems is emphasized by D.-Y. Kim and Morosan (2006).

Evaluations of recommender systems have largely focused on impacts on decision quality (Ansari, Essagier, & Kohli, 2000; Ariely, 2000; Benbasat & Nault, 1990; Häubl & Trifts, 2000; Todd & Benbasat, 1992). In contrast, Aksoy, Bloom, Lurie, and Cool (2006) urge that recommender systems should be understood in terms of the quality of the interaction with the user. Similarly, Gretzel (2004) calls for evaluation of the persuasiveness of recommender systems and suggests that recommender system–user interactions are fundamentally social relationships to which social theories and concepts apply. Xiao and Benbasat (2007) argue that recommender systems research needs to include not only recommender system characteristics, use, and decision outcomes but also user–recommender systems interactions and user evaluations of recommender systems.

Recommender Systems as Social Actors

Although recommender systems have been identified as important sources of advice for consumers, most existing recommender system studies have viewed them as information search tools and have largely neglected their social role in the interaction with users. More recent studies, however, argue that computer applications like recommender systems need to be understood as “social actors” (Nass & Moon, 2000). Nass and Moon found that people construct social relationships with machines including computers, and apply social rules in their interactions with technology.
Recommender system studies also support this argument. Findings by Aksoy et al. (2006) suggest that a consumer is more likely to positively evaluate and use a recommender agent when it generates recommendations in a way similar to the consumer’s decision-making process. Morkes, Kernal, and Nass (1999) demonstrated that computer agents that use humor are rated as more likable, competent, and cooperative. In addition, the recommender system’s role as a trusted friend was found to be important to support system users’ decision making (Bauernfeind & Zins, 2006; Wang & Benbasat, 2005). Consequently, recommender systems have to be understood as communication sources to which theories developed for human–human communication apply. One set of such theories relates to the impact of source characteristics on persuasion.

**Determinants of Recommendation Source Preference**

In 1968, Andreasen proposed four types of communication sources: 1) Impersonal Advocate (e.g., mass media); 2) Impersonal Independent (e.g., Consumer Reports); 3) Personal Advocate (e.g., salesman); and 4) Personal Independent (e.g., friends). Sénechal and Nantel (2004) applied and adapted Andreasen’s typology to computer-mediated environments. They asserted that online users can obtain information and recommendations from four groups: 1) personal source providing personalized information; 2) personal source providing nonpersonalized information; 3) impersonal source providing personalized information; 4) impersonal source providing nonpersonalized information. Sénechal and Nantel classify recommender systems as impersonal information sources that provide personalized information to consumers.

Not all recommendation sources are equally liked and used. White (2005) found that recommendation receivers preferred friends over expert sources when the decision was highly emotional. Similarly, Smith et al. (2005) found that recommendations provided by online peers were highly preferred over editorial recommendations. In contrast, Sénechal and Nantel (2004) reported that recommender systems are the most influential source, outperforming friends and experts. Studies conducted by Wærn and Ramberg (1996) also indicate contrasting findings. In one study, they found that advice receivers generally trust human beings more than computers as advice givers. However, the results of a second study indicated that computers received higher attributions regarding trust and understanding. Wærn and Ramberg (1996) explained that these differences can be caused by different initial attitudes towards sources and also by prior experience of the user. Smith et al. (2005) argued that online consumers’ preference for recommendation sources depends on the specific nature of their shopping goal.

The traditional persuasion literature has argued that source characteristics are influential factors in determining recommendation source preference. For instance, likability of the source has been suggested as an important factor in determining what makes a source preferred over another (O’Keefe, 2002; Shavitt & Brock, 1994). Similarity of the source has also been found to have an impact on recommendation recipients’ favorability towards a source (Brock, 1965; Woodside & Davenport, 1974). In addition, the effect of physical attractiveness on source preference has received considerable research attention (Chaiken, 1986; Horai, Naccari, & Fatoullah, 1974; Widgery & Ruch, 1981). Among the suggested possible causes, however, communicator credibility has been recognized as the factor that has the greatest effect on the persuasive power of a source (O’Keefe, 2002).

