Understanding and Increasing Social Production Online

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mikhil Masli

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Loren G. Terveen

September 2013

Acknowledgments

Throughout this thesis, I use the word “we” to represent the researchers carrying out the research described. Not only is this the convention in my field, Human-Computer Interaction and Social Computing, but it is an explicit acknowledgment of the crucial role my colleagues and co-authors have played in the pursuit of this research. I would like to thank my adviser, Dr. Loren Terveen for his faith in me and the guidance he provided throughout the creation of this thesis. I’d also like to thank my committee members, past and present, Dr. Jaideep Srivastava, Dr. John Riedl, Dr. Mark Snyder, Dr. Brent Hecht, Dr. Yuqing Ren and Dr. Nicholas Hopper, who provided timely advice and guidance, without which, this dissertation would not have materialized. I also want to thank all my co-authors and mentors, Dr. Joseph Konstan, Dr. Reid Priedhorsky, Katherine Panciera, Landon Bouma, Andrew Owen, Dr. Edward Cutrell, Dr. Werner Geyer, Casey Dugan, Beth Brownholtz, and Dr. David Millen for partnering me with on several research projects.

Along with professional colleagues, I would also like to specially thank my family and friends for their support. I’d like to thank my father, Nandkishore Masli for providing the inspiration to pursue a Ph.D. and I wish he was here to see me achieve it. I would also like to thank my mother, Sandhya Masli and aunt, Dr. Sharmila Masli for setting the right examples and guiding me to where I am now. I am thankful beyond words to my girlfriend and soon-to-be-wife, Dr. Amruta Bamanikar for putting up with me through the last five years and relentlessly staying by my side. Finally, I would also like to thank all my friends and classmates during my time as a graduate student for helping maintain an excellent environment for me to grow as a student and researcher.

Thirdly, I would like to thank all the members of the GroupLens research lab and the Cyclopath team for creating an incredible atmosphere to do research and be creative and innovative, specially Angela Brandt and Alison Amundson for helping manage us. I would also take this opportunity to thank all the members of the Cyclopath community, without whom the resource would not have sustained and my dissertation would not have materialized. The department of Computer Science and Engineering is an excellent place for graduate education, and I would like to thank our department head Dr. Vipin Kumar for excellent leadership and all the support staff, specially Georganne Tolaas, our graduate programs coordinator.

Thank you all!

Dedication

To Nandkishore Masli and John Riedl, two fathers who never cease to guide.

Abstract

Social production is a phenomenon that lets a large number of people work together to produce common resources. The advent of the Internet has enabled this phenomenon to scale to large proportions, with participants coming from a wide variety of places and demographics. Many such communities have become a part and parcel of our daily lives today. For example, for reference, we turn to the collaboratively-built encyclopedia Wikipedia, and for getting our questions answered, we turn to Q&A sites such as Yahoo! Answers. The last few years have seen this phenomenon merge ways with advancements in online mapping technologies, giving rise to a new class of applications that may be termed geographic social production communities. Resources such as OpenStreetMap are becoming increasingly popular and even Google Maps has added a feature called MapMaker to help crowd-source geographic data. From a human-computer interaction and social computing perspective, understanding how social production happens in such communities and findings ways to increase and improve it are both important and interesting research topics. Accordingly, this dissertation makes contributions in both these directions. While the research and findings of this dissertation are particularly applicable to geographic social production communities, they can be extended to other applications as well.

We first study how contributors in Cyclopath—a geographic social production community—participate by analyzing behavioral log data using visualization and statistical methods. Specifically, we investigate how Cyclopath contributors *specialized* in the tasks they choose to do. We find evidence for specialization by *work type*: Most users edit a single type of map feature, such as points of interest or roads and trails. We also see a user life-cycle effect: as users gain experience, they specialize in editing roads and trails. Our findings suggest more effective ways to organize social production interfaces, compose units of work, and match them to users who want to help. However, matching tasks to people is, at its core, dependent on compliance with requests to contribute on the part of the users. Therefore, as a next step, we investigate into techniques that may help us increase the chances of this happening.

