

My research goal is to better **integrate technical activities** such as behavior modeling, interface design, and system building **with conceptualizations of social dynamics** as expressed by social science theories—and as an information scientist, I hope to contribute on both the technical and the social fronts. For instance, understanding the working of memory might help designers build better memory support tools; knowing how interfaces affect observed behavior may lead to more accurate models of that behavior; observing how the behaviors people exhibit online accord (or don't) with existing social science theory can advance work in the social sciences. To do this, I use my computer science background in human-computer interaction, computer supported cooperative work, recommender systems, and artificial intelligence to **build and study systems in collaboration with social science researchers**. My specific research domain is helping people and groups make sense of information, particularly information they create themselves. Below, I'll explain what this means, using a metaphor of data as a door, as a window, and as a mirror.

Data as a door. People leave traces of their interaction with digital objects (Hill et al., 1992). Often these traces give information not just about the use of the object but about the people themselves, such as their interests, knowledge, and preferences. Collaborative filtering-based recommender systems (Resnick et al., 1994), which use data such as movie ratings, the reading of news articles, or purchases on an e-commerce site to generate recommendations for new movies, articles, or purchases, have been successful both as a research area and a practical tool. When I entered grad school with the GroupLens lab at Minnesota, much of the algorithmic work had already been done. Thus, my early work focused on **making recommender systems more useful**, including recommending for groups (O'Connor et al., 2001), supporting new users (Rashid et al., 2002), and evaluating systems based on recommendation adoption rather than the common machine learning approach based on testing the algorithms offline, against previously collected datasets (Cosley et al., 2002).

This was also where I first developed the goal of **bringing system design and social science theory together**, in the context of the CommunityLab collaboration between Minnesota, Michigan, and CMU. For instance, theories about conformity (Asch, 1951) and the idea of computers as systems that people respond to socially (Nass & Moon, 2000) suggest that the recommender interface itself can influence people's ratings of items by showing the ratings the system predicts. This is true, and has negative consequences both for the systems themselves and studies based on data collected by these systems such as the MovieLens datasets (Cosley et al., 2003). Likewise, theories of motivation in group settings such as the collective effort model (Karau and Williams, 1993) and the idea of public goods (Hardin, 1982) suggest a number of **strategies to encourage people to contribute more data to the system or the community**. These including making the value of one's contributions clear (Ludford et al., 2004; Ling et al., 2005) and reducing the cost of contributing through *intelligent task routing*, or matching people with appropriate work (Cosley et al., 2006, 2007).

The recommender systems, user modeling, and e-commerce approach to personalization often treats a person's interaction trace and self-revelation data as the key to a locked door. Understanding a person's preferences unlocks the door, allowing the **owner of the system to use this knowledge** to make recommendations and affect our behavior (not unlike the idea of persuasive interfaces). This often benefits the owner in e-commerce scenarios; it may benefit a group of people in the community settings described above; and it may also benefit the individual by reducing their search costs or increasing their happiness through effective recommendations. But the person who generated the data is rarely considered an active agent participating in the process of understanding, and the soul is in the machine.

Data as a window. Another way to see these data is in the aggregate: an often large, but noisy, record of people's interactions that can be used to **test or generate theories and models of social behavior**. This

work often goes under the rubric of computational social science (Lazer et al., 2009), and tools as diverse as social network analysis, automatic annotation and natural language processing, and PageRank have leveraged these data to advance both our understanding of social phenomena and our ability to build systems that support fundamental tasks such as information finding. When I entered Cornell, a group of researchers at the Institute for Social Sciences was interested in exploring social networks at a large scale. My work on intelligent task routing and knowledge of Wikipedia as both a dataset and a community made me a natural collaborator, leading to work on a number of fronts that **used social network approaches to understand collaboration in Wikipedia**. These included exploring how well theories of social influence affect the diffusion of tools in Wikipedia (Yuan et al., 2007, 2009a, 2009b), inferring roles based on the activities people engage in (Welser et al., 2008, 2011), and understanding discourse patterns in discussions around article production (Black et al. 2008, forthcoming; Thom-Santelli, Cosley, & Gay, 2009). The most interesting part of this work to me combined large-scale modeling with close analysis of interactions to understand **how homophily of interests and social influence interact** in leading people to start conversations and choose articles to edit in Wikipedia (Crandall et al., 2008). Most collaboration is planned: people talk about what to work on, then work on it. In distributed and decentralized collaborations like Wikipedia, however, the process appears to be reversed: people to initiate conversation when they need to coordinate their independent work around the same articles.