**Source Credibility**

O’Keefe (2002) argued that credibility is not an intrinsic characteristic of a source; rather, the decision regarding a communicator’s credibility depends on how the message recipient perceives the source. Thus, source credibility can be defined as judgments made by a message receiver concerning the believability of a communicator (Fogg, Lee, & Marshall, 2002). Reviews of source credibility studies by Anderson and Clevenger (1963) and McGuire (1968) concluded that a more credible source is preferred and also more persuasive. A number of recent studies confirm that source credibility is positively correlated with influence on message recipients’ attitudes and behavioral intentions as well as behaviors (Gilly et al., 1998; Har-
mon & Coney, 1982; Lascu et al., 1995; Sénécal & Nantel, 2003, 2004). Credibility of information sources has also received limited attention in the realm of tourism. Kerstetter and Cho (2004) found that the perceived credibility of a travel information source is the strongest predictor for its actual usage. Teichmann and Zins (2006) investigated the perceived credibility of travel information over the course of different trip stages and found differences for pretrip and posttrip contexts. Importantly, Fogg (2003) and Fogg et al. (2002) suggest that source credibility also matters when people interact with computers and argue that credibility is particularly important when computers give advice or provide instructions to users. Credibility has also been identified as a critical factor in the context of receiving advice from recommender systems (Swearingen & Sinha, 2001; Urban, Sultan, & Qualls, 1999; Westerink, Bakker, De Rijder, & Siepe, 2002).

Credibility is described as comprising multiple dimensions (Buller & Burgoon, 1996; Gatignon & Robertson, 1991; Petty & Cacioppo, 1981; Self, 1996). Although the literature suggests various dimensions of credibility, most researchers agree that it is comprised of two key elements: trustworthiness and expertise (Fogg, 2003; Fogg et al., 2002; O’Keefe, 2002; Rhoads & Cialdini, 2002). Figure 1 illustrates the dimensional nature of credibility.

**Trustworthiness.** Trustworthiness of a source refers to aspects such as character or personal integrity (O’Keefe, 2002). Intentions are also seen as instrumental in determining the trustworthiness of a source. A source whose intent it is to persuade is perceived as less trustworthy than one without persuasive intent (Petty & Cacioppo, 1981). Consequently, trustworthiness is often described by terms such as well-intentioned, truthful, and unbiased (Fogg et al., 2002). Mayer, Davis, and Schoorman (1995) conceptualized benevolence and integrity as dimensions of trustworthiness. Delgado-Ballester (2004) identified reliability and intentions as important trustworthiness dimensions. Fogg (2003) identified key points that affect the perceptions of trustworthiness: 1) a source is fair and unbiased; 2) a source would argue against their own interest; and 3) a source has perceived similarity. According to O’Keefe (2002), the trustworthiness dimension is commonly represented by scales such as honest–dishonest, trustworthy–untrustworthy, just–unjust, fair–unfair, and unselfish–selfish. These items are related to the assessment of whether the communicator will likely be inclined to tell the truth (O’Keefe, 2002). In the context of recommender systems, Xiao and Benbasat (2007) propose to test benevolence and integrity of recommender systems, with benevolence being defined as the recommender system caring about the user and acting in the user’s interest, and integrity being described as the recommender system’s adherence to a set of principles (e.g., honesty) that the user finds acceptable.

**Expertise.** Mayer et al. (1995) describe expertise as the ability of a source to have influence in a certain domain. Fogg et al. (2002) conceptualize it using terms such as knowledgeable, experienced, and competent; thus, this dimension seems to capture the perceived knowledge and skill of the source. Similarly, O’Keefe (2002) referred to expertise as competence, expertness, or qualification. Fogg (2003) provides many examples for cues that lead to perceptions of expertise such as labels that proclaim one as an expert, appearance cues, and documentation of accomplishments. In expertise research, this dimension is commonly represented by scales such as experienced–inexperienced, informed–uninformed, trained–untrained, qualified–unqualified, skilled–unskilled, intelligent–unintelligent, and expert–not expert (O’Keefe, 2002). These items are related to the assessment of whether the communicator is in a position to

![Figure 1. Key dimensions of credibility (adapted from Fogg, 2003).](image-url)
know the truth, to know what is right or correct (O’Keefe, 2002). Xiao and Benbasat (2007) describe the competence of a recommender system as the system’s ability, skills, and expertise to perform effectively.

O’Keefe (2002) insists that sources are only perceived as credible if they are identified as being high in both trustworthiness and expertise. Fogg (2003) argues that trustworthiness and expertise do not necessarily go hand in hand. According to Fogg, one can perceive a source as trustworthy but without expertise and as an expert source that is not trustworthy; yet, both dimensions have to be positively evaluated for a source to be perceived as credible. However, if one dimension of credibility is strong while the other dimension is unknown, the source still may be perceived as credible, due to the so-called “halo effect” (Fogg, 2003).