Social psychology offers several theories of potential use for designing techniques to increase user contributions to online communities. Some of these techniques follow the "compliance without pressure" approach, where users are led to comply with a request without being subjected to any obvious external pressure. We evaluate two such techniques—*foot-in-the-door* and *low-ball*—in the context of Cyclopath and report that while both techniques succeeded, low-ball elicited more work than foot-in-the-door. However, we find that these effects were one-shot and contribution levels drop back to pre-request levels soon after. We also find that while these techniques have the potential to succeed in the short term, they could cause long-term harm, because users may feel manipulated and lower their sense of belonging to the community. We believe that one of the reasons for this was the inherent *unnaturalness* in the process: users were being coerced to contribute, i.e., do something that they typically do not. Therefore, as a next step, we explore ways of eliciting contributions by leveraging natural processes.

One such natural process is *information consumption*. Accordingly, we explore the feasibility of using the act of *consuming* information as a gateway to *contributing* information; specifically, we investigate semi-automated means to extract useful information from standard types of explicit *user feedback*. We analyze naturally occurring textual route feedback in Cyclopath, finding that the feedback was rich in information such as bikeability ratings, tags and notes that are useful to improve the system's route finding and navigational assistance capabilities. We also present a technique to extract such information by engaging users in dialogue immediately after they obtain a route.

Finally, we explore the utility of online social production through the lens of an application that has not been well explored before: how citizen-driven online social production helps the government activity of transportation planning. We described the design of a novel route analysis tool based on Cyclopath to assist transportation planners to make better planning decisions. We highlight the advantages of using a online social production system over other, similar ones through a real-life usage scenario. We believe that the results and ideas in this dissertation are applicable to a broad class of online social production systems.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Social Production Online

1.1.1 What is Social Production?

Social production is a phenomenon [7, 35] that lets loosely connected people work together to produce shared services and resources—termed as artifacts of lasting value [21]—that are useful to more than those who created them. This phenomenon has been in existence in offline, real life communities for centuries, but the advent of the Internet in the last two decades has taken it to a new level. The Internet has suddenly abolished geographical boundaries and has enabled people from different parts of the world to contribute to, maintain and consume a shared resource with relatively little effort and cost. Social production in the online world has created some of the most powerful publicly-maintained resources and services of modern life, and enabled new forms of communities to thrive around them.

Such communities have become a part and parcel of everyday life. For example, recommender systems like Amazon leverage users' ratings of consumer products to enable personalized recommendations. Q&A sites like Yahoo! Answers[1] and Stack Overflow[2] form knowledge economies, where users spend

[1] answers.yahoo.com

[2] stackoverflow.com

points to ask or boost the priority of questions and earn them for answering: If our car is broken, we ask for advice on how to fix it on Yahoo! Answers, and if we cannot make our computer programs work, we turn to Stack Overflow. Breaking news often appears on Twitter and Wikipedia before being covered by traditional news organizations. The last few years have seen the phenomenon of online social production merge ways with advancements in online mapping technologies, giving rise to a new class of applications that may be termed geographic social production communities. Resources such as OpenStreetMap are becoming increasingly popular and even Google Maps has added a feature called MapMaker to help crowd-source geographic data. Consequently, information such as available parking spaces, cycling and running routes and several other geographic information is shared and consumed via Google Maps-based mash-ups. In other words, social production has played an important role in improving our everyday lives.

Going beyond everyday conveniences, healthy social production online is often found to be a key force driving the development of offline societies themselves. The importance of social support in patient communities is a well-established phenomenon that is increasingly taking the online route. Online communities like PatientsLikeMe[3] and Inclusive Planet[4] are increasingly being used to complement offline social support for people with health ailments or disabilities. Cities and local governments are using social media to interact with citizens and engage in a two-way dialogue with an aim to enriching civil life. For example, New York City hired a "Chief Digital Officer", Rachel Haot, to spearhead efforts to integrate social media more tightly into the urban living experience with projects such as "Made in NY"[5]. Lastly, we have seen several examples where collection and exchange of information online has improved crisis response and saved lives [77]. The Ushahidi project[6] has enabled crowd-sourcing and social production to be used for social activism, public accountability and crisis response, including the 2007-2008 elections in Kenya

[3] www.patientslikeme.com

[4] www.inclusiveplanet.com

[5] wearemadeinny.com

[6] www.ushahidi.com

and earthquakes in Haiti (2010) and Christchurch, New Zealand (2011) [76]. Social media has helped communities in Iraq [63] as well as Egypt [1] be more resilient to the crisis around them.

