Leveraging online behavior to support knowledge and memory

This proposal describes computational work on models, algorithms, and interfaces that leverage people's current online activities and insights from social science to create valuable individual and public goods. The work focuses on two socially valuable domains: (1) improving the quality of open content systems such as Wikipedia by motivating people to contribute to them, and (2) supporting people's self-concepts and relationships with others by using content they create in social media to support reminiscence. My prior work with large behavioral datasets, algorithms that motivate contributions to public goods, social network modeling, and interface design will inform the proposed work. I also plan educational initiatives to help students become interdisciplinary scholars who contribute to the research. The proposal will advance the science and engineering of social systems and set the direction for my career as an excellent researcher and teacher.

1 Introduction and motivation

Traces of people's activity in the physical world tell us useful information. Dog-eared book pages point out key ideas, a crowded restaurant is probably good, a dirt path shows us how people really get around on campus. People leave similar traces around digital artifacts, an idea called "edit wear and read wear" [89]. This insight lies behind a number of innovations in human-computer interaction (HCI), including collaborative filtering (e.g., [88,150,158]), social navigation (e.g., [161]), social accounting (e.g., [56, 62, 105]), website analysis tools (e.g., [35, 36]) and search [142].

Most prior work has focused on read wear because until recently, people primarily consumed information online. Things have changed. People post pictures to Flickr [93, 126] and videos to YouTube, blogs about the world and themselves [133], transcribe books [137] and write encyclopedias [60], articulate and use their social networks [23,59], and tag everything in sight [5,78,126,155]. It is not exaggerating to call this an explosion of user-created content. By 2007, over 35% of all U.S. Internet users, including two thirds of teens, had posted content to the Internet—and this does not count comments on or ratings of other content [92, 119].

1.1 Computation, social science, contribution, and reminiscence

A question with important practical implications and great relevance to computer science, then, is how to make use of these data. One way is to use the data to understand human behavior. Much of this online data can be modeled as interactions with other people, leading to the use of social network analysis techniques [175] and other algorithms to explore how people interact both online (e.g., [9, 63, 181] and in the world [58]. Social networks also support applications such as facilitating introductions [167], finding experts [127, 157], and detecting suspicious activity [39]. A second common use of the data people generate is to post a problem that allows the data to be mapped onto a known machine learning paradigm, as with the influential paper "Recommendation as Classification" [12]. This allows these algorithms to glean new value from old data.

A promising approach for helping these tools perform even better is to consider that the data are generated by people who are *motivated* to undertake certain *behaviors* in support of their *goals*. A number of studies by the PI and others have shown that theories from social psychology, economics, sociology, and other social sciences can inform computational work around models, algorithms, and interfaces that use these data [7,13,23,25,27,30,34,47,50,53,60,61,84,103,123,124,149,169,170]. A

common thread in this work is to help groups attract active contributors, applying theories about motivation (e.g., [82, 106, 146]) to the design of software that supports them.

Motivating contributors is particularly important for the open source and Wikipedia communities, which create valuable public goods through individual volunteers' effort. Thus, I propose to extend my earlier work around motivating people to contribute to public goods [44, 45], building on the SuggestBot interface that uses people's prior editing behavior in Wikipedia to recommend other articles that both need improvement and match people's interests [46]. I will develop models of people's Wikipedia careers that reflect how milestone events, interactions with other people, and transitions between roles affect contributions over time. These models will inform an improved version of SuggestBot that increases people's motivation to contribute to Wikipedia and related communities by recommending activities and interactions suggested by the models.

Thinking about people's behavior, motivation, and goals can also stimulate ideas for new uses of data. In many cultures, custom demands that hunters "use every part of the whale", yet most social media make little use of user-created content beyond targeted advertising. In particular, people often create content that reflects important, memorable life events. This content could be re-purposed to support reminiscence, which serves valuable psychological purposes throughout life [179]. I propose to build models that identify content that is likely to be useful for reminiscence and use conversations about that content to identify strong relationships between people. This work will inform algorithms that support reminiscence, choosing times, people, and content that align with people's current practices. I will study those practices through a series of interviews and prototypes that probe how reminiscence fits into daily life, and develop applications that use these insights with content people already create online to support valuable reminiscence experiences.

This work will require educational initiatives to support students in becoming effective, motivated interdisciplinary scholars. I will develop courses that help students approach online social systems from both technical and social perspectives while giving them tools to work with the data people create in these systems. I will also connect undergraduates with researchers who have different perspectives on this work through conferences participation and summer research opportunities at other labs, broadening students' interdisciplinary experience and increasing their interest in scholarly activity as a career. Finally, I will deepen the introductory HCI curriculum at Cornell by adding short practicum courses around qualitative methods for usability and interface development, and provide mentoring opportunities that will allow students to grow as interdisciplinary scholars.