In contrast, if one dimension is known to be weak, credibility suffers, regardless of the other dimension (Fogg, 2003). Importantly, expertise and trustworthiness are conceptually distinct aspects of credibility; thus, it is possible to manipulate and measure them separately to examine their individual effects on persuasive outcomes (O’Keefe, 2002).

**Conceptual Model and Hypotheses**

Based on this review of the literature, credibility of a recommender system was conceptualized as the perceived expertise and perceived trustworthiness of a recommender system. Although these two aspects both signify credibility, they are assumed to be theoretically different and, thus, should be measured as two separate dimensions (Fig. 2). Taking into account the importance of perceived credibility for recommender system evaluations stressed in the literature, it is proposed that higher perceptions of recommender system expertise and trustworthiness lead to greater preference of recommender systems over other recommendation sources such as experts or friends. Therefore, the following hypothesis was formulated:

**H1:** The greater perceptions of trustworthiness and expertise of recommender systems, the greater the likelihood that they will be preferred over human sources of advice.

In addition, gender differences in advice-seeking behavior, technology use, and overall credibility perceptions are expected to influence ratings regarding the expertise and trustworthiness of recommender systems. Because Fogg’s (2003) credibility model served as the basis for this study, the results of the Fogg et al. (2001) study, which pointed towards lower credibility ratings for men, were deemed to be most relevant. Based on these theoretical assumptions, the following hypothesis is proposed:

**H2:** Trustworthiness and expertise ratings will be lower when assessed by men.

**Methodology**

The methodology applied in the context of this study involved a paper-based survey that was administered to a sample of undergraduate students...
enrolled at a university in the US. Such a sample is appropriate as the purpose of the study is not the generalization of results to an empirical population but rather theoretical generalization, and the students who participated in this study represent a theoretical population (Webster & Sell, 2007).

Survey Design and Administration

The survey instrument to measure the two dimensions of credibility was designed based on the credibility model proposed by Fogg (2003) and the various descriptions of trustworthiness and expertise found in the literature. Expanding on Fogg’s credibility model, trustworthiness was defined as encompassing reliability and intentions because existing research provides evidence that both are important aspects of trustworthiness (Deldago-Ballester, 2004). Stepp (1990) conceptualizes expertise as the breadth and depth of knowledge, reliability as consistent and continued performance with respect to reasonable solutions, and intentions as the system’s purpose and motives of its designers. Using these definitions, a list of 25 expertise, reliability, and intentions questions was created. All questions were measured on a 7-point Likert scale ranging from 1 = strongly disagree to 7 = strongly agree.

An initial version of the survey was pretested and the result showed problems with reverse scored items and with the wording of some questions (most notably, subjects seemed to not have a clear understanding of the term “objective”). The survey was revised to include a total of 20 items: nine expertise items, six reliability items, and five intentions items (Table 1). In addition to the credibility-related items the survey included questions regarding respondents’ familiarity with recommender systems (measured on a 7-point scale ranging from not at all familiar to very familiar) and their preferences for getting advice from: 1) either a recommender system or a friend; 2) a friend or an expert; or 3) an expert or a recommender system to test the proposed model and hypotheses. Preferences for sources were measured using 7-point semantic differential scales (Recommender System–Friend; Expert–Friend; Recommender System–Expert). Subjects were also asked to indicate their gender to be able to test for gender-specific differences in credibility perceptions.

The study aimed at testing perceptions of and preferences for recommender systems as a type of source for advice in general, rather than for a specific system or type of system. To ensure that survey respondents did not think of a specific system, they were presented with examples of recommender systems at the beginning of the survey and instructed to think about these in general. Although no tourism-specific example was included as no prominent system was available at the time of the survey, travel recommender systems fall under the general category of recommender systems. Consequently, the survey findings apply to tourism-specific systems. The survey was administered to a total of 109 students for partial course credit.