In short, social production online has tremendous potential to complement the offline lives of societies. Much of this potential is being tapped using contemporary research and efforts are on in full swing to improve existing social production systems and design better ones. This dissertation contributes to this effort by studying online social production in the context of geographic social production communities.

1.1.2 How does Social Production Happen Online?

Understanding how social production occurs online happens has been a focus of scholarly research in both computer science and the social sciences for the past few decades. A lot of research has studied the participants in social production online, trying to answer questions such as who participates, why and how. A majority of the contribution in online social production communities is done by a few contributors [84]. Most participants are information consumers, some of whom take on contribution and organization-related activities as they gain more experience in the community [82]. For example, in the context of Wikipedia Bryant et al. found that Wikipedia editors shifted concerns as they became more experienced, evolving from a focus on topics about which they had some personal expertise to taking on different types of "community maintenance" tasks, e.g. monitoring for vandalism and enforcing policies like "Neutral Point of View" [10]. Some of those who never contribute (termed as "lurkers") might just be free-riding [51] whereas others might be learning about the community [72].

Online social production communities often survive and sometimes even thrive using a few dedicated set of volunteers [19]. These volunteers could be grown internally from within the community and assigned formal leadership roles [15, 8]. For example, the Apache project has a strong merit based structure wherein a developer rises in rank as his contributions increase and is more valuable. This rise in rank or any other form of formal recognition that a person could receive from the community is a strong motivation for continued participation [45].

Along with studying the participants, scholarly research has also studied the social production infrastructure itself, trying to identify what works and what does not. To support healthy social production online, attention has to be given to both human and system/interface factors. Support mechanisms and scaffolding like maintaining past records of conversations, an easy-to-use interface, rating systems, top contributor lists, message filtration systems etc. play an important role in keeping an online community healthy [8]. Also, most experts agree that holding offline community activity helps boost online community activity [20, 54, 57].

1.1.3 Motivations and Barriers to Contribution

A vast body of research has investigated the reasons why users participate and contribute in online social production systems. Filling observed gaps or fixing broken computations and making the system work better for themselves is one of the most important motivations for participants to contribute information. For example, Bryant et al. found that new users primarily use Wikipedia for information gathering and identify problems and mistakes in passing and fix them [10]. In the context of open source software development, Hertel et al. discovered that the most significant predictors of engagement in the development of the open source Linux project were gaining identification as a Linux developer and pragmatic reasons such as to make the software work for their own, specific needs and for career-related benefits [48]. Interviews with users of Cyclopath, a geographic wiki (introduced in more detail later), revealed that one of the main reasons why they edit the map and submit bike-friendliness ratings of roads is to have Cyclopath compute the route they desire [79]. An important motivator for workers to perform tasks on crowd-sourcing platforms such as Amazon's Mechanical Turk is the monetary benefit behind them.

Participants also contribute information to make the system a better resource and benefit the community that uses it. Experienced users of Wikipedia (a.k.a. *Wikipedians*) contribute to make Wikipedia a better public resource and serve the larger community that reads it [10]. Similar findings were observed among experienced Cyclopath users: their motivations evolved from simply benefit themselves to benefiting the cycling community at large as they gained experience [79]. Finally, participants contribute because of several intrinsic motivations as well. For example, enjoying the coding process and the mental stimulation it provides is an important motivation for developers to contribute code to open source projects [59]. Motivations to contribute to OpenStreetMap ranged from belief in free availability of mapping data to feeling a part of a community [41]. Research on Everything2, a peer production community that encourages participants to share articles and stories, also revealed the effect of habit: users often participate due to a habitual response developed out of routine [108] whereas general volunteer motivations, pro-social behavioral history, and community-specific motivations were found to be predictors of both quantity and nature of participation of users of MovieLens, a movie recommender system [34].

On the other hand, research has discovered several barriers to contributing. For example, a survey of MSN bulletin board users revealed the following reasons for not contributing: not needing to post; needing to find out more about the group before participating; thinking that they were being helpful by not posting; not being able to make the software work (i.e., poor usability); and not liking the group dynamics or the community was a poor fit for them [74]. Similar research on a few Usenet groups found that lurkers did not contribute because they wished to preserve their privacy and remain anonymous, had work-related constraints on sharing of information and online participation or were shy to post [73]. In enterprise settings, employees may hesitate to share information out of fear of criticism, or of misleading community members through sharing possibly inaccurate information [4].