2 Research overview and broader context

I will use a number of methods to attack these problems, both because multi-method approaches generally improve research [128] and because different questions demand different tools. I will start by identifying theories of behavior that are likely to pertain to the problem at hand, applying them to model people's behavior and the data it generates. These models may be inspired by statistics, machine learning, or agent-based simulation, tying important model parameters to theoretical constructs. The models will inform algorithms that use features that prove important to accomplish goals in the domain, such as motivating contributions or identifying content useful for reminiscing. For instance, theory predicts that people who initiate conversations with others early in their Wikipedia career will tend to contribute more frequently than people who don't. If the models bear this out, they would suggest algorithms that look for occasions to start conversations around events such as working on an article together. These algorithms will then be tested in systems that

I will deploy in the communities for long-term evaluation. Throughout these other activities, I will be in contact with community members, which proved valuable in my earlier work in Wikipedia [46] as well as in my work designing for social and spiritual museum experiences [49].

The chosen domains support multi-method approaches. Both provide vast amounts of rich data for the modeling and algorithmic portions of the work. Both also exist as real communities where I can deploy systems. Wikipedia has a number of distinct instances, one for each language Wikipedia supports. The people who work on these instances are quite different; this diversity will allow me to validate the work in a number of contexts and discover cultural differences that matter both for the work and for the social science theories it uses. The public nature of the data generated in these domains ameliorates (but does not eliminate) issues about privacy, ownership, and copyright issues that pertain to collecting and aggregating activity data.

These domains also matter in the world. Wikipedia is an important social good and improving it will benefit society. Likewise, helping people reminisce has the potential to improve their lives. My students and I care about making a difference and have personal interest in these domains. Female students in particular have gravitated toward my preliminary work around reminiscing, suggesting that this work may bring under-represented groups to HCI, computing and information science. As for me, when you're planning your career, it helps to do something you love.

3 Motivating contributors in volunteer communities

3.1 Background and motivation

The first general problem is understanding how to motivate people to contribute more to communities such as open source software and Wikipedia that produce public goods. One path to this understanding directly asks people why they contribute to these systems, revealing motivations including interest in the software or topic, self-improvement, altruism, ideology (it's the "right" way to build software), reputation, and future benefits (e.g., [87, 113, 162]). Computer-supported cooperative work, HCI, and sociology have looked at the high-level decisions that all such communities must face and that affect people's willingness to be part of the community, decisions around fundamental issues such as identity, privacy, governance, policy, managing wrongdoing, and providing community leadership [30, 109, 141, 144, 145]. Finally, research in psychology and HCI has looked to understand why people act (or fail to act!) [54, 83, 86, 106, 108, 115], and how computers can use social characteristics [66, 134–136] and persuasive techniques [64, 65] to change people's behavior.

My contributions to this field use these insights to inform system and algorithm designs that increase participation around creating public goods. This work blends analysis, theory, tool-building, and evaluation, and has generated a number of useful outputs:

> Algorithms that exploit homophily and uniqueness to create good discussion groups [123,124].
>
> Mechanism designs based on the collective effort model of motivation [106] that support efficient review of contributions to public goods [44].
>
> Models of the effect of review mechanisms on the quality of a community's public goods over time [45], suggesting that compared to experts reviewing contributions before they go live, Wikipedia-style review is faster in the short run and just as good in the limit.

Algorithms for *intelligent task routing*, recommending tasks that increase motivation by reducing people's cost of contributing [45].

SuggestBot, an effective implementation of intelligent task routing in Wikipedia [46].

Models that explore how similarity of interests and interaction between people shapes their future behavior in Wikipedia [53].

Visualizations of effective and ineffective conversational patterns in Wikipedia governance, augmenting social network diagrams inferred from conversation with attributes based on theories of deliberative discourse [181].

This work has convinced me that tight integration of theory, analysis, and design is crucial. For instance, people in Wikipedia typically interact for the first time when they encounter each other through working on the same article. This was suggested by the ideas of distributed cognition [91] and object-centered sociality [33, 61], and we empirically observed it by coding randomly sampled first instances of interaction [53]. This insight was important in interpreting the results of algorithms that predict people's future contributions to Wikipedia based on who they are similar to and who they interact with. It was also useful in developing process models of people's Wikipedia activity.

Wikipedia is an interesting domain because of its social value and because the openness of the community's operation supports research. A pleasant overview of research in Wikipedia is available in [17]; at a high level it attempts to answer three main questions. The first focuses on quantitative analysis of article creation and quality, motivated by the question "Does Wikipedia work?" (e.g., [76, 122, 147, 164, 173]). A second broad research theme focuses on understanding what motivates individual contributors and how individuals coordinate and communicate, asking "Why and how do people work on Wikipedia?" (e.g., [27, 37, 60, 169–171]. The third set of studies focuses on Wikipedia governance: how are disputes resolved, vandals repelled, and policies created and used, getting at the question of "Why does Wikipedia work?" (e.g., [2, 17, 20, 68, 111]).

Most of these studies are analytical, though some work has gone into building tools that people can use to investigate conflict and activity in Wikipedia [2, 169], investigate whether authors have ulterior motives when contributing [80, 165], or, as with my own work, using people's existing activity to help them find useful work to do [46]. Likewise, much of the work is atheoretical. Notable exceptions include applying activity theory [132] and communities of practice [116, 117] to understanding how people grow as Wikipedia contributors [27], the use of ideas about genre in organizational contexts [186] to analyze language use [60], and the appropriation of ideas around peer-based common goods [15] and distributed self-governance in communities [141] as a lens for exploring how Wikipedia comes to resolve disputes and govern itself [68, 171].

