My overall goal as a researcher is to help groups make sense of information. This leads me to a number of subfields of computer science, including human computer interaction, computer-supported cooperative work, and recommender systems, and to a lesser extent, artificial intelligence, information retrieval, and computer science education. As a researcher, I am primarily a builder and experimenter, developing algorithms and systems that are informed by social science theories about how people interact with computers, information, and each other. Ideally, these efforts lead to systems that are effective at meeting users' needs and improving their lives, while helping designers and researchers to understand both when theories apply to design problems and how to apply them.

This overall goal, plus my experiences as a member of the GroupLens research lab, the CommunityLab collaboration between Minnesota, Michigan, and Carnegie Mellon, and the HCI Group at Cornell have shaped my work. I have made a number of contributions:

1) Creating a better user experience in recommender systems, by making recommendations for groups and helping new users quickly teach a system to make accurate recommendations.

2) Making recommendations for research papers that combine content-based and collaborative filtering approaches, while exploring how to evaluate live recommendation systems.

3) Understanding how interfaces affect people's contributions to a community, including the effect of displaying predictions on rating accuracy and the effect of matching people with similar peers on posting to discussion forums.

4) Encouraging people to provide public goods in online communities, by using social mechanisms such as peer review to inhibit bad behavior along with recommendation algorithms that match people with tasks they prefer.

5) Exploring how technologies can support social interactions, using the records of people's activity in social systems ("digital traces" or "read and edit wear") to understand behavior, to make people aware of their behavior, and to help them change it when appopriate.

Below I discuss each in more detail.

When I arrived at GroupLens, the lab was moving toward the problem of **creating a better user experience in recommender systems**. We attacked two specific questions.

- How can systems make recommendations for groups? **(O'Connor et al. 2001)** This raised issues including representing groups, developing group recommendation algorithms, and managing privacy. We found simple algorithms that minimize any group member's "misery" were appropriate for our mostly-small groups, and most users are willing to trade privacy for functionality.

- How can algorithms for choosing items to rate best support new users? **(Rashid et al. 2002)** Making recommendations for someone the system knows nothing about is a classic recommender systems issue. We studied how several item selection algorithms affected a system's ability to learn quickly and well about new users. Algorithms that balance a user's ability to rate an item with the amount of information a rating will provide to the system do best in creating a painless signup process that results in accurate recommendations.

One criticism of recommender systems research is that most deal with taste-based domains such as movies, books, and restaurants. In response, we worked toward **making recommendations for research papers**. During a summer internship with Steve Lawrence at NEC Research, I added personalized recommendations to the CiteSeer system **(Cosley et al. 2002)**. This presented two challenges.

- How can one evaluate recommendation algorithms in a live system? Most evaluation measures recommendation accuracy in offline analysis of previously collected datasets. We argued that measuring utility directly by observing user behavior is more effective than offline accuracy metrics because behavior captures important aspects of recommender effectiveness such as decision support.

- How can a collaborative filtering-based recommender incorporate content? Standard collaborative filtering algorithms ignore content, considering only user ratings. In domains such as movies this makes sense, but research papers offer rich content in the form of text and through links to other papers. A number of hybrid approaches that combine content and ratings information exist (Burke 2002). I developed an *in context* algorithm that considers the user's current context and the citation network to limit the population of documents, then ranks them using the Personality Diagnosis algorithm (Pennock 2000). This was efficient—important on a heavily-used site like CiteSeer—and reasonably useful. GroupLens has continued to study using citation information to recommend reseafrch papers; I helped this effort by developing an in context recommender that uses text similarity to rank selected documents—an approach many users like **(McNee et al. 2002)**.

In CiteSeer, the interface strongly affected recommender performance; page position had a huge effect on recommendation use. GroupLens had also begun to turn toward understanding why people contribute to online communities, working with colleagues from Carnegie Mellon and the University of Michigan. Marrying these concerns brought us to the issue of **understanding how interfaces affect people's contributions to communities.**

- Does showing a predicted rating for an item influence how people rate it? **(Cosley et al. 2003)** The theory of conformity suggested they might (Asch 1951), especially since people often use their social reactions to people when interacting with computers (Nass and Moon 2000). We presented people with movies they had already rated, using their old rating as a prediction, and with movies they had not yet rated, using an actual collaborative filtering algorithm's prediction. Sometimes we deliberately altered this prediction. Whether the prediction was accurate or altered, people who saw predictions tended to rate toward them.