Analysis

First, the expertise and trustworthiness scales were tested for their reliability and dimensionality. Cronbach’s alpha was used to test the reliability of the scales and to purify the scales by excluding...
Table 1
Survey Items

<table>
<thead>
<tr>
<th>Expertise: Ability to Provide Valuable Recommendations</th>
<th>Trustworthiness</th>
<th>Reliability</th>
<th>Intentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Make decisions easier.</td>
<td>• Are reliable.</td>
<td>• Are consistent in the recommendations they provide.</td>
<td>• Are not biased.</td>
</tr>
<tr>
<td>• Can provide me with more accurate recommendations than human beings.</td>
<td>• Are consistent in the recommendations they provide.</td>
<td>• Do not make mistakes.</td>
<td>• Want me to find an option that best fits my needs.</td>
</tr>
<tr>
<td>• Can provide me with more valuable recommendations than human beings.</td>
<td>• Can be trusted.</td>
<td>• Can be trusted.</td>
<td>• Are designed with the best intentions in mind.</td>
</tr>
<tr>
<td>• Help me find things I really like.</td>
<td>• Are predictable.</td>
<td>• Are dependable.</td>
<td>• Are there to help me.</td>
</tr>
<tr>
<td>• Have access to and can process more information than human beings.</td>
<td>• Are dependable.</td>
<td>• Can provide me with more accurate recommendations than human beings.</td>
<td>• Are good way to get suggestions from a neutral source.</td>
</tr>
<tr>
<td>• Provide useful suggestions.</td>
<td>• Are predictable.</td>
<td>• Can be trusted.</td>
<td>• Are a good way to learn about different product options.</td>
</tr>
<tr>
<td>• Have better insight in what I like than most of my friends.</td>
<td>• Are predictable.</td>
<td>• Can be trusted.</td>
<td>• Can provide me with more accurate recommendations than human beings.</td>
</tr>
<tr>
<td>• Offer suggestions that I might not have thought of.</td>
<td>• Are consistent in the recommendations they provide.</td>
<td>• Do not make mistakes.</td>
<td>• Are designed with the best intentions in mind.</td>
</tr>
<tr>
<td>• Are a good way to learn about different product options.</td>
<td>• Are predictable.</td>
<td>• Can be trusted.</td>
<td>• Are there to help me.</td>
</tr>
</tbody>
</table>

those variables that significantly lowered the alpha statistics. Factor analyses with Principal Components extraction and Varimax rotation were conducted to examine the dimensional structure of the scales. Second, correlation analyses were conducted for familiarity with recommender systems and the credibility measures to ensure that it would not have to be considered in the modeling of the relationships. Third, the hypothesized relationships between the model constructs were tested. Because the sample was rather small and the scales contain a fairly large number of variables, the two credibility scales were constructed a priori by adding the scores and dividing them by the respective number of items and then entered into a path model using LISREL 8.52 (Jöreskog & Sörbom, 2000). The analysis was based on a polychoric correlation matrix because gender was measured as a dichotomous variable. Maximum likelihood was employed as the estimation method as there is considerable evidence that maximum likelihood estimation is robust with respect to many types of violation of the multivariate normality assumption (Chou & Bentler, 1996). Two separate models were tested: one with preference for recommender systems over experts and one with preference for recommender systems over friends as the dependent variable. Due to missing data the sample size for the two models was 99 and 98 respondents, respectively.

Results

About 57% of the respondents were male and 43% female. All respondents were between 20 and 25 years old. Most respondents were moderately familiar with recommender systems and about 25% indicated they were very familiar with such recommender technologies. Because familiarity could potentially influence perceptions regarding recommender systems, Pearson product-moment correlation coefficients were calculated for familiarity and the credibility items. Among the 40 items, familiarity was only significantly ($p < 0.05$) correlated with two items, with correlation coefficients of 0.250 and 0.201, respectively. Because the correlation was very weak and most likely due to chance, it was not considered for further analyses.

Credibility Scale Development

Based on Cronbach’s coefficient alpha scores the scales were further purified. Two items were excluded from the expertise scale ("can provide me with more accurate recommendations than human beings" and “have better insight in what I like than most of my friends”) and two items also had to be excluded from the trust scale (”do not make mistakes” and “are predictable”) because they significantly lowered the alpha score. From these efforts a nine-item trustworthiness scale and a
seven-item expertise scale emerged and showed very good measurement properties (Table 2). The scales were further subjected to factor analyses. The eigenvalues, factor loadings, and variance explained suggest that both scales are unidimensional. Thus, it was decided to have only one trustworthiness scale rather than separate scales for reliability and intentions.