3.2 Research activities around motivating contributors

The work I propose will integrate quantitative analysis and tool-building with the theory-driven work on understanding behavior. My initial work around intelligent task routing and SuggestBot will serve as a springboard to look more deeply at questions around motivating participation in volunteer communities. How does contributors' motivation change over time, based on their activities and interactions with other members? How do quality of public goods and motivation interact? Can communities exploit these predictors of motivation to create more active, effective, satisfied members? How can implicitly inferred profiles of users' interests and activities be improved?

How does contributors' motivation change over time? I plan to model the effect of pivotal events and role transitions on how people contribute to Wikipedia. Both are potentially powerful for understanding how behavior changes over time. Analysis of thread structures in discussion groups shows that reaction to a user's first post predicts future contributions [7,103], while activity theory [132] predicts and interviews confirm that encountering aspects of the community at appropriate times helps Wikipedians transition from newbies to experts [27]. My own preliminary analysis shows that people who receive simple template welcome messages as their first interaction with Wikipedia contribute more, probably because these messages make other people salient and lead to social bonds with other members [146]. Likewise, the idea of communities of practice [116,117] points to scaffolding role transitions as an important tool for helping people grow in communities. Recent work has shown that social network diagrams can give clues as to a user's roles in a community [63], and Wikipedians in fact recognize a number of formal and informal roles (e.g., [27,68,170,171]).

I will build on this work by modeling the effect of a number of potential turning points in the lifecycle of a community member. In Wikipedia, this might include welcomes or warnings, creating a user page, joining a sub-project, participating in discussions of Wikipedia governance, receiving unofficial recognition, talking with other users, or reaching edit-count milestones (e.g., 10th, 100th, or 1000th edit). A number of theories from social science suggest that these events should affect people's motivation to contribute (e.g., [86, 96, 106, 146, 174]). I will also model role development in Wikipedia. Roles already identified by the community, combined with techniques that consider both the kinds of activity people do and the people who they interact with in detecting roles [181], will help me detect behaviors that predict when a person is making a role transition.

How do quality of public goods and motivation interact? I will also expand on my prior work around modeling contribution quality in public goods [45]. This work suggested that Wikipedia's model of review is likely to be as good in the end as a more structured, expert-centric peer review process. However, the model fails to account for the fact that people's motivation changes based on their experiences in a community. For example, a person whose first contribution is rejected may not come back [103]. My current plan is to adapt current modeling work in Wikipedia [10, 53] that takes inspiration from agent-based modeling techniques used in computational social science (e.g., [8]). I will create process models where agents' contributions are influenced by their level of motivation, which is in turn affected by the critical events identified earlier. After validating these models against actual contribution behavior in Wikipedia, I will use them to explore the conditions under which Wikipedia improves or declines.

I plan to develop these models in the English Wikipedia and test their generality against other language versions of Wikipedia, related projects such as Wikibooks and Wiktionary, and Wikia's communities of interest, all of which are backed by the MediaWiki platform. I focus on communities backed by MediaWiki so that the models, algorithms, and tools to be built can assume a common environment, which will allow me to make progress most quickly. I expect many of the techniques and findings to generalize based on their grounding in theory, and in the future I hope to find support for work that extends these findings to other contexts.

Can communities exploit causes of motivation to increase participation? My intelligent task routing work leverages knowledge of tasks that need to be done, people's history of doing tasks, theories of motivation [42,82,106], and recommendation algorithms [3,88,150] to match people with specific tasks they are likely to do [45, 46]. Expanding this work beyond tasks promises to increase

its value. New algorithms will incorporate the triggers of motivation identified in the modeling work. New user profiles will consider interaction, discussion, and other activities besides editing articles, supporting richer notions of similarity between contributors and providing new elements to use in making recommendations. These recommendations will also go beyond tasks. Based on the models of motivation, they may introduce people to other members, suggest projects that reflect their interests, help people create personal profiles, point people to information about the community, or highlight useful tools [27].

There is no guarantee that helping a user create a profile page will *cause* people to be motivated; self-revelation may be an *effect* of already being a motivated contributor. But there is a good chance that these predictors can be used to motivate. I will test this hypothesis through both offline and online experimentation. I have created a dataset of over 100 million Wikipedia edits that can be used to test recommendation algorithms with standard offline methods [85]. I will deploy algorithms that perform well in the offline testing live in Wikipedia through an updated version of SuggestBot, cleverly named SuggestBot2. I will then observe how the recommendations affect people's behavior: compared to people who receive random or no recommendations, do people who receive these recommendations tend to act on the suggestions more often, stay active for longer periods, and show other indicators of increased motivation to contribute? I have long argued for testing recommendation algorithms in real contexts of use, because the gold standard for whether a recommender system is effective is if it improves people's outcomes [48]. The openness of Wikipedia and my good working relationship with the community based on my SuggestBot work mean that Wikipedia will be an excellent testbed for the work in this proposal.

How can we improve implicit profiles? SuggestBot used implied interest profiles that it inferred based on people's article editing behavior, and SuggestBot2 is likely to do the same, at least at first. Implicit user profiles have known problems: noise[1], failure to detect changing interests, and context-blindness [77, 107]. I plan to investigate two approaches for improving user profiles: allowing users to create profiles that reflect salient interests, and developing algorithmic approaches for removing noise from profiles.

