Studying the Effect of Similarity in Online Task-Focused Interactions

Dan Cosley, Pamela Ludford, Loren Terveen
Department of Computer Science and Engineering
University of Minnesota
4-192 EE/CS Building, 200 Union St SE
Minneapolis, MN 55455
{cosley, ludford, terveen}@cs.umn.edu

ABSTRACT
Although the Internet provides powerful tools for social interactions, many tasks—for example, information-seeking—are undertaken as solitary activities. Information seekers are unaware of the invisible crowd traveling in parallel to their course through the information landscape. Social navigation systems attempt to make the invisible crowd visible, while social recommender systems try to introduce people directly. However, it is not clear whether users desire or will respond to social cues indicating the presence of other people when they are focused on a task. To investigate this issue, we created an online game-playing task and paired subjects to perform the task based on their responses to a short survey about demographics and interests. We studied how these factors influence task outcomes, the interaction process, and attitudes towards one's partner. We found that demographic similarity affected how people interact with each other, even though this information was not explicit, while similarities or differences in task-relevant interests did not. Our findings suggest guidelines for developing social recommender systems and show the need for further research into conditions that will help such systems succeed.

Categories and Subject Descriptors
H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces---Computer-supported cooperative work, synchronous interaction; H.5.2 [Information Interfaces and Presentation]: User Interfaces---Evaluation/methodology; H.1.2 [Models and Principles]: User/Machine Systems---Human factors

General Terms
Measurement, Design, Experimentation, Human Factors

Keywords
Social navigation, recommender systems, demographics, similarity, friendship, community, matchmaking

GROUP'03, November 9–12, 2003, Sanibel Island, Florida, USA.

1. INTRODUCTION
In *Bowling Alone* (2000), Robert Putnam argues Americans have become less socially involved, and there is evidence that this is true throughout the developed world. He claims that computer use has contributed to significant declines in community participation.

The computer itself, however, is a neutral tool. For example, the Internet provides opportunities for social interaction via email, instant messaging, chat rooms, bulletin boards, and multiplayer online gaming. People expect to have social interactions when using these technologies, and behavioral norms have emerged surrounding their use. Acceptable behaviors vary by application and often by user community.

Other Internet tasks are generally perceived to be solitary in nature. People don't expect to encounter others when checking their library accounts, using Google to find research papers, scanning newspaper headlines, or buying camcorders. Yet at any given time, others are visiting the same web sites and performing the same tasks: how many people are reading cnn.com right now? These hidden crowds—people whose online "daily rounds" overlap—represent both a resource and an opportunity.

This phenomenon raises a key question: would it be useful to make the hidden crowds visible? By helping people realize that others who might share their interests are "nearby", computers could help people form new social bonds. Computers might assist people in forming both weak bonds, such as knowing an expert in a topic of interest, or strong bonds, such as lasting friendships.

1.1 Our question: what would make a social recommender work?
Using the hidden crowds to help people connect is not a new idea. Social navigation helps reveal the presence of others by providing representations of fellow participants and their activities. Online gaming sites, for instance, allow users see how many people are playing, discover whether other players are free or active, and browse game history and personal information revealed by others.

Social navigation systems typically provide visualizations of the aggregate activity of users of a system (e.g., visitors to a site) to aid decision-making, usually the traversal of a complex information space. In contrast, "social recommender" systems explicitly match pairs or groups of people to facilitate collaboration or social interaction. Matchmaking sites are a canonical example, asking users a series of questions and then introducing them to other users who give similar responses.

This paper focuses on social recommender systems that use similarity to select others to recommend. We concentrate on similarity because it has several desirable properties. Behavioral science studies show that in everyday life, people choose friends with similar age, income, gender, marital status, and/or ethnicity, and that similarity of interests is an important factor in liking others. Similarity also is computationally attractive. Besides directly asking users for information, software can observe users—for example, the documents they read or the music they listen to—and automatically build interest profiles. Then techniques from information retrieval and machine learning can be used to match people with similar profiles. This allows designers to build social recommenders that operate unobtrusively in a number of situations, for instance, recommending others interested in topics that a user is researching.

But—is the promise of social recommenders real? Will people be interested in meeting others while engaged in online information-seeking tasks? On a visit to Manhattan, you might pass 10,000 strangers. How many of them would you want to talk to? Would they reply? Will the conversation be any more than "where's a good place to eat?" For a social recommender to succeed, it must overcome these barriers: people must be willing to meet others while in an information-seeking mode.

Similarity is also not a panacea. We don't know if similarity between people engaged in solitary tasks online will play the same role in fostering friendships as it does in other settings. Even if a social recommender succeeds at pairing individuals based on similarity, it may fail on other levels. For example, in the course of performing tasks, teams often benefit by having a diverse set of skills and characteristics. Building groups tightly connected by interest similarity may harm their long-term success. More globally, using similarity to introduce people could induce "balkanization", a fracturing of community into small, disconnected components.

Instead of building a system for recommending people and seeing what happens, we believe it makes sense to first evaluate the conditions necessary for such a system to succeed. In this paper, we study how similarity affects the way people perceive others online. In particular, we investigate the effect of similarity on two people engaged in a task, when the similarity is left implicit and the communication bandwidth is limited. We believe these conditions represent those a social recommender would need to deal with, now and in the near future.

After surveying related work, we describe our experiment. In the experiment, 45 pairs of subjects played a team question-answering game. We found people preferred others with demographics similar to their own, and, to our surprise, task-relevant interests did not have a significant effect. We then discuss the implications of our results, particularly for the design of social recommender systems, and where such systems might go from here.