The work above was primarily observational and analytical in nature. However, as a **builder and designer**, I tend to favor research where a major component is developing systems that both give insight into behavior and may solve real problems. As a postdoc with Geri Gay's Interaction Design Lab, I led several projects where we designed systems to **understand and support tagging behavior**, while providing museum experiences that go beyond the standard “provide information” model that dominates museum informatics to focus on social and spiritual aspects of the experience instead (Bell, 2002). Both the Artlinks (Cosley et al., 2008) and MobiTags (Cosley et al., 2009) systems allowed people to label objects in a museum. In contrast to the typical information science view of tags as a tool primarily for categorizing and finding things, our designs did succeed in **supporting these social and reflective aspects of experience**. People used the tags to reflect on and think about objects in new ways as well as thinking about the people who created unexpected tags—not unlike Wash and Rader (2007) finding that tags primarily help experienced del.icio.us users find people who add interesting content, not to categorize their own links. Further, people used the tags in communicative ways, as tools to signal their knowledge and to reach out to other, like-minded experts (Thom-Santelli, Cosley, & Gay, 2010). In another system, kultagg, we found that making the tagging system itself more expressive through allowing people to choose the location and color of tags was valuable and led to a number of potential **uses of tags beyond categorization** (Cheng & Cosley, 2010).

This kind of research, which uses trace and self-revelation data as a tool to answer social science questions and probe at behavior, treats these data as a window. Here, the goal is to allow the **researcher to create knowledge** based on these data to advance research agendas and to come to better understandings of human behavior. Again, this may have side benefits—these understandings might then feed into later system design, while systems designed specifically to answer questions may support new needs and behaviors. But again, the people who generate the data are typically objects of study, not agents in the research, and the soul of this research is in the hypotheses.

Data as a mirror. A less common, but increasingly plausible use for these data is to **present it back to the people who created it**. In HCI, this idea goes under various names and purposes, including personal informatics, to support self-analysis (Li et al., 2010); lifelogging, to support memory (e.g., Gemmell, Bell, & Lueder, 2006); and conversational visualization, to support reflection on

relationships (Donath, Karahalios, & Viégas, 1999). However, the goal of such systems is often not clear (Sellen & Whittaker, 2010). One notable counterexample is the GroupMeter system that I worked on with Gilly Leshed, which helps members of a workgroup become aware of social aspects of teamwork through reflecting the language they use in work conversations (Leshed et al., 2007). It turns out that creating awareness of language use is not too hard, but changing it substantively is difficult (Leshed et al., 2009) and that **careful attention to design and providing task-appropriate normative guidance are critical elements** in this kind of system (Leshed et al., 2010).

My main impact so far in this area is around Pensieve (Cosley et al., 2009). Pensieve supports reminiscence, an activity with important psychological benefits (Webster & McCall, 1999) that is becoming more prevalent in HCI, as evidenced by the strong response to a CHI 2011 workshop I organized. Pensieve's goal is to **support both thinking and writing about the past** by reflecting the data people create in social media back to them through either email or Facebook, and providing a convenient diary interface for writing about their memories. A large-scale deployment showed that **Pensieve supports common goals of reminiscing** such as maintaining social connectedness and working through the past and encourages writing stories (Peesapati/Schwanda, et al., 2010). Analysis of people's reactions to the system led to follow-on work that showed the need for presenting culturally congruent stimuli for reminiscing (Peesapati, Wang, & Cosley, 2010), the potential value of place as a subject of and trigger for reminiscing (Peesapati, Schwanda, Schultz, & Cosley, 2010), and the value—and pitfalls—of supporting reminiscence **not just for individuals, but for groups** (in submission to *HCI*).