Most behavior in Wikipedia, including article editing, discussion, personal communication, and participating in community governance, is expressed by editing pages. Thus, a straightforward way to represent a user profile is a set of pages, a representation that should be useful for people to work with. Wikipedia's API allows easy retrieval of various set of pages: a user's contributions, the pages a user monitors for changes, pages that belong to a given category, and the set of pages linked to by another page. This last method may be particularly useful because people often list the articles they consider most important on their profile pages. I will build interfaces that help people create a starting set of pages using the Wikipedia API, then edit that set of pages to create profiles that best represent their current or long-term interests. I will evaluate the effectiveness of these interfaces by seeing if people who receive recommendations based on their manipulated profiles act more often on those recommendations then users who are allowed to manipulate their profile but receive recommendations based on a baseline implicitly inferred profile.

I will also investigate ways to algorithmically detect noisy inferences. This may involve adapting attribute selection techniques from machine learning [21] to work at the level of subgroups or individuals, detecting anomalous behaviors that don't match the rest of a person's profile, weighting

[1] Famously, in the *Wall Street Journal* article "My TiVo Thinks I'm Gay".

behaviors differently when computing similarity between users (unlike most collaborative filtering algorithms, which treat all ratings as equally important), or changing the weight of recent behaviors. The algorithms might also use Wikipedia-specific indicators that a given edit does not reflect one's primary interests to filter the profile. Observing how people manually choose items to include and exclude from profiles may also suggest improvements. I will develop algorithms that use these techniques to make more effective recommendations and test them against both the Wikipedia dataset and against publicly datasets of ratings information such as the Netflix prize dataset.

In summary, the work in Wikipedia will allow me to explore the two key motivations for the proposal: leveraging behavior and theory in order to increase the value people derive from participating in online communities. Learning more about motivating participation will lead to better outcomes for participants in Wikipedia and greater value in the public goods they produce.

4 Leveraging personal sharing to support reminiscence

4.1 Background and motivation

The other main domain of work in this proposal concerns using people's online activity to support reminiscence. I became interested in this because I rarely remember the past, even though I've kept a blog for ten years designed specifically to help trigger memories. Two months ago, I wrote a program to remind me to look at the blog once in a while by sending me snippets as text messages on my cell phone. This program, though simple, gives me great value by helping me make use of those memory triggers. In this part of the proposal I look at how I can help other people increase the value they derive from reminiscing.

As I sat down to write this section I received a text message from the program: "Creating a directory called 'thesis' is, I suppose one small step on the way." This took me back to July 2004, when I was almost at the stage when I knew what I needed to do to write my dissertation. I thought about the people I was working with: my advisors and a few of my friends from Minnesota, including one I hadn't thought of for a while. Then my thoughts turned to last month, when after a period of alternating thinking and worrying, I created a directory called "CAREER". This caused me to reflect on how I get large tasks done, and that starting big tasks with small symbolic-but-concrete gestures is an effective strategy for me. I wrote a new blog post to help me remember this, and when I was done with that I felt recharged and turned back to the writing.

Reminiscence serves a number of purposes illustrated in this (true) story, including reviewing pleasant memories, reflecting on and learning from past events, maintaining connections to other people, and helping to regulate moods (e.g., [32, 40, 176, 178]). Though reminiscence driven by regret correlates with negative psychological consequences [129], and in fact reminiscence is still a loosely defined concept in psychology [41], structured reminiscence often has beneficial effects as a therapeutic tool for the elderly (e.g., [31, 95, 166, 188]). Younger people also benefit from reminiscing [26], though they use it for different purposes than the elderly [179], and reminiscing across the lifespan often focuses on events from early adulthood in a phenomenon known as the "reminiscence bump" [101]. Therapeutic reminiscence is generally structured, but in everyday life, reminiscence is often spontaneously triggered through external cues such as photos and music [183], mementos [26, 143], topical themes [28, 188], and nostalgic cultural reminders [110].

Dreams of marrying memory with technology date back to the invention of computers [29]. Advances in ubiquitous computing have supported the development of "memory prostheses" for

both everyday needs (e.g., [114, 151]) and for assistive technologies targeted at the cognitively impaired and elderly (e.g., [104, 184, 185]), a line of work that stands to grow more important as lifespans increase. The rise of digital photography, combined with the power of photos as memory cues, has spurred work on understanding how people use, organize, and share photos (e.g., [5, 52, 93, 131]) and building tools to support photo sharing and storytelling (e.g., [6, 11, 51, 159]). Photos are only one memory cue, however. The Ambient Ink project display people's written notes as a screensaver to facilitate opportunistic reminding [94]. Other work inspired by tangible computing [100] looks to attach memory-laden content to physical objects (e.g., [70, 71, 163]), an approach supported by recent studies of how people use mementos and physical artifacts to reminisce [143, 182]. Life-logging research looks to capture, organize, and manage all of the sensory data and documents a person is exposed to [74, 75, 154] in the course of their entire life.