- Can we use people's ratings to encourage them to contribute to conversations? **(Ludford et al. 2004)** We formed people into discussion groups with more and less similar members based on movie ratings. It turns out that while people tend to *like* similar people more (Verbrugge 1977), they have more to talk about with *dissimilar* people. We also selected movies that were relevant to a question and told people about "unique" movies they had seen but few others had. The collective effort model (Karau & Williams 1993) predicts that people are more likely to contribute when they know their contribution matters. "Uniqueness" turned out to be an effective way to operationalize that prediction, increasing contributions.

In the above studies, people were making contributions primarily for their own benefit. Many communities, however, are creating group-specific resources, which we call Community Artifacts of Lasting Value, or CALVs. People have an incentive to not contribute information to a CALV, because contributing requires effort while they already possess the information. But, if everyone contributes, the community as a whole benefits. This is a public goods problem (Hardin 1982), a variation of the classic tragedy of the commons (Hardin 1968). This brings me to my dissertation work: **encouraging people to provide public goods in online communities**.[1]

- How does review by other community members affect the quality and quantity of contributions? **(Cosley et al. 2005)** I developed a mechanism that allowed MovieLens members to contribute movies to its database, dividing members into three groups. One group's contributions were checked by an expert before being added, a second group's contributions were checked by peers, and a third group's were not checked. People in the no checking group did the least and the worst work. Peer checking, however, was as effective as expert checking for both quantity and quality. Since there are few experts, using peers for review will allow communities to scale while being robust against the departure of key members.

- How does reviewing contributions before making them available to the community affect quality? **(Cosley et al. 2006)** I have developed a mathematical model of how review affects the quality of a CALV. The model claims that all else being equal, reviewing contributions before including them does worse in the short term and no better in the long term. A field experiment supports the model's predictions.

- Can algorithms that intelligently match people with tasks increase contributions? **(Cosley et al. 2006, 2007)** Simple task routing algorithms that use how much someone might like a task, how easy a task might be, and how important the task is to the CALV all appear to increase willingness to contribute, with ease having a huge effect. Slashdot, which assigns meta-moderation randomly, and Wikipedia, which suggests new articles to create based on community desire, might increase contributions with task routing.

As people act in digital contexts, they leave what Marc Smith calls *digital traces* and Hill et al. call *read and edit wear* (Hill et al. 1992). These traces can be used to understand behavior, as in the modeling of quality work above, and to influence it, as in the intelligent task routing work. At Cornell, I have focused on **exploring how technologies can support social interactions** through exploiting these digital traces**.** Unlike the work in MovieLens and Wikipedia, this work has taken place in the relatively intimate confines of the workgroup and the museum.

- Can we use dynamic behavioral feedback to improve people's ability to collaborate using computer-mediated communication systems?[2] **(Leshed et al., 2007)**. Group work is often

[1] I played a significant role in writing two successful NSF grants that build on this work, NSF 0534420 ("Helping Hands: Computer Support for Community-Maintained Artifacts of Lasting Value") and NSF 0729344 ("DHB: Collaborative Research: Solving Critical Problems in Online Groups").

[2] I contributed to an NSF grant, under review, that is based on this work.

conducted online, though email, instant messaging, discussion forums, and wikis, but participants are rarely trained in effectively using these tools for collaboration. In this work, we found that giving people feedback during the course of a task changes their communication behavior, while simple linguistic measures of word count and pronoun use correlate with explicit feedback provided by peers.

- How can technology support social connection in public places?[3] Ubiquitous computing will move digital traces from the web to the physical world, and a likely emerging application will be supporting social goals in public places. We developed, deployed, and evaluated an interactive visualization of museum visitors' reactions to an exhibit, highlighting social connections between people through their reactions. Although people expect technologies in the museum to provide exhibit information, they were receptive to the idea of exploring social connections both for their own sake and as a way to experience the museum and its exhibits in new ways.