2. RELATED WORK

A number of researchers have created systems aimed at making people aware of the hidden crowds online. Wexelblat and Maes (1999) built the Footprints system, which helps people visualize and exploit others' histories of interaction with web pages. Svensson et al. (2001) created a site for recipe exchange and grocery shopping where any user's activity was visible to other users. Livemaps (Cohen et al., 2002) let users see other web surfers, grouped by topic. I2I (Budzik et al., 2002) is similar to Livemaps in that it groups web users by topic, allowing them to leave "calling cards" when they are interested in a topic and making these cards visible to other users who are browsing related web pages. Our research can inform such system-building efforts, identifying factors they can use to increase their chances of success. Specifically, we are investigating how various types of similarity (of interests or on demographic factors) influence people's willingness to interact with others, their attitude towards those they're put into contact with, and the success of these opportunistic interactions.

Similarity-based social recommenders are a kind of recommender system (Resnick et al., 1994). These systems garner users' opinions about items in a domain, such as their ratings of movies. The system recommends other items that the users might like based on similarities between the users' opinions. Yenta (Foner 1996) is a social recommender system that explicitly attempts to recommend people instead of things, using a similarity-based approach. Expertise Recommender (McDonald & Ackerman, 2000) and ReferralWeb (Kautz et al., 1997) also recommend people and use social network information in deciding whom to recommend to a particular user. While most research in recommender systems has focused on developing effective algorithms, a few researchers have studied the problem of how to present recommendations so that users can make better-informed decisions (Herlocker, 2000; Swearingen and Sinha, 2002). We hope that studying the role of similarity in online interactions will help us design appropriate interfaces to help users evaluate social recommendations.

User modeling research (Rich 1979, Kobsa 2001) aims to create models of users in order to adapt system interaction to each user, tailoring the functionality that the system provides or customizing the information that is presented. Issues of interest for our work include what is represented in a model, how it is represented, and how it is built. For our current purpose, we directly ask users for demographic information and for their interest in subject areas. We expect that most social recommender systems will choose instead to infer users' interests from their behavior, e.g., from the web pages they read or the music they listen to.

Understanding the factors that lead individuals to like each other is a core issue for the social sciences. For example, social psychologists have found that people typically like others who are similar to themselves, who are good-looking, intelligent, or have other positive social attributes, who have provided them favors, and with whom they have a history of interaction (Berscheid & Reis, 1998 gives an extensive overview of the basis of interpersonal attraction). Sociologists have shown friends are likely to be quite similar to each other on a range of demographic factors, such as ethnicity, income, education level, religion, and profession (e.g., Fischer 1977, Fischer et al. 1982, Verbrugge 1977). Similarity is a major factor in two phases of relationship formation, which Verbrugge called "meeting" and "mating". First, people who encounter each other in their daily rounds are likely to be quite similar—in effect, social contexts such as places of work, commuter trains, churches, and community groups serve as filters. From this pool of already similar candidates, people tend to form friendships with those others who most share their values and interests. Another interesting result is that the context in which a friendship is formed—for example, at work vs. the community one grew up in—influences the dimensions along which similarity is highest. For example, friends made at work are more similar in their occupation level and economic sector than on ethnicity (Jackson 1977).

This work is a basic foundation for our research. Questions it raises for us include: Which dimensions of similarity matter for online interactions? Jackson's work suggests that the characteristics, which matter in an online context, may be different from those, which matter in offline interactions. Which characteristics do people become aware of through typical online interaction when explicit user profiles are absent? According to the classic New Yorker cartoon, "On the Internet, no one knows you're a dog." We ask whether that is true, or whether people soon figure out that you're a dog. Do they even care, if they are busy with their own concerns? And—one way or another—do the dogs form themselves into packs?

These questions point us to additional work. Van Alstyne & Brynjolfsson (1996) carried out a theoretical analysis of the prospects for online communities to "balkanize", i.e., to self-segregate into tightly defined communities of interest. This work relates to ours in a number of ways. First, we represent user interest in a set of topics as they did, using a numeric vector where element *i* represents a user's degree of interest in topic *i*. Second, we investigate a major possible consequence of balkanization: namely, whether groups composed of individuals with similar interests perform more poorly on a range of tasks that require diverse knowledge. A third relevant issue is one we already have touched on—to what extent does balkanization occur in online behavior? That is, to what extent do people segregate themselves into narrow communities of interest? If it turns out that people do prefer others who share their interests, it supports the hypothesis that communities will, in fact, balkanize.

Research in organizational behavior enriches our perspective with its focus on working relationships and work teams. Since working relationships exist to accomplish tasks, factors such as skills and successful task outcomes are more important, while disclosure of personal information is less important (Gabarro 1990, p. 79). However, task-centered relationships exert their own influence on relationship development. Completing a task successfully can lead team members to like each other more (Farris & Lim 1969) and cause greater satisfaction and team cohesion (Staw 1975). Further, there is much evidence that the "more similar two people are in background and attitudes, the easier and more satisfying a task-based relationship will become" (Gabarro, p. 102). However, there is a potential problem: a balkanization argument suggests that if teams are too similar, their knowledge, skills, and perspectives will be too limited, causing task outcomes to suffer. Taking a broader perspective, in many nations diversity of work groups is not just a goal, but a fact (Ayman 2000). Balancing similarity and diversity in relationships that have both task-oriented and social components may be a complex issue. We consider both demographic factors (age, education, gender, etc.) and task-relevant knowledge, and investigate the conditions under which successful task completion and the development of satisfying personal relationships can go hand in hand.