I'm especially interested in work that **combines multiple perspectives**, as with IdeaExpander, a system that Hao-Chuan Wang, Sue Fussell, and I collaborate on (Wang, Cosley, & Fussell, 2010) with the goal of better understanding and supporting creativity and cross-cultural collaboration. In IdeaExpander, intelligent agents monitor an ongoing conversation, extract interesting ideas and keywords, and find relevant pictures to present during the conversation. Here, the computer uses its (imperfect) understanding of the current state of the conversation to present data that the humans can reflect on and integrate into the conversation. It can **increase both the quantity and breadth of ideas generated during brainstorming** by choosing related pictures tagged with multiple, relatively rare concepts and by **leveraging cultural differences in knowledge and perception** (Wang, Fussell, & Cosley, 2011a) suggested by theories of associative memory and studies of cognition across cultures. Further, it can be used to **give people the benefit of speaking their own language** in multi-lingual conversation, using the pictures to support grounding by repairing translation errors and presenting visual referents to anchor new conversational topics (Wang, Fussell, & Cosley, 2011b). Our current work includes using the system to support storytelling using a person's own photos and blog entries, and we expect a number of future applications both of this system and the more general idea of CMC tools where the computer uses partial understanding of people's state, knowledge and goals to actively participate in the conversation.

I have a number of other projects coming up to speed that look at people's trace and self-revelation data from multiple perspectives:

- RegulationRoom, with Cynthia Farina, Mary Newhart, and Claire Cardie. The goal here is to **increase effective public commenting on federal government regulations**, using our site RegulationRoom.org. We are using insights from NLP about opinion mining and from conflict resolution research about effective moderation and productive discourse practices, developing algorithms, interfaces, and policies that both help people write more thoughtful, evidence-based, and useful comments and to productively engage with divergent points of view (Farina et

al., 2011a, 2011b).

- Goalmometer, with Johnathon Schultz. The goal here is to help people **track, monitor, and reflect on their goals**, on the progress they are making toward them, and on obstacles in their way. We are leveraging theories of motivation and attitude and studies of existing practice around goal management to develop enhanced to-do lists and visualizations that we will deploy to examine people's uptake of and reaction to making goals and progress more explicit and accessible.
- Being Heard, with Victoria Schwanda and Xuan Zhao. Here, the goal is to **help people reflect on their relationships through exploring archives of email**, chat, and other conversation by presenting these data in engaging visualizations. Much earlier work in this space focuses on aesthetic goals; we hope to make serious use of communication theories to help guide design choices that lead to tangible, useful outcomes and encourage reflection on the relationship itself rather than specific past incidents. We also plan to go beyond earlier work by **connecting these archives to present context**, and by exploring interfaces designed to be **used together as pairs** rather than as individuals.

The future of this kind of work is bright, and it will be a **major growth area in HCI research**, as people create increasing amounts of data online, as sensors collect increasing amounts of data about people's presence and context, and as people continue to integrate technology use into their daily lives. I hope that careful attention to the psychological and social processes and constraints involved in generating and collecting these data, combined with thoughtful system development and evaluation, will help us understand how these data can be used **for computational understanding, for fundamental social science research, and for the benefit of the people who generate it.**

References

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership, and Men*, 177–190.

Bell, G. (2002). Making Sense of Museum: The Museum as 'Cultural Ecology': A study. *CIMI whitepaper, Intel Corporation*.

Black, L. W., Welser, H. T., DeGroot, J. M., Cosley, D. (2008). 'Wikipedia is not a democracy': Deliberation and Policy-Making in an Online Community. *International Communication Association*, Montreal.

Black, L. W., Welser, H. T., DeGroot, J. M., Cosley, D. (forthcoming). Self-governance through Group Discussion: Wikipedia's Policy Making Implications for Virtual Teams. *Small Group Research*.

Cheng, J., Cosley, D. (2010). kultagg: ludic design for tagging interfaces. *Proceedings of GROUP 2010*.

Cosley, D., Akey, K., Alson, B., Baxter, J., Broomfield, M., Lee, S., Sarabu, C. (2009). Using Technologies to Support Reminiscence. *BCS HCI 2009*. Cambridge, UK.

Cosley, D., Baxter, J., Lee, S., Alson, B., Adams, P., Nomura, S., Sarabu, C., Gay, G. (2009). MobiTags: Supporting Semantic, Spatial, and Social Interaction in Museum Spaces. *CHI 2009*.

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., Terveen, L., Riedl, J. (2007). SuggestBot: Using Intelligent Task Routing

to Help People Find Work in Wikipedia. *IUI 2007*.

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.

Cosley, D., Lewenstein, J., Herman, A., Holloway, J., Baxter, J., Nomura, S., Boehner, K., Gay, G. (2008). ArtLinks: Fostering Social Awareness and Reflection in Museums. *CHI 2008*.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S. (2008). Feedback effects between similarity and social influence in online communities. *KDD 2008*.