A special class of barriers are those that newcomers face. These are participants who want to begin contributing to the social production community, but face difficulties doing so. For example, a qualitative study of readers and infrequent editors of Wikipedia revealed strong negative perceptions that prevented them from identifying with active contributors. Other research on

Wikipedia has reported that an increasing number of newcomers editing in good faith are having their edits reverted by automated vandalism-fighting tools, originally developed to maintain the quality of information in the encyclopedia [42]. These rejections have a severe negative effect on newcomers who are likely to stop editing and never return [44]. Another noteworthy set of barriers are those faced by contributors belonging to a minority demographic. For example, female editors on Wikipedia have rated their satisfaction with editing lower than their male counterparts citing conflicts and hostility [60].

1.1.4 Increasing Social Production

Despite the motivations explored above, under-contribution is a major problem for several online social production communities. For example, over 80% of collaboratively-made animations on the Newgrounds[7], are incomplete. Even popular ones like Wikipedia are seeing a decline in their active contributor counts. The success of community-maintained, public resources lies in the contributions of its many members who may volunteer their knowledge, time and effort for the greater good of all. As a result, much scholarly research has studied techniques that can motivate participants to graduate beyond consumption and contribute information into the shared resource.

Perhaps, one of the most straightforward ways of increasing social production is to provide monetary awards to contributors. For example, About.com first attracted subject experts or "guides" to create and publish content under its banner by paying them a small fee, and later sharing part of the advertising revenue gathered with them [105]. Amazon's Mechanical Turk provides a scalable infrastructure to enable task creators to disburse monetary payments to the workers who complete the task. However, in many social production situations, this is either not a viable or a scalable solution. Another basic solution is to make the system easy to use for newcomers and experts alike. Given that most of the contributors are either newcomers or inexperienced, this includes doing sufficient usability testing to ensure that participants do

[7] www.newgrounds.com

not face any hurdles while attempting to contribute information. In situations where the social production system is setup to serve as a support group to people with certain disabilities, accessibility should be ensured and tested for: One of the reasons why Inclusive Planet found success as a community for the visually-impaired because it was easy for users to consume as well as contribute content via screen-readers. Similarly, given that most of the contributions are submitted by a few highly active participants, the system must provide specialized tools to enable such experts to operate efficiently and in a frictionless manner.

Beyond the above straightforward solutions, prior research has investigated and evaluated several infrastructural innovations designed to increase social production in online communities. For example, communities like eBay and Stack Overflow award points and badges to encourage specific forms of participation [2] and enable participants to build a reputation or a name for themselves within the community by contributing information. Sometimes, these reputation scores unlock special areas of the community; for example, in Stack Overflow, one needs a minimum of 15 points to be able to vote an answer higher or lower and at least 10,000 points to gain moderation capabilities such as deleting a question[8]. Similarly, gamification of the contribution tasks is another popular technique that appeals to the fun-related motivational factors; the most notable one being the ESP game which helped label images and related "games with a purpose" [100]. Another infrastructural technique is that of personalization: Intelligent task routing- based systems that assign contribution tasks to users who are likely to be familiar with the topic and type of the task have been shown to boost contributions [22, 85]. Past research has also shown that visualization of information can also be manipulated to improve community health and boost member participation [98].

Social psychological research has been a strong source of motivation for designing techniques to boost contributory participation in online communities. A lot of empirical research has evaluated techniques derived from social psychological theories and models and have succeeded in showing boosts in

[8] stackoverflow.com/help/privileges

contribution levels. For example, the Collective Effort Model, proposed by Karau and Williams [51] to explain the phenomenon of *social loafing*—people exert less effort on a collective task than they do on a comparable individual task—inspired Ling et al. to carry out a series of experiments evaluating the effects of interventions such as informing users of the possible uniqueness of their contributions, their similarity to others within the community, and benefits imparted to themselves and others [62]. Ling et al. also evaluated theories of goal-setting and found that while goals that were too small or too large did not produce increased contribution, intermediate levels did. Experiments based on the Social Identity Theory and the Self-Categorization Theory have shown that a shared group identification positively influences motivation to contribute to common information pools [31]. Harper et al. studied the design pattern of social comparisons —presenting a comparison of one's participation with others within the community—and found that while it succeeded in increasing contribution by focusing user attention to particular system features highlighted, it did not succeed in raising the overall interest levels in the system [47]. Other researchers have used social psychological research to develop new designs for online social production systems [25, 28].