The effect of various novel technologies on small group interactions is a core CSCW concern, and much prior research in the area has influenced us. Bradner & Mark (2002) recently carried out a noteworthy study. Subjects were given a set of tasks to carry out with a partner whom they had never met, interacting through either text messaging or video conferencing. The partner actually was a confederate of the researchers who followed an interaction script. In addition to the medium of communication, the main variable manipulated was perceived distance between the subject and the confederate. The confederate actually was in the same building as the subjects, but half the subjects were told their partner was in the same city, and the other half were told she was in a remote city. The experimental tasks gave subjects the opportunity to persuade each other to change their minds and to be relatively honest or deceptive with each other. The results showed that subjects were less deceptive with and more persuaded by their partner when they thought she was located in the same city. On the surface, our research differs significantly: we are studying similarity, not distance, and our experimental tasks and design are rather different. However, we are interested in the same core issue: when people are online, they bring along a set of social norms and behaviors formed from their prior experiences. We believe it is important to understand how these norms and behaviors adapt to online communication, in order to build more effective social technologies. How people react to similarity in others is one aspect of this much larger problem.

3. EXPERIMENTAL DESIGN

We now turn to the design of our current experiment, first describing the task we asked subjects to perform, then outlining the hypotheses we investigated.

3.1 Let's play… the Family Feud!

We faced a number of constraints when designing the task. First, we needed to create a task with a measurable outcome that was suitable for two people to perform as a team. We wanted the task to encourage cooperation, teamwork, and discussion, while incorporating real-world CSCW activities such as brainstorming and decision-making. The task also had to be something people actually do online. We wanted the task to be primary focus, not the partner. Finally, the task had to be engaging and short enough so users would be willing to participate.

We eventually settled on a team online game as a way to meet our constraints. We designed a game based on the American television show *Family Feud*. Prior to the game, the producers conduct surveys, asking people to respond to questions such as "Name an item of clothing worn by the Three Musketeers." Participants on the show guess the most popular responses, scoring points based on the popularity of their choices. Prior to the experiment, we posted a survey consisting of 73 questions on our web site, advertised it in various online forums, and received responses from 227 people. We asked questions in six knowledge categories inspired by popular trivia games. Some questions turned out to be unsuitable for use in the game, as everyone agreed (e.g., "Name a band from the Sixties."[1]) or had a different opinion (e.g., "Name your favorite movie."). From each category, we chose one question with a distribution of responses we thought would encourage brainstorming and discussion. Table 1 shows the categories and the questions we chose for the game.

[1] The Beatles, of course.

Table 1. Topic categories and questions for the experiment.

Category	Question
Popular (U.S.) Culture	Name a TV comedy show that ended before 2000.
Science and Nature	What is the most important discovery of the 20th century?
U.S. History	Name a famous woman in U.S. history.
Geography	Name a country formed from the former Soviet Union.
Sports	Name a track and field athlete.
Arts & Entertainment	Name a book written by Charles Dickens.

The interface (see Figure 1) showed players a question and asked them to guess the three most popular responses. A basic chat interface allowed partners to communicate. One player, designated as the “captain”, was responsible for entering the team’s answers and submitting them after the team reached agreement. The captaincy alternated between the players every round in order to encourage cooperation. Once the captain submitted answers, the team received points for guessing popular responses, with bonus points for getting the top response or all three responses in order. The game consisted of one practice round and five scoring rounds.

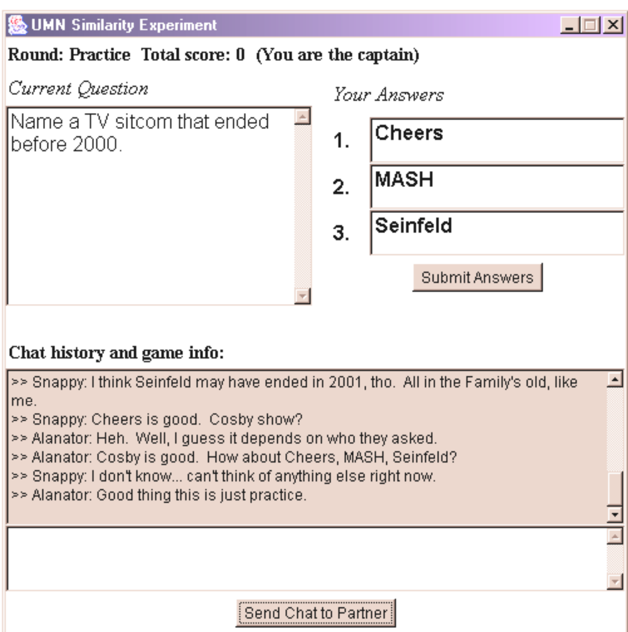

Figure 1. The experimental interface. The top portion of the interface supports answering Family Feud-style questions; the bottom portion allowed teammates to communicate.

3.2 Subjects

We recruited subjects from a number of online forums. Most came from one of two sites: a movie recommendation web site and a site devoted to finding bargains on the Internet. Both sites allowed us to recruit users who were actually engaged in an information-seeking task. The link took people to the experimental site, which provided information about the experiment and listed specific times when the game would be played. Subjects had to return to the site at a scheduled game time. We offered subjects a $5 gift certificate from amazon.com as compensation for completing the experiment. We thought that offering a prize for good performance would encourage people to focus on the task, so we promised $20 to the top six teams as an additional incentive.