A theme in this work is that one of the goals is usually to develop a device that supports the capture or reproduction of memory-related data. A recent paper critiques the memory-as-data approach that many such technologies use, claiming they do not accord well with theories of autobiographical memory [43] that could inform the design of technologies for reminiscence [168]. In particular, memories are actively constructed and experienced differently every time they are remembered, especially in the case of autobiographical memories that are most likely to have value for reminiscing [152]. Empirical research using life-logging technologies points out this difference between recording and experience, suggesting that life-logging tools may work better as cues to trigger "knowing" that events happened, rather than seeking to support detailed "remembering" of those events [154]. Using specialized devices is also problematic because making and reviewing memories is not a planned task; rather, it is a continuing activity that is part of everyday life. Creating specialized devices to address these everyday computing activities fails when the device is not handy. More generally, building tools that demand people's attention does not accord with the vision of ubiquitous computing, where the technology fades into the background [1, 180].

4.2 Research activities around supporting reminiscence

My goal is to integrate technologies that support reminiscing with people's current practices. People write blog posts, take digital pictures, and update their Facebook and twitter statuses to record and share their thoughts, moods, feelings, and experiences. A straightforward technological approach, then, is to write variations on my personal blog-scanning text-messaging agent and declare victory. However, in the immortal words of Don Norman, "designers are not users", or at least, not typical users [139]. It is not clear that other people share my problem of failing to remember the past. In fact, it is not clear that for most there is a "problem" to solve; rather, the right approach may be to focus on how HCI can enrich people's existing activities [81]. Recent design stances urge designers to look beyond users' tasks, considering their values [69], experience [67], reflections [156] and fun [73] as important elements in the design process.

The research plan around reminiscence will thus unfold along two parallel tracks. The first explores how people reminisce and the role technologies might play in supporting reminiscence through a series of interviews and prototype deployments, loosely inspired by the idea of cultural probes [72]. The second examines how reminiscing behavior manifests itself in social media, learning to identify content that is likely to support reminiscence while modeling patterns of communication around this content. These tracks will then converge around the development of algorithms and systems that reuse user-created content to help both individuals and groups reminisce.

4.2.1 Track 1: Understanding and prototyping for day-to-day reminiscence

Supporting reminiscence raises a number of questions. What are people's current reminiscing practices, and what opportunities do they afford for designs that supplement them? What kinds of triggers prompt reminiscence? Which ones are most evocative? What roles do other people play in reminiscence? Can technologies effectively prompt, remind, or support spontaneous reminiscence? How much control do people want over these technologies, and how will they impact reminiscing practices and outcomes? Can these technologies encourage people to create memory triggers for themselves, for later in life?

The initial interviews I have conducted show that people are very positive about the idea of using technology to help them reminisce. People have expressed a desire for spontaneity and interruption, including requests for reminders delivered by text message, email, or less plausibly but more interestingly, fortune cookies. They also would like relatively frequent prompts, feel little need for control, want support for social aspects of reminiscing, and desire multimedia memory cues. We have deployed initial prototypes within the team to probe some of the issues raised by the interviews. Some informants have no easily accessible digital content, so we are exploring whether non-personalized prompts based on common reminiscence "topics" or themes [28] support reminiscing. The initial experiences with these prototypes suggest this is a viable approach as well. I hypothesize that personalized content will be more powerful for triggering reminiscence, but that both approaches will be effective, greatly broadening the potential base of users.

A credible observational study of reminiscing in daily life would be difficult because of the disruption and the cost of shadowing participants. Instead, we will deploy prototypes, using techniques borrowed from the ideas of cultural probes [72] and technology probes [98]. Our interviews are informing the design of the initial set of prototypes, which will both prompt reminiscence and provide regular, natural opportunities for participants to tell us more about their reminiscing practices. Collecting these reactions and evaluating "success" is a key issue in conducting this research. It is easy to fall into evaluation traps that focus on countable, separable aspects of using a system. It would be a mistake, for instance, to ask people to rate each memory trigger on a Likert scale for its emotional impact, duration of remembering, understandability, and perceived usefulness for other people. Some aspects of usability such as learnability, efficiency, and error rate are naturally captured by quantitative metrics [138]. However, the danger is that you get what you measure. Asking people to boil down complex reactions to numbers might tell us something about individual prompts but very little about how the technology impacts people's experience of reminiscing.

Instead, our evaluation will focus on collecting rich stories of use to drive a series of prototypes that will gradually address our questions around supporting day-to-day reminiscence. We would like to collect as much data as we can about how people interact with the system without creating distractions [97] or obligations to respond [118] that would disrupt the experience. Thus, our tentative plan is that the prototypes will send text, photos, and other memory cues through email, gathering data by soliciting replies to those mails. At the time we give people the prototype, we will tell them they can reply to any prompt with anything they would like to tell us. We are still deciding whether to give people direction or examples of the kinds of things we would like to know. My inclination is to avoid giving these examples because they may limit or shape people's reactions.

Occasionally, instead of a memory cue we will send a probing question that encourages people to reflect on the way they reminisce and their experiences with the technology. HCI practice sometimes adopts techniques such as cultural probes without attending to the context in which

these techniques are made and the assumptions that underlie them [22], which can dilute the value of the methods. Fortunately, I can draw on my colleague Phoebe Sengers for expertise around these techniques and design activities. Ideally, we will design probe questions that lead people to see both reminiscing and the technology from fresh perspectives, generating new design insights while challenging their, and our, ideas about the role reminiscing plays in their lives [156].