The availability of this digital information about behavior opens a number of frontiers that I expect to push forward over the coming years. Much of my work so far has been rooted in design: how can I exploit behavioral data to build better systems, experiences, and communities? This is a question that I am passionate about, whether the context is social activity on the web, in the workplace, or in the world, and I will continue to be a builder who really cares about evaluation and outcomes. To effectively exploit these data however, I will need to be involved in *computational social science*. In theory, social science theory should be useful for informing technology design. In practice, doing this requires a lot of practice because the mappings from theories to designs are difficult **(Ling et al. 2005)**. Digital traces can help build bridges between theory and design. Examining digital behavior can allow us to better understand the effective contexts for theories built primarily on control lab experiments; it can lead us to operationalize the theories in ways that support design activities; and it may reveal patterns of behavior that lead to new theories. It is an exciting time to be someone who knows, and cares about, both the computer and the human sides of human-computer interaction.

[3] Submitted to CHI 2008.

References[4]

Asch, S.E. (1951). Effects of group pressure upon the modification and distortion of judgements. In *Groups, Leadership, and Men*, pp. 177-190.

Burke,R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4), pp. 331-370, November 2002.

Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. (2007). SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of IUI 2007*, Honolulu, HI, pp. 32-41.

Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2006). Using Intelligent Task Routing and Contribution Review to Help Communities Build Artifacts of Lasting Value. *CHI 2006*.

Cosley, D., Frankowski, D., Kiesler, S., Terveen, L., & Riedl, J. (2005). How Oversight Improves Member-Maintained Communities. *CHI 2005*, Portland, OR.

Cosley, D., Lam, S.K., Albert, I., Konstan, J., & Riedl, J. (2003). Is Seeing Believing? How Recommender Systems Influence Users' Opinions. *CHI 2003*, Fort Lauderdale, pp. 585-592.

Cosley, D., Lawrence, S., & Pennock, D.M. (2002). REFEREE: An open framework for practical testing of recommender systems using ResearchIndex. *VLDB 2002*, Hong Kong, pp. 35-46.

Fogg,B.J. (1998). Persuasive computers: perspectives and research directions. In *CHI 1998*, pp. 225-232.

Hardin, G. (1968). The tragedy of the commons. *Science*, 162, pp. 1243-1248.

Hardin, R. (1982). *Collective Action*. Johns Hopkins, 1982.

Hill, W., Hollan, J., Wrobleski, D. and McCandless, T. Edit Wear and Read Wear. In *Proc. CHI'92 Human Factors in Computing Systems* (Monterey, 1992), Addison-Wesley, pp. 3-9.

Leshed, G., Hancock, J., **Cosley, D.**, McLeod, P., & Gay, G. (2007). Feedback for Guiding Reflection on Teamwork Practices. *Group 2007*, Sanibel Island, Florida.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., **Cosley, D.**, Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P., and Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4), article 10.

Ludford, P., **Cosley, D.**, Frankowski, D., & Terveen, L. (2004). Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity. *CHI 2004*, pp. 631-638.

Karau, S.J., & Williams, K.D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality & Social Psychology*, 65(4), pp. 681-706.

McNee, S., Albert, I., **Cosley, D.**, Gopalkrishnan, P., Lam, S.K., Rashid, A.M., Konstan, J., & Riedl, J. (2002). On the Recommending of Citations for Research Papers. *CSCW2002*, pp. 116-125.

Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103.

O'Connor, M., **Cosley, D.**, Konstan, J. A., & Riedl, J. (2001). PolyLens: A Recommender System for Groups of Users. *ECSCW 2001*, Bonn, Germany, pp. 199-218.

O'Donovan, J. & Smyth,B. (2005). Trust in recommender systems. In *IUI 2005*, pp. 167-174.

Palen, L. & Dourish, P. (2003). Unpacking "privacy" for a networked world. In *CHI 2003*, pp. 129-136.

Pennock, D.M., Horvitz, E., Lawrence, S., & Giles, C.L. (2000). Collaborative filtering by personality diagnosis: A hybrid memory and model-based approach. In *UAI 2000*, San Francisco, CA, pp. 473–480.

Rashid, A.M., Albert, I., **Cosley, D.**, Lam, S.K., McNee, S., Konstan, J.A., & Riedl, J. (2002). Getting to Know You: Learning New User Preferences in Recommender Systems. *IUI 2002*, pp. 127-134.

Verbrugge, L. M. (1977). The Structure of Adult Friendship Choices. *Social Forces*, 56(2), Dec. 1977.

[4] I have participated in a number of other research activities not listed here; see my CV for a full account.