When subjects returned to the site for a game, they first took a pre-survey. This survey asked subjects for demographic information (using questions based on the U.S. Census) and ratings of their interest in each of the six topic categories. After subjects completed the pre-survey, they received instructions and played a single-player practice version of the game.

Once enough subjects arrived at the site, the system paired them into teams. We tried to create teams with either very similar or very different interests. We were interested in extreme values for similarity because we believe that a real social recommender will most likely recommend the most similar people. As mentioned earlier, we represented subjects’ self-ratings of interest in our topic categories as a vector in a 6-dimensional space. We computed similarity between subjects by using cosine correlation (Salton & McGill, 1983). Cosine correlation is commonly used in information retrieval to evaluate similarity between queries and documents; the similarity of two vectors increases as the cosine of the angle between them decreases.

We asked each team to start by chatting for a few minutes, then to play the game. We imposed no time limit on the game, although the instructions stated that the experiment would likely take 45 to 60 minutes to complete. Upon completing the game, players were instructed to say goodbye to their teammate, then completed a post-game survey about their reactions to their partner and to the game. The survey questions asked subjects how well they iked their partner (e.g., “My partner and I had a lot in common”) and how well they collaborated with their partner (e.g., “My partner and I were able to agree on answers easily”).

3.3 Hypotheses

We hypothesized both demographic and interest similarity would be visible in thin online interactions (people *can* tell you’re a dog). We also believed people would prefer more similar partners.

- *H1d. People prefer to interact with demographically similar partners.*
- *H1i. People prefer to interact with people who have similar interests.*

Organizational psychology research and balkanization principles suggest teams achieve superior outcomes when team members have a variety of skills. Therefore, we believed teams with more diverse knowledge would perform better.

- *H2d. Teams perform better when users are less similar demographically.*
- *H2i. Teams perform better when users have less similar interests.*

We expected that both doing well and liking one's partner would affect how well people collaborated with their partner. If our hypotheses above are true, then similarity should have both a positive and a negative effect on how well people collaborate. These effects would tend to cancel each other out, so we adopted null hypotheses about how similarity affects collaboration.

- *H3d. Demographic similarity has no effect on quality of collaboration.*
- *H3i. Similarity of interests has no effect on quality of collaboration.*

3.4 Measures
The pre-survey responses provided demographic information and subject self-ratings of interest in topic categories. We built regression models to study the effect of demographic variables on subjects' evaluation of their partners. We also correlated similarity of interests with game outcomes and subjects' evaluations of their partners. Finally, we used subjects' text comments about the game, along with content analysis of the conversations between partners, to help us understand and interpret the results of the surveys. In content analysis, we grouped conversation into three broad classes: social (general social statements plus exchanges of personal information), game play (suggestions or discussion of answers), and meta-game (statements about the way the game worked, as well as post-mortem discussion of their performance on a question).

4. RESULTS
A total of 90 subjects (41 male, 49 female) participated in the experiment. Subjects were moderately diverse in their demographic features. Most were between the ages of 18 and 64. Subjects were evenly distributed across income levels from $10,000 to $75,000+. All but eight had attended college. 30 had attended without earning a degree, 8 had received associate degrees, 21 had earned bachelors degrees and 23 had received graduate or professional degrees. Almost all subjects were white, and all subjects lived in either the United States or Canada. Two pairs of subjects had talked online with each other before. When subjects rated their interest in the subject categories, most rated themselves modestly, claiming their interest was slightly above average (except for sports, which was slightly below average).

Table 2. Descriptive statistics for communication between partners, game performance, and game time. Some teams played quick games while others were much more social.

	Mean	Median	Minimum	Maximum
Team gameplay turns	75	67	13	237
Team meta turns	27	23	1	140
Team social turns	25	19	4	88
Team score	196	200	71	293
Game time (minutes)	27	23	7	68

To give an idea of the general flavor of interactions during the game, Table 2 shows mean, median, low, and high values for game time, score, and types of conversation. Games ranged from quick events with few suggestions and almost no socializing to much longer, conversational affairs. In general, communication was balanced, with neither partner dominating the conversation.

4.1 Demographic similarity affects measures of liking
We found demographic similarity affected how well partners liked each other. As summarized in Table 3, when teammates were similar in education level or gender, they conversed more overall, talked more about social topics, and played the game longer.

Table 3. Similarity of gender and education led to increased interaction.

	Gender	Education
Overall conversation		Similar
More social exchanges	Similar	
Game length		Similar
Personal info exchanges	Similar	

Gender. Same-gender pairs interacted more than mixed gender teams. Thirteen of the teams consisted of two male players, fifteen of a male and female, and seventeen consisted of two female players. Figure 2 summarizes these results.

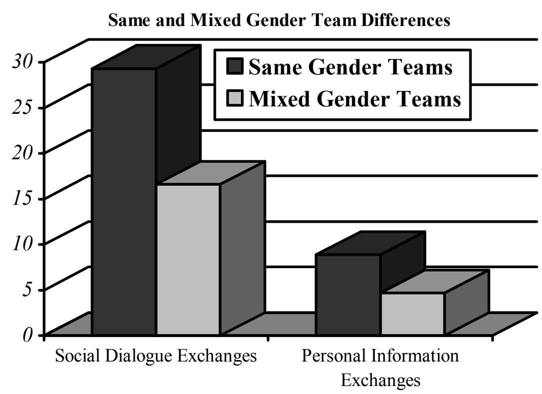

Figure 2. Differences between the amount of information exchanged by same-gender and mixed-gender teams. Same-gender teams tended to have more social exchanges, including explicit requests for personal information.