Donath, J., Karahalios, K., Viégas, F. (1999). Visualizing Conversation. *Journal of Computer-Mediated Communication*. 4(4).

Farina, C. R., Newhart, M., Cardie, C., Cosley, D. (2011a). Rulemaking 2.0. *University of Miami Law Review*, 65(2).

Farina, C. R., Newhart, M., Miller, P., Cardie, C., Cosley, D., Vernon, R. (2011b). Rulemaking in 140 Characters or Less: Social Networking and Public Participation in Rulemaking. *Pace Law Review*, January 2011.

Gemmell, J., Bell, G., Lueder, R. (2006). MyLifeBits: a personal database for everything. *Communications of the ACM* 49(1), 88– 95.

Hill, W. C., Hollan, J. D., Wroblewski, D., McCandless, T. (1992). Edit wear and read wear. *Proceedings of CHI 1992*. 3–9.

Lazer, et al. (2009). Computational Social Science. *Science* 323, 721–723.

Leshed, G., Cosley, D., Hancock, J. T., Gay, G. (2010). Visualizing Language Use in Team Conversations: Designing Through Theory, Experiments, and Iterations. *CHI 2010 Design Case Study*.

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

Leshed, G., Perez, D., Hancock, J. H., Cosley, D., Birnholtz, J., Lee, S., McLeod, P. L., Gay, G. (2009). Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. *CHI 2009*.

Li, I., Dey, A., Forlizzi, J. (2010). A stage-based model of personal informatics systems. *Proc. CHI 2010*, 557– 566.

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

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 and Social Psychology*, 65(4):681–706.

Nass, C., Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 60(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.

Peesapati, S. T., Schwanda, V., Schultz, J., Cosley, D. (2010). Triggering memories with online maps. *Proceedings of ASIST 2010*.

Peesapati, S. T., Schwanda, V., Schultz, J., Lepage, M., Jeong, S., Cosley, D. (2010). Pensieve: Supporting Everyday Reminiscence. *CHI 2010*.

Peesapati, S. T., Wang, H.-C., Cosley, D. (2010). Intercultural human-photo encounters: How cultural similarity affects perceiving and tagging photographs. *Proceedings of ACM International Conference on Intercultural Collaboration (ICIC 2010)*.

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.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *CSCW '94: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*. Chapel Hill, North Carolina, United States: ACM Press, 175–186.

Sellen, A. J., Whittaker, S. (2010). Beyond total capture: a constructive critique of lifelogging. *Commun. ACM*. 53(5), 70-77.

Thom-Santelli, J., Cosley, D., Gay, G. (2009). What's Mine is Mine: Territoriality in Collaborative Authoring. *CHI 2009*.

Thom-Santelli, J., Cosley, D., Gay, G. (2010). What Do You Know? Experts, Novices and Territoriality in Collaborative Systems. *CHI 2010*.

Wang, H.-C., Cosley, D., Fussell, S.R. (2010). IdeaExpander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli. *CSCW 2010*. Savannah, GA.

Wang, H.-C., Fussell, S., Cosley, D. (2011a). From Diversity to Creativity: Stimulating Group Brainstorming with Cultural Differences and Conversationally-Retrieved Pictures. *CSCW 2011*.

Wang, H.-C., Fussell, S., Cosley, D. (2011b). Using Language-Retrieved Pictures to Support Multilingual Brainstorming. Demo at *CSCW 2011*.

Wash, R., Rader, E. (2007). Public Bookmarks and Private Benefits: An Analysis of Incentives in Social Computing. *Proceedings of the American Society for Information Science and Technology (ASIS&T) Annual Meeting*.

Webster, J. D., McCall, M. E. (1999). Reminiscence functions across adulthood: A replication and extension. *J. Adult Dev.*, 6(1):73–85.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M. (2008). Finding social roles in Wikipedia. *American Sociological Association*, Boston.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M. (2011). Finding Social Roles in Wikipedia. *iConference 2011*.

Yuan, Y. C., Cosley, D., Ling, X., Welser, T. (2009). The Diffusion of a Task Recommendation System to Facilitate Contributions to an Online Community. *International Communication Association*.

Yuan, Y. C., Cosley, D., Welser, H. (2007). The Impact of Network Relations on the Diffusion of SuggestBot in Wikipedia. *National Communication Association*, Chicago, IL.