Influence of Credibility Perceptions on Source Preference: Test of Hypotheses

Mean ratings for the credibility items were generally not very high (Table 2), suggesting that recommender systems are only perceived as moderately credible providers of recommendations. Respondents rated the expertise of recommender systems slightly more favorable than their trustworthiness. As far as preferences for sources of advice were concerned, a clear majority of the respondents (83%) preferred receiving advice from a friend over advice from a recommender system, with 11% being undecided. When asked to decide between recommender system and human expert, 78% preferred the human expert and 9% were undecided. Finally, when requested to make a decision between friend and human expert, about the same percentages of respondents chose friend (44%) and expert (43%), while 13% remained undecided. Consequently, a clear preference for human sources over recommender systems seems to prevail and the dependent variables for the model testing were formulated in terms of preferences for human sources over recommender systems.

The results of the path analyses provide important insights regarding the factors that drive these preferences for friends and experts as sources of advice. Two separate models were estimated: one with preferences for friends over recommender systems and one with preferences for experts over recommender systems as the dependent construct. In order to improve model fit, the initial models were modified to include a relationship between trustworthiness and expertise. This suggests that the two constructs are conceptually different but not independent from each other and confirms Fogg’s (2003) hypothesis that evaluations of one dimension influence perceptions of the other. Further, insignificant relationships were deleted from the final model estimates. Both final models showed acceptable fit with high goodness of fit indices, an insignificant chi-square statistic, and RMSEA and RMR values close to or below 0.05 (Fig. 3).

Table 2
Measurement Properties of Modified Scales

<table>
<thead>
<tr>
<th>Construct Name and Items</th>
<th>Mean Ratings</th>
<th>SD</th>
<th>Factor Loadings</th>
<th>Eigen Value</th>
<th>Variance Expl.</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendation systems . . .</td>
<td>4.64</td>
<td>0.86</td>
<td>3.57</td>
<td>51.01%</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide useful suggestions</td>
<td>4.82</td>
<td>1.17</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Make decisions easier</td>
<td>4.59</td>
<td>1.30</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are a good way to learn about different product options</td>
<td>5.03</td>
<td>1.10</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offer suggestions that I might not have thought of</td>
<td>5.08</td>
<td>1.20</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have access to and can process more information than humans</td>
<td>5.28</td>
<td>1.26</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help me find things I really like</td>
<td>4.26</td>
<td>1.31</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can provide me with more valuable recommendations than humans</td>
<td>3.33</td>
<td>1.17</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trustworthiness</td>
<td>4.40</td>
<td>0.81</td>
<td>3.94</td>
<td>43.77%</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Are reliable</td>
<td>4.27</td>
<td>1.10</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are dependable</td>
<td>4.19</td>
<td>1.11</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are designed with the best intentions in mind</td>
<td>4.80</td>
<td>1.40</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can be trusted</td>
<td>4.11</td>
<td>1.26</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are not biased</td>
<td>3.82</td>
<td>1.61</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Want me to find an option that best fits my needs</td>
<td>4.33</td>
<td>1.11</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are a good way to get suggestions from a neutral source</td>
<td>4.37</td>
<td>1.39</td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are consistent in the recommendations they provide</td>
<td>4.74</td>
<td>1.02</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are there to help me</td>
<td>5.00</td>
<td>1.22</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The results indicate a significant influence of gender on perceptions of credibility and significant influences of perceptions of credibility on preferences for recommender systems as recommendation sources. However, the results are somewhat different for the comparison of recommender systems with friends as opposed to with experts. Preferences for friends over recommender systems are driven by perceptions of lack of expertise of recommender systems whereas preferences for experts are influenced by perceptions of lack of trustworthiness. The weaker the perceptions of recommender system expertise, the greater the preference for friends as sources of recommendations; and the weaker the perceptions of recommender system trustworthiness, the greater the preference for experts over recommender systems. Also, the preference for experts as recommendation sources is indirectly and directly influenced by gender while the preference for friends is only indirectly influenced by gender. Women tend to rate the trustworthiness (and indirectly also the expertise) of recommender systems higher and, consequently, are more likely to prefer recommender systems as sources of advice. In addition, males prefer experts as recommendation sources regardless of their perceptions of recommender systems.

Consequently, the results confirmed Hypothesis 1 that higher ratings of recommender system credibility lead to greater preferences for them as sources of advice, although the specific dimension of credibility that drives this influence differs for preferences for recommender systems over experts versus over friends. The results also partially confirm Hypothesis 2 in that males tend to rate the trustworthiness of recommender systems as lower but no influence of gender was found for ratings of expertise.