Chat content analysis revealed same gender teams engaged in more social dialogue than mixed gender teams, with subjects in same gender pairs contributing 29.33 social turns per game versus 16.6 for subjects in male/female pairs (T-test, $p<0.03$, $t=2.32$, $df=43$). Same-gender teams also exchanged more personal information than mixed-gender teams. We counted the number of times subjects asked, offered, or inferred information about their geographic location, gender, age, subject knowledge, or other personal information during the game. Same gender teams exchanged personal information an average of 8.9 times per game, while mixed gender teams averaged 4.7 exchanges. The difference was significant (T-test, $p<0.05$, $t=2.03$, $df=43$).

We believe that gender affected the way teammates interacted because it is often explicitly visible. Players often (39 of 90) revealed their gender through their nickname. Choosing a gender-specific pseudonym could have been a subconscious effort by subjects to protect themselves from consequences of defying gender-based interaction norms (Jaffe et al., 1995). In addition, while playing the game, three people explicitly asked their partner for gender information, with six volunteering it.

Education. Subjects also interacted more with partners who had an educational background similar to their own. When partners had similar levels of education, they talked more overall than pairs with less similar educational backgrounds (one-way ANOVA, $p<0.04$, $F=3.53$). Teammates with similar educational status also played the game longer than those who had dissimilar backgrounds (one-way ANOVA, $p<0.02$, $F=3.69$). This makes sense, since more conversation would lead to longer game times. Unlike with gender, however, we did not find that the quantity of social exchanges increased with educational similarity.

We wondered whether subjects explicitly mentioned their educational status, or if educational similarity was revealed implicitly as partners played the game. Only six subjects shared their educational status; subjects who did not share their status showed the same effect of similarity as those who did. Burgoon and Miller (1987) found individuals use a conversant's language choice to form attributions about their education level. Thus, subjects' level of education was likely reflected in properties of the interaction perceived by their partners. This shows that some demographic features do "come through" in online interaction and shape its character, just as they do in face-to-face interaction.

Other demographic factors. We collected additional demographic data about test subjects, including age, racial/ethnic background, income level, and location.

We found a trend relating number of social interactions and age similarity. When team members were in a similar age range, they tended to participate in more social conversation than those who were not. (T-test, $p=0.055$). Subjects were curious about their partner's age, as well. Eight explicitly asked their partner for age information, 19 provided their age, and four more made inferences about their age while playing the game.

We could not study the effect of race and ethnicity, as our subjects were largely homogenous. We did not find any correlation between income level and interaction preferences. We also tested whether geographic distance between players affected their interest in each other; it did not. We note that none of the teams included two subjects who lived in the same metropolitan area or locale. Bradner & Mark (2002) found that perceived distance can influence interaction, so this issue deserves further study. It is interesting to note subjects exchanged location information more often than any other type of personal information (26 asked, 40 provided).

Task-relevant interests. Our second hypothesis about liking was that subjects would prefer to interact with others who had similar task-relevant interests (H1i). Our metric for similarity of task-relevant interests was the similarity we computed between partners' interest in our topic categories. This metric did not correlate with overall conversation, social conversation, or game length, and thus did not predict how well two subjects would like each other.

4.2 Task performance and similarity

We had also hypothesized (H2i) that similarity of interests would negatively affect a team's performance in the game, since we thought doing well at the game would require interest and/or skill in the various topic categories. This was not the case. Overall score was independent of computed similarity between partners, being predicted instead by the team's education level and the amount of effort they dedicated to brainstorming and discussing answers to trivia questions. Interest in a given subject area did not predict scores on questions in that subject area. So, self-rated interest did not correlate with ability to predict popular answers to trivia questions on a given topic.

Our other hypothesis about performance was that teams with diverse demographics would achieve higher game scores (H2d), which also turned out not to be the case. This may have been because knowledge leading to correct answers did not favor individuals from a given demographic group. For example, we might speculate that younger subjects might do better on a question such as "Name a show on the WB[2]".

4.3 Similarity and quality of collaboration

We now turn to the effect of similarity on the quality of collaboration. We had adopted a null hypothesis suggesting demographic similarity would not affect how partners collaborated (H3d). We expected tension between liking one's partner and doing well on the task would negate each other, and believed ability to collaborate would be unaffected by similarity. Our results suggest similarity in education or age improves the quality of collaboration, while being the same gender lowers it.

Education. Teams with similar levels of education had more brainstorming conversational turns (T-test, $p<0.02$, $t=2.28$, $df=88$). In addition, teams with higher education levels overall dedicated more time to suggesting and discussing answers than those who were less educated.

Age. Teammates of similar ages were more likely to have discussion about the game, 19.14 turns on average, than teammates who were not in the same age demographic, who averaged 12.14 turns (T-test, $p=0.02$, $t=2.34$, $df=88$). This discussion included comments about how the game worked, reactions to their performance, and comments on answers other people had chosen.

Gender. Mixed gender teams showed signs of collaborating better than same gender teams. According to the post-survey, players on mixed-gender teams rated their partner's trivia knowledge higher than individuals on same-gender teams did (T-test, $p<0.03$, $t=2.26$, $df=88$). We observe although same-gender teams were more willing to exchange social information, mixed-gender teams had a higher opinion of their partner as a teammate.