4.2.2 Track 2: Supporting reminiscence in social online spaces

The work above should help us answer our questions about the role of reminiscing in people's lives and point out promising directions for leveraging people's online activity. I expect that most such directions will point toward social applications such as connectedness [160], maintaining group memories (a complex topic rife with tension between individual and group interests, explored in the context of families in [112]), and maintaining common ground for shared understanding [38]. Connectedness, social presence, and awareness are important concerns in HCI and CSCW [99,125], both for group work (e.g., [19,57]) and for individuals (e.g., [55,59]), while maintaining relationships is an important function of reminiscence [32,176,178].

Thus, the second main track of research around reminiscence involves studying how it is practiced in online domains such as Facebook where the content is visible to other people and people interact around that content. This involves a number of questions. How often do people reminisce online? What kinds of content support reminiscence, and with whom? Does reminiscing strengthen online relationships, and how can we measure that? Can we build tools that support reminiscing online, increasing interpersonal reminiscence and encouraging the capture of context such as the people in a photo? Can we use context to effectively help people choose memories to reminisce about, either for themselves or for a group?

How do people create content to reminisce online? My plan is to choose representative samples of Facebook users and members of a blog community, then carefully examine their reminiscing activity. This will start with manual coding of their comments and posts, indicating whether a given piece of content serves a reminiscing function (e.g., [176,178]), the general topics of the content (e.g., people, places, events, times) and the type and media of the content (photo comment, wall post, blog entry). I will then adapt recent work by other Cornell researchers around topic classification [90] and sentiment analysis [24] that uses supervised learning methods to supplement human codings of text corpora. These methods often agree with human coders about as well as humans agree with each other, allowing us to label, then model, reminiscing behavior on a grand scale. These models will help us identify valuable content to use in later design activities.

How do social networks support reminiscence? Another important prerequisite for supporting social reminiscence is understanding how communication around reminiscing takes place. My recent work around understanding Wikipedia as a social network [53, 181, 187] considers the evolution of social networks across time as well as using coded behavioral information and visualizations to understand patterns of interaction in these networks. Here I will extend these techniques to explore the relationship between reminiscence and communication. In particular, I will explore ways to include artifacts such as photos and Facebook groups directly in the social network analysis, incorporating elements of the theory of object-centered sociality [33]. As with the idea of distributed cognition [91], this theory posits that people interact through "objects", though more

broadly construed to include groups, places, activities, and interests. One implication of this theory is that social network sites should treat objects as first-class parts of the system [61]. I will try to apply this insight to the tools of social network analysis, representing content in Facebook that people interact around, such as photos, groups, and events, as part of the network. The immediate goal of representing content as part of the network is to make sense of interactions around reminiscence. More generally, if this approach proves effective it will inform social network analysis methods, which too often abstract away context and content and focus solely on relationships.

Can systems and algorithms support social reminiscence, online and off? Track one and two converge here, combining insights from the prototypes with the modeling work to develop useful systems. The form of these systems will be driven by our findings. It might involve Facebook application that presents a group of friends with content that prototypes have shown to trigger powerful memories and that modeling has shown often trigger conversations online. The content could be chosen specifically to leverage events salient to the group such as the joining of a new friend, anniversaries, upcoming events of interest to group members, or content created by individual members. It might involve creating an environment for sharing, reminiscing around, and capturing metadata on photos and other content, perhaps incorporating motivational elements, a la the ESP Game [172]. Maybe it will even be a machine that bakes fortune cookies that contain powerful memory cues and invites groups of friends to lunch at a favorite Chinese restaurant.

In any case, the end results will: (a) be evaluated in the context of a much longer field trial than most technologies typically enjoy; (b) serve to validate our findings; (c) suggest new directions for studying reminiscence; and (d) have the potential to improve millions of people's lives.

5 Education that integrates research, computing, and theory

The students I am working with come from a number of disciplinary backgrounds. The educational challenge for this proposal is to help those students acquire the interdisciplinary attitudes and skills this work demands. One key component of the planned educational activities is to provide research and teaching experiences that help undergraduates value interdisciplinary approaches to research and system design, as well as sparking their interest in scholarship as a potential career. The other main component is based on courses that emphasize interdisciplinary approaches and skills that will be useful to students in this work, or any other work they choose to tackle.

5.1 Rich, authentic, interdisciplinary research experiences

Students I work with are not "interviewers" or "programmers"; they are *researchers* who choose questions, review literature, interview people, design experiments, create software, perform usability testing, conduct experiments, analyze data, and write papers. I encourage cross-training—programmer-interviewers, artist-analysts, psychologist-designers—welcoming students both from disciplines common to HCI (computer science, information science, psychology) and from less common cognates such as landscape architecture, symbolic systems, neuroaesthetics, and operations research. This approach has served me well in integrating quality interdisciplinary research with quality education; here, I propose three ways to strengthen this integration.

First, I will take undergraduate students to conferences in order to broaden their research experiences. Talking about research, seeing other cool research, and meeting other cool researchers

is one of the most stimulating aspects of being a scholar. I believe that giving students these experiences is likely to encourage students to consider academic careers, and that funding conference travel is a cost-effective way to bring students to scholarship. This grant deadline is particularly opportune because the CHI, C&T, ASSETS, and ICMI conferences are all located near Cornell in the next 18 months. I will also organize sessions at these conferences that help undergraduates make the most of the experience, meeting each other and talking about their experiences with more senior researchers from labs around the country.