Discussion

Although the generalizability of the study is limited due to its specific sample and its general reference to recommender systems rather than a specific type of system, the results provide important theoretical implications regarding factors that influence recommender system adoption and use, methodological implications for the further development of instruments to measure the credibility of recommender systems, as well as practical implications regarding the design of recommender systems for travel and tourism contexts.

Werthner and Ricci (2004) identified the development of recommender systems as one of the...
most important and promising trends in the realm of information technology and tourism. As suggested by Xiao and Benbasat (2007), the future design of such recommender systems needs to be informed by research that takes into account not only the characteristics of the recommender system and their influence on decision outcomes but also other factors that can potentially determine recommender system use. Following Xiao and Benbasat’s (2007) framework, this study was able to provide insights regarding the influence of user characteristics (gender), user–recommender system interactions in terms of user familiarity, and the influence of user evaluations of recommender systems (specifically credibility evaluations) on intentions to use them as sources of advice. As such, it represents an important contribution to the recommender systems literature overall and specifically in the tourism domain.

The developed credibility scale provides an important instrument to measure the credibility of recommender systems, which should be applied in future recommender systems studies and further tested regarding its reliability. Concerning future credibility scale development, the current expertise construct seems to cover breadth of knowledge and access time but not depth of knowledge, because questions related to accuracy and insight had to be excluded from the scale. Future research should explore whether developing items to reflect this aspect of expertise can improve credibility measurement. Also, it is not clear what drives the gender differences in credibility perceptions. It could be a greater openness to accept advice in women that leads to more favorable evaluations. In contrast, gender-specific attitudes toward technologies could also lead to such differences. Or some aspects of existing recommender systems are more attractive for women (e.g., the inherently social nature of collaborative filtering applications). These factors need to be identified and successfully measured to provide more specific insights.

As mentioned above, the study findings refer to recommender systems in general. For measuring the impact of credibility on preferences for a specific source, such as a destination recommender system, it might be necessary to specify the type of advice (i.e., destination recommendation). Also, given White’s (2005) findings that emotional decisions might lead to different source preferences and given the fact that travel and tourism products are often high in emotional content, specific references to the decision-making context need to be made in future studies in order to truly test the relative importance of expertise versus trustworthiness in determining preferences for travel recommendation sources.

As far as the design of recommender systems is concerned, the results of the study call for consideration of credibility as it drives a user’s likelihood to accept advice from the system. Future research should focus on identifying specific cues in every aspect of recommender system design (interface, presentation of recommendations, etc.) that most strongly increase perceptions of credibility. General trust cues such as quality seals (see Fogg, 2003) appear to be rather unsuitable for the context of recommender systems. Further, addressing gender-specific differences appears to be an important strategy for recommender systems to become preferred sources for travel advice. Ratings of trustworthiness and expertise were generally low for both males and females. Because women are often the main information seekers and decision makers for travel-related decisions (Austrian Tourist Office, 2006), designing systems that cater specifically to females could be a possible design strategy. On the other hand, males currently do not have favorable attitudes toward recommender systems and need to be convinced that they are trustworthy recommendation sources.

Recommender systems will become ever more important in providing tourists with intelligent recommendations for various travel products (Ricci, 2002). It is crucial for effective recommender system design to consider those factors that have an influence on a user’s propensity to accept recommendations from a specific system. Recommender system research has so far only focused on determining the persuasiveness of recommendations and has largely neglected aspects of the recommender system as a source of advice. Credibility seems to be an important concept to consider; however, there might be other factors that could potentially increase the persuasiveness of recommender systems. These factors need to be identified and integrated into a persuasive framework for travel recommender system design.
Biographical Notes

Kyung Hyan Yoo is a Ph.D. student in the Department of Recreation, Park & Tourism Sciences at Texas A&M University and a research assistant for the Laboratory for Intelligent Systems in Tourism. Her research interests include online tourist information search and decision making, persuasive technology, word of mouth, destination recommender systems, and source factors in online advice seeking relationships.

Ulrike Gretzel is an Assistant Professor in the Department of Recreation, Park & Tourism Sciences at Texas A&M University and Director of the Laboratory for Intelligent Systems in Tourism. She received her Ph.D. in Communications from the University of Illinois and holds a master’s degree in International Business from the Vienna University of Economics and Business Administration. Her research focuses on persuasion in human–computer interactions and the communication of sensory experiences in online environments.

References


Franke, T. (2002). Integration of group support tools into an online tourism advising system. eCTRL Workshop on User Modeling and Decision Making in Travel and