Task-relevant interests. As with liking and task performance, similarity of interests did not affect the quality of collaboration (H3i). We believe the game questions may have been too easy for the level of interest to make a difference: even subjects who said

[2] The WB is a United States television network whose shows are aimed at teenagers and young adults.

in chat that they had limited interest or knowledge in a given subject area were usually able to contribute answers and comments to discussion of questions in that area. Additionally, subjects might have misunderstood the pre-survey. Although subjects were asked to rate their interest level in the topic area, they might have instead rated their knowledge level on the topic.

5. DISCUSSION

Table 4 summarizes our findings. Demographic similarity affected how well people like and collaborate with each other, while similarity of interests did not. Neither sort of similarity affected the outcome of the game. We now discuss possible reasons for these results and suggest implications.

Table 4. Summary results for our hypotheses. Demographic similarity affected both liking of and collaboration with one's partner, while similarity of interest had no effect.

Hypothesis	Supported
People prefer to interact with demographically similar partners.	Yes
People prefer to interact with people who have similar interests.	No
Teams perform better when users are less similar demographically.	No
Teams perform better when users have less similar interests.	No
Demographic similarity has no effect on quality of collaboration.	No
Similarity of interests has no effect on quality of collaboration.	Yes

We start by discussing whether our task effectively measured behavior in the way we had anticipated. We had a number of requirements. Some were easy to meet, e.g., designing a two-player task with a measurable outcome. Others included making the task acceptable to subjects, encouraging conversation, incorporating elements of real CSCW tasks, and having users focus on the task rather than the social aspects.

The task met these goals. Most subjects commented that the task was fun and engaging, even if they did not perform well relative to other teams. No one dropped out during the game—in fact, a number of subjects wanted to play multiple times. Most subjects also said they were able to effectively communicate with their partner. Communication, in turn, was an important component of success at the task; teams that brainstormed and discussed more ideas tended to do better (linear regression, r^2=0.38). This also suggests that we met our goal of representing CSCW skills such as brainstorming and consensus formation. People did appear to be focused on the task, and were more likely than not to say winning the game was important.

5.1 You're a dog, aren't you? Me, too.

We next discuss why subjects preferred to interact with demographically similar partners, even when their attention was focused on performing a task and the similarity was not made explicit.

An advantage often claimed for text-based computer mediated communication is that people can present themselves with a "face" or persona (Turkle 1995) that obscures basic characteristics like gender, race, and age that are immediately apparent in face-to-face interaction. Some information seems to leak through, however. Choice of nickname often reveals gender, or at least the gender someone is pretending to be. Educational level appears to show through in one's conversational style. In our experiment, it appears that people can often tell that you are, in fact, a dog.

Further, dogs form packs: people preferred others like themselves. We found it interesting that people actively asked for and offered personal information, even though it wasn't required for the task. On average, teams had 7.5 such exchanges (asking for, giving, or inferring a piece of personal information). This behavior implies people are interested in learning about unseen others in the online world. One subject even jokingly wondered whether his partner was human, and he and his teammate conducted a mock Turing Test on each other.

In addition to explicitly revealing otherwise invisible demographic information in chat, people often make personal profiles and web pages available to others they meet. People also seek information about others they have met or may meet. Subjects sometimes asked for information directly, and other times mentioned inferring it from the conversation. More generally, searching for information about someone you encounter online is so common that it has a special term: "Googling" someone means to feed his or her name to a search engine.

These behaviors suggest that when people meet online they actively seek to gather, and are willing to provide, much of the information that would be apparent in face-to-face interaction[3]. Apparently, they consider this information useful in evaluating and relating to others. This is consistent with communication theories such as Social Information Processing Theory (Walther 1992) and Uncertainty Reduction Theory (Lawrence & Mongeau 1996), which emphasize the need for people to reduce uncertainty about their communication partners. We contend that social recommenders will have the best chance to succeed if they provide as much information as possible about the people recommended. As we mentioned earlier, users of recommender systems want more information about a recommended item than just the systems' prediction of how much they'll like it. Such information helps users evaluate the recommendations. We expect this to be even truer for systems that attempt to bring people together, since choosing to interact with someone has more serious implications than choosing a song to listen to. McDonald's work in expertise recommendation (McDonald & Ackerman, 2000) offers one approach: by letting users filter a recommended set of experts with a social network, it enables them to locate experts who have some sort of relationship to the user.

People often claim to want to guard their personal information closely, which presents an interesting design challenge because of the tradeoff between the desire for privacy and the benefits to be gained from revealing personal information. The situation is complicated by that fact that different contexts call for differing

[3] One exception in our study was race: people never asked their partner's race, nor volunteered their own.

levels of disclosure. The success of social recommenders will depend on designers' ability to bring together people willing to show their online faces, rather than use them as masks. Understanding how to do this provides a fertile area for future research (Lederer et al., 2002; Palen and Dourish, 2003).

5.2 Why didn't interest similarity matter?

We found our measure of interest similarity did not affect how much subjects liked their partners, how well subjects collaborated, or how well they performed the task. In retrospect, we see our task did not measure this effect in the way we had anticipated. With Family Feud-style questions, interest in an area did not turn out to predict performance in the area. Most teams were able to generate multiple responses for each question, even when both players claimed in chat and in self-rating of interest that they were not good in the topic area. Success in the task was also only partially based on knowledge of the topic; it also depended on people's ability to guess how others would respond.

We had considered a task that required more expertise, such as answering trivia questions. However, we were afraid that such a task would devolve into a test of skill in using search engines. This fear was well founded, since several players said that they were doing research on the Internet during the game.