Second, I will implement an "exchange program" with researchers at these other labs who have related interests but different strengths. My work the past two summers with students from other disciplines and universities has been valuable for everyone involved. My plan is to find opportunities for undergraduates to switch institutions for the summer, somewhat like a summer internship. These experiences will give students new research relationships and perspectives that may increase their motivation to pursue graduate education.

Third, I will help students consider the issues around interdisciplinary work through providing mentoring opportunities. Last spring, two Cornell undergraduates who had completed the HCI sequence and helped write a CHI paper [49] acted as mentors for project groups in the introductory HCI course. They guided groups through the semester-long keystone design and evaluation projects, advising students on potential project topics, work practices, and designs. All parties benefited: students in the class received more personal attention while mentors valued both working with other students and the opportunity to participate in interdisciplinary activities without the pressure of being typecast by their current skills. Funding from this proposal will allow me to expand this beyond what Cornell already supports.

Graduate students will also benefit from leading undergraduate project teams working on this proposal. Mentoring and directing novice researchers is an important but sometimes neglected aspect of graduate education, and as with the advanced undergraduates mentoring class project groups, this activity should benefit both mentors and mentees. I will also use lessons learned from the undergraduate experiences outlined above to develop a similar but more immersive program for graduate students in a future IGERT grant. Finally, I will aggressively pursue REU supplements and other funding to complement funds from this proposal.

5.2 Interdisciplinary approaches and skills in the curriculum

Student interest in HCI has grown tremendously. Enrollments in the introductory HCI course I have taught the last two years grew from 40 in 2006 to 55 in 2007 and 80 in 2008. This growth plus recent HCI faculty hires will allow me to develop courses that broaden our HCI curriculum while supporting the specific research activities of this proposal. Two courses will address the challenges of intertwining computation and social science, one focusing on the forces that shape behavior in systems, the other on working with the data these systems generate. Two other short practicum courses will complement the current introduction to HCI, providing a deeper look at skills relevant to social scientists on the one hand and computer scientists on the other.

5.2.1 Techno-social analysis of systems

This graduate-level Information Science course is open to students from all disciplines. Students will study how the social and technical aspects of social systems shape behavior, then use that understanding in design. The course is a seminar based half on readings and half on case studies.

Students will look at psychological, economic, and sociological forces such as theories of motivation, innovation adoption, and network effects. At the same time, they will look at how the design of these systems shapes behavior: how do the algorithms that process data, the interfaces that present it, and the social norms that govern the community affect what people do?

Students will achieve an understanding of how these forces work together and be able to reason about the design of social systems: suppose Wikipedia required real names or social news aggregators tried to mute their loudest voices? Students will also identify a set of social science theories that seem most useful for informing computational work. They will complete a series of small exercises that apply theory to design and analysis, present an original case study of an existing system, and develop a project plan with the goal of jump-starting a piece of publishable research. The case studies will inform the proposed research as well as serving as tangible outputs to be disseminated to the research and design communities.

Lessig's *Code* [121] and Benkler's *The Wealth of Networks* [16] are plausible choices for a foundation text, with supplementary papers added as needed. Phoebe Sengers teaches related courses with an emphasis on cultural critique while John Riedl teaches a more technical look at the social web; both will be useful resources as I build a course that walks the path between.

5.2.2 Managing massive, messy interaction data

This upper-division undergraduate course will be cross-listed between Computer Science and Information Science. The main objective will be to learn to manage, understand, and exploit the data people generate as they use social systems, with an eye toward using these data to support research. This involves acquiring and processing the data in ways that address research questions while enabling future research, exploring the data through statistical and visualization techniques, creating models that predict behavior, and working with algorithms for learning from the data.

Students will become familiar with a number of useful tools for working with data. They will use screen-scrapers and APIs to acquire data, discussing the legal and ethical issues involved. They will begin to learn the python scripting language and its libraries for data processing, analysis, and modeling. They will also use visualization tools such as ManyEyes [177], learn basic methods for modeling activity data, and receive introductions to useful machine learning algorithms. I will structure each topic as a two-to-four week module organized around a hands-on assignment. Students who already have skills in a given area will work on advanced projects or mentor other students. For a final project, students will be required to generate a dataset from a social media site, posing and answer research questions about that dataset. These datasets and analyses will be publicly disseminated, again, with the hope of leading to research projects relevant to the proposal.

Segaran's *Programming Collective Intelligence* [153] or the forthcoming *Collective Intelligence in Action* [4] are plausible foundation texts, while Lada Adamic's courses on data processing and Judith Donath's courses on social visualization will also be valuable resources.

5.2.3 A better introduction to HCI

A recent critique claims that focusing on easy-to-teach usability evaluation leads to a failure to effectively teach more difficult aspects of interaction design [79]. Cornell's introductory HCI course, INFO/COMM 345, is designed to be accessible to students in a number of disciplines. This is good for teaching interdisciplinary values, but does cause the course to focus on usability and other topics that have few discipline-specific prerequisites. I will design two one- or two-credit practicum

courses to be taken in parallel with 345. These courses will support and supplement the design activities in 345 with deeper, discipline-specific work on interaction design skills, such as learning qualitative methods for user research or interface development. These classes will be taught in a practicum style with minimal lecture, following constructivist theories [14] and active learning approaches [102] that seem particularly appropriate in courses with heavy skill components.