Another possible reason we saw no effects of interest similarity is that our notion of interest may have been too broad. In a world of increasing specialization, it may be necessary to go deep into specifics before finding areas of interest that help people connect. Not "sports", not "football", not "English football", but "Manchester United" might be the granularity at which systems will need to detect similarity.

This is a problem for social recommender systems. In our experiment, we considered asking people to answer a number of trivia questions and building keyword profiles based on the results, but rejected that approach as too time-consuming for subjects. Real recommenders will also need to carefully consider how much of a burden to impose on users—another point in favor of automatically extracting profiles from observing users' behavior. Hopefully, such approaches will allow systems to use finer-grained similarity than we were able to capture, and this finer similarity will lead to positive results.

Of course, finer-grained similarity is more likely than coarse similarity to lead to balkanization issues. In the limit, like the apocryphal PhD student who knows everything about nothing, fine-grained similarity could lead to special interest groups of size one. Finding the right balance between social needs and system success would be an interesting issue to pursue.

Finally, one might conclude that people simply won't respond to interests they share with others while engaged in solitary tasks. We don't agree with this conclusion. Interests, ideas, and opinions are a primary currency on the Internet. Communities on the Internet often form around shared interests, and it's only a mild stretch to claim that interest on the Internet plays an analogous role to location in physical space[4]. Since location is a key element in the first, "meeting" step of Verbrugge's two-step process of forming bonds with others offline, we might expect interests to play a similar role in online interactions. In fact, the model online might be exactly backwards from Verbrugge's. Instead of filtering people by demographics and then using shared interests to select who to interact with, it may be that people online will filter by interests and then gravitate toward those with similar demographics and backgrounds.

It is an open question how best to use similarity. Our results suggest that recommendations that take into account both interest and demographic similarity will do better than recommendations based on similarity of interest alone. Sensitivity to the user's current context and goals when making recommendations should also improve the success of social recommenders.

The best method for making the introductions is also an open question. One way would be to explicitly generate recommendations for immediate, synchronous conversations. Another is to recommend asynchronous events where like-minded people can meet. For example, a system could observe that a number of Pittsburgh Steelers fans were online and schedule a chat for later that day. Finally, a system could make others visible in a non-intrusive way, opportunistically capturing users' attention when they pause in their current tasks while presenting enough information to help users decide whom to contact. All of these approaches are interesting, and worthy of further study.

6. CONCLUSIONS AND FUTURE WORK

People are sensitive to the demographic traits of others, even when encountering them online during the course of a task, and even when these characteristics are not explicit. Further, in our task, demographic similarity did not directly affect task performance. These results can and should be verified in other settings and exploited by designers of CSCW systems in general and designers of social recommenders in particular. Using demographic similarity in building teams and other groups online appears to be a promising approach.

Despite our failure to observe significant results in our task, using similarity of interest as a strategy for introducing people seems promising. We envision a number of approaches for studying how best to use similarity. For example, we could follow the approach of Bradner and Mark (2002), which studied the effect of proximity, but substituting a measure of interest similarity.

Another area of research is how—and how much—similarity information to reveal. In our experiment, we chose to reveal no similarity information directly. Another possibility would be to study the effect of providing varying levels of information in a profile of the partner. An interesting alternative would be a "good host" interface that picks a specific point of similarity as a starting point to encourage the conversation. Interfaces that allow people to control how much to reveal—managing their online face—while satisfying their needs to discover information about others is another fertile area for research.

Bringing people together and helping them build relationships in the course of their online experience is an opportunity for computer science to make a real difference in the quality of people's lives. We share the excitement of those who have already built systems that strive toward this goal. By carefully studying the conditions that foster the formation of social bonds online, we hope to progress toward a world where the people on the screen are just as real, important, and valued as people nearby.

[4] The work of Harrison & Dourish (1996) provides a useful starting point for developing this theme.

7. ACKNOWLEDGMENTS

We thank Greg Grossmeier and Mike Cassano for their help in building the software used to conduct the experiment, and thank Greg, Mike, and Sean McNee for their comments on drafts of the paper. This work was funded by a University of Minnesota Grant-In-Aid, "Facilitating Participation in Online Communities".

8. REFERENCES

[1] Ayman, R. (2000): 'Impact of team diversity on collaboration dynamics', in *Collaborating Across Professional Boundaries*, available at http://www.stuart.iit.edu/ipro/papers/pdf/ayman.pdf.

[2] Berscheid, E., and Reis, H. T. (1998): Attraction and close relationships. In D. T. Gilbert & S. T. Fiske et al. (Eds.), *The handbook of social psychology, Vol 2* (4th ed.), New York, NY, US: McGraw-Hill, pp. 193-281.

[3] Bradner, E., and Mark, G. (2002): 'Why distance matters: effects on cooperation, persuasion and deception', in *Proceedings of CSCW 2002*, ACM Press, 2002, pp. 226-235.

[4] Budzik, J., Bradshaw, S., Fu, X., and Hammond, K. J. (2002): 'Clustering for Opportunistic Communication', in *Proceedings of WWW 2002*. ACM Press, 2002.

[5] Burgoon, J. K., Bengtsson, B., Bonito, J. A., Ramirez, A., & Dunbar, N. E. (1999): 'Designing interfaces to maximize the quality of collaborative work', in *Proceedings of HICSS 1999*, IEEE, Maui, HI, 1999.

[6] Burgoon, M., and Miller, G. R. (1987). An expectancy interpretation of language and persuasion. In H. Giles & R. N. St. Clair (Eds.), *Recent advances in language, communication, and social psychology*. Hillsdale, NJ, Lawrence Erlbaum, 1987, pp. 199-229.

[7] Cohen, D., Jacovi, M., Maarek, Y. S., and Soroka, V. (2002): 'Livemaps for collection awareness', *International Journal of Human-Computer Studies*, vol. 56, no. 1., 2002, pp. 7-23.

[8] Farris, G. F., & Lim F. G., Jr. (1969): Effects of Performance on Leadership, Cohesiveness, Influence, Satisfaction, and Subsequent Performance, *Journal of Applied Psychology*, vol. 53, pp. 490-497.

[9] Fischer, C., et al. (1977): *Networks and Places: Social Relations in the Urban Setting*. (with Jackson, Stueve, Gerson, Jones, and Baldassare). New York: Free Press.

[10] Foner, L. (1996): A Multi-Agent Referral System for Matchmaking, in *Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, London, April 1996, pp. 245-261.

[11] Gabarro, J. J. (1990): The Development of Working Relationships, in Galegher, J., Kraut, R.E., and Egido, C. (Eds.), *Intellectual Teamwork*, pp. 79-110, Lawrence Erlbaum Associates.

[12] Fischer, C. (1982): *To Dwell Among Friends: Personal Networks in Town and City*. University of Chicago.

[13] Harrison, S., and Dourish, P. (1996): Re-Place-ing Space: The Roles of Place and Space in Collaborative Systems, in *Proceedings of CWCW 1996*, ACM Press, Boston, Ma, 1996, pp. 67-76.

[14] Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000): Explaining collaborative filtering recommendations, in *Proceedings of CSCW 2000*, pp. 241-250.

[15] Jackson, R.M. (1977): Social Structure and Process in Friendship Choice, in Jackson et al, pp. 59-78.

[16] Jaffe, J. M., Lee, Y., Huang, L., and Oshagan, H. (1995): Gender, Pseudonyms, and CMC: Masking Identities and Baring Souls. Presented at the *45th Annual Conference of the International Communication Association (ICA)*, Albuquerque, New Mexico, USA.

[17] Kautz, H., Selman, B., and Shah, M. (1997): ReferralWeb: Combining Social Networks and Collaborative Filtering, in *Communications of the ACM*, 30(3), March 1997.

[18] Kobsa, A. (2001): Generic User Modeling Systems, in *User Modeling and User-Adapted Interaction* 11. Kluwer, 2001, pp. 49-63.

[19] Lawrence, J. A and, Mongeau, P. (1996): The Role of Image Anticipation of Future Interaction in Computer-Mediated Impression Development. presented at the International Communications Association, Chicago, IL.

[20] Lederer, S., Dey, A., and Mankoff, J. Everyday Privacy in Ubiquitous Computing Environments, in *Ubicomp 2002 Workshop on Socially-informed Design of Privacy-enhancing Solutions in Ubiquitous Computing.*

[21] Lieberman, H. (1997): 'Autonomous Interface Agents', in *Proceedings of CHI 1997*, ACM Press, Atlanta, GA, 1997, pp. 67-74.

[22] McDonald, D. W., and M.S. Ackerman (2000): Expertise Recommender: A Flexible Recommendation Architecture. In *Proceedings of CSCW 2000*, ACM Press, 231-240.

[23] Olson, J. S., Olson, G. M., and Meader, D. K. (1995): 'What mix of video and audio is useful for small groups doing remote real-time design work?', in *Proceedings of CHI 1995*, ACM Press, Denver, CO, 1995, pp. 362-368.

[24] Palen, L. and Dourish, P. (2003): Unpacking "privacy" for a networked world, in *Proceedings of CHI 2003,* ACM Press, Ft. Lauderdale, FL, 2003, pp. 129-136.

[25] Putnam, R. (2000): *Bowling Alone*. New York, Simon & Schuster.

[26] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994): 'GroupLens: an open architecture for collaborative filtering of netnews', in *Proceedings of CSCW 1994*, ACM Press, Chapel Hill, NC, 1994, pp. 175-186.

[27] Rich, E. (1979): 'User Modeling via Stereotypes', in *Cognitive Science*, vol. 3, pp. 329-354.

[28] Salton, G., and McGill, M. (1983): *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.

[29] Svensson, M., Höök, K., Laaksolahti, J., and Waern, A. (2001): 'Social navigation of food recipes', in *Proceedings of CHI 2001*, ACM Press, 2001, pp. 341-348.

[30] Swearingen, K., and Sinha, R. (2002): Interaction Design for Recommender Systems, in *Proceedings of DIS2002*, ACM Press, London, 2002.

[31] Staw, B.M. (1975): Attribution of the 'Causes' of Performance: A General Alternative Interpretation of Cross-Sectional Research on Organizations, *Organizational Behavior and Human Performance*, vol. 13, pp. 414-32.

[32] Turkle, S. (1995): *Life On The Screen: Identity in the Age of the Internet*. New York: Touchstone Books.

[33] Van Alstyne, M. and Brynjolfsson, E. (1996): 'Electronic Communities: Global Village or Cyberbalkans?' In *Proceedings of 17th International Conference on Information Systems*, 1996.

[34] Verbrugge, L. M. (1977): 'The Structure of Adult Friendship Choices', *Social Forces*, vol. 56, no. 2, Dec. 1977.

[35] Walther, J. B. (1992). Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research*, 19(1), 52-90.

[36] Wexelblat, A. and Maes, P. (1999): 'Footprints: History-Rich Tools for Information Foraging', in *Proceedings of CHI 1999*, ACM Press, 1999, pp. 270-277.