6 Work plan and institutional context

	2009	2010	2011	2012	2013
Motivating Contributions	Interface for current SB	SB2 architecture and prototypes	Finish SB2 (Spring)	Noise reduction in inferred profiles	Tie up loose ends; react to new directions and ideas suggested by earlier work.
	Modeling aspects of motivation	Algos for motivation	Deploy long-term in Wikpedia (Fall)	Deploy to other Mediawiki groups	
Supporting Reminiscence	Model content	Model interaction			
	Probe prototypes	Probe prototypes	Develop system	Deploy long-term	
Teaching	Fa:TSAS	Sp:345, Fa:MMMID	Fa:TSAS	Sp:345, Fa:MMMID	
Other Activities	Ugrad conference travel, REUs	Ugrad summer exchange, REUs	Ugrad summer exchange, IGERT	Ugrad, grad summer exchange	
Outputs	Initial WP datasets	Cleaned case studies from TSAS	Datasets from MMMID	Develop a comprehensive resource based on materials	

Figure 1: A provisional work plan for the activities in this proposal. TSAS: Techno-social analysis of systems, MMMID: Managing massive, messy interaction data, SB: SuggestBot

Figure 1 shows the planned schedule for this work. The research activities in both domains start with parallel work on interfaces and modeling content, moving through algorithmic and system development to long-term deployments of the systems. The teaching plan is structured to generate useful datasets and case studies relatively early in my career. Other activities include the planned research experiences for undergraduates and graduates, along with seeking additional funding to support those activities. The outputs show planned deliverables beyond the usual dissemination of results through papers; toward the end of the grant period I expect to have enough material to produce a comprehensive resource for doing design of social systems driven by data and informed by social theory. Note that the fifth year is intentionally left blank. I expect that some activities will take longer than planned; this gives those activities room to breathe. It is also likely that lessons from our initial system deployments in years three and four will suggest ways to improve the systems, and that we will want to re-deploy them in the communities. Finally, I expect other opportunities to arise based on work from the first four years, as well as from broader changes around people's attitudes toward existing systems and new technologies that arise.

The educational and research objectives of this proposal closely fit the Information Science mission at Cornell: "Information Science brings together faculty, students and researchers who share an interest in combining computer science with the social sciences of how people and society interact with information." Cornell has invested substantial resources in developing the IS program, allocating faculty lines in multiple departments, endowing a directorship, and providing lab space

and computing resources. The soon-to-be-built William H. Gates Hall will substantially expand the resources available to IS and bring together researchers from the broader Faculty of Computing and Information Science. Further, Information Science has brought in several HCI and CSCW researchers in the past year, including Jeremy Birnholtz, Sue Fussell, and François Guimbretière.

I am the first faculty member to be hired directly into Information Science precisely because I have a record of doing work that integrates technical and social perspectives. Over the last two years here as a visiting assistant professor in Communication, I have formed excellent working relationships with researchers across Cornell including Jon Kleinberg, Phoebe Sengers, Jeff Hancock, Michael Macy, Dan Huttenlocher, and Geri Gay [49,53,120,157,181]. I also continue to have strong relationships with experts in recommender systems and social psychology from my work with John Riedl, Loren Terveen, Bob Kraut, and Sara Kiesler [44–46, 50, 123, 130, 140]. These collaborators, and the broader resources available to Information Science at Cornell, will provide valuable support to complement the resources requested by this proposal.

7 Summary: educational, intellectual, and social merits

This work will have important educational, intellectual, and social impacts. The research activities will improve undergraduate education and motivate future scholarly activity. The new coursework will give HCI students a more thoughtful perspective on the relationship between human behavior and system design. It will also attract students who enjoy this work and further the research goals of the proposal. A disproportionate number of these students so far have been female, suggesting that choice of domains and the combination of people- and technology-work may serve to recruit women to participate in HCI and computer science.

Intellectually, the deep intertwining of computational work with social science theory will push on the boundaries of how HCI does interdisciplinary work while generating a number of useful research outputs. One outcome will be the development of the useful models, algorithms, and systems developed around the domains of work. Both the research and the process of teaching people how to do it will result in insights about using social science theory both for data anaysis and system design that will be shared with practitioners and researchers. The activities in the proposal will also generate rich behavioral datasets to share with other researchers and support this kind of work [18].

Finally, the work will benefit society broadly. Encouraging contribution to knowledge-producing communities will promote individual learning, increase the quality of the goods produced [122,164], and support involvement in communities [148]. The work in reminiscence also promises broad social impact. As a scholar, I value good research; as a teacher, I value good education; as a person, I value contributing to society. This work has the potential to achieve all three. I hope that you agree, and thank you for the time you have taken to consider this proposal.

Results From Prior NSF Support

I have not yet received NSF support as a PI or Co-PI, though I gratefully acknowledge the opportunities NSF funding gave me through grad school to pursue the work described above.