1.2 Organization of this Thesis

This dissertation contributes to the on-going efforts of understanding online social production and designing and evaluating techniques to increase it. All of the research within the next four chapters has been conducted in the context of Cyclopath, a live social production system. The following sub-sections provide an introduction to Cyclopath and an outline of this dissertation.

1.2.1 Research Platform: Cyclopath

Cyclopath (http://cyclopath.org, Figure 1.1) is a geographic wiki (or *geowiki*) offering route-finding services for bicyclists in the 7-county metropolitan area of Minneapolis-St. Paul, Minnesota, USA, an area of roughly 8,000 square kilometers and 2.3 million people. It went live in the summer of 2008. The

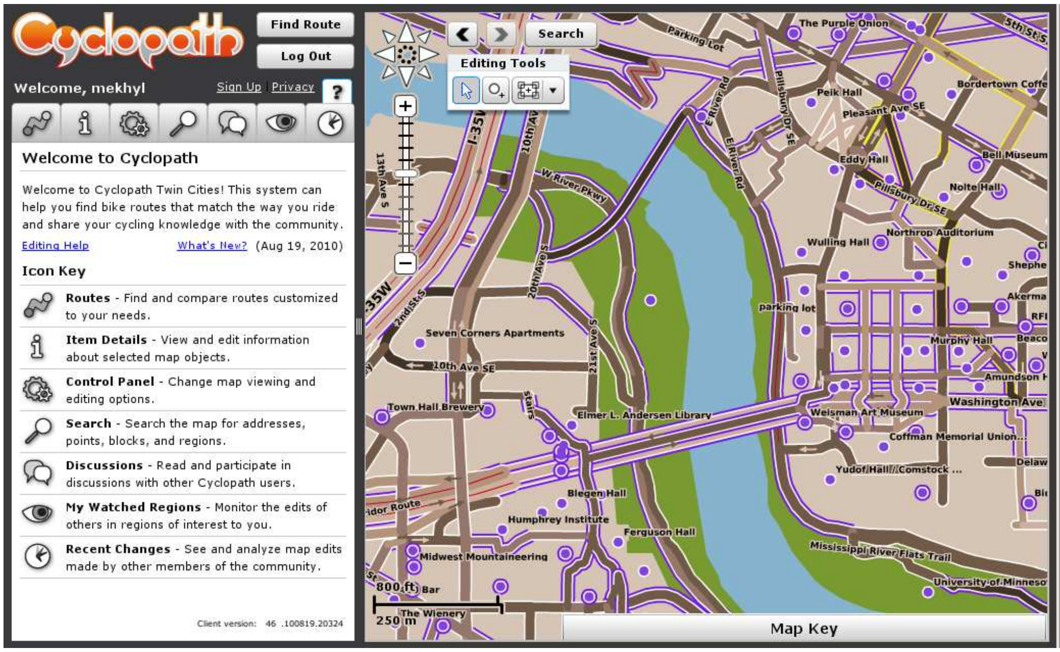

Figure 1.1: The Cyclopath geographic wiki. The right side of the interface is the map which shows roads, points, and regions with highlights if they have annotations. The left panel is used for a number of associated functions including editing properties of the item currently selected in the map.

database was populated initially with road and bicycle trail data from the Minnesota Department of Transportation (Mn/DOT). The initial release supported the map features blocks, points, and notes. Tags were added in April 2009, and regions were added in November 2009.

Users contribute to Cyclopath by editing (any number and any type of) map features, then clicking "Save Changes." This sends the set of edits to the Cyclopath server, which saves it as a *revision*. The Cyclopath server logs various information about each revision, including who did it (username if available and IP address) and a timestamp. The results of the revision are immediately visible to all Cyclopath users. Also, since Cyclopath is a wiki, all prior map states are retained; users can monitor the Recent Changes List for revisions of interest and revert any that are problematic.

The following types of map features can be edited in Cyclopath (the counts are as of April 16, 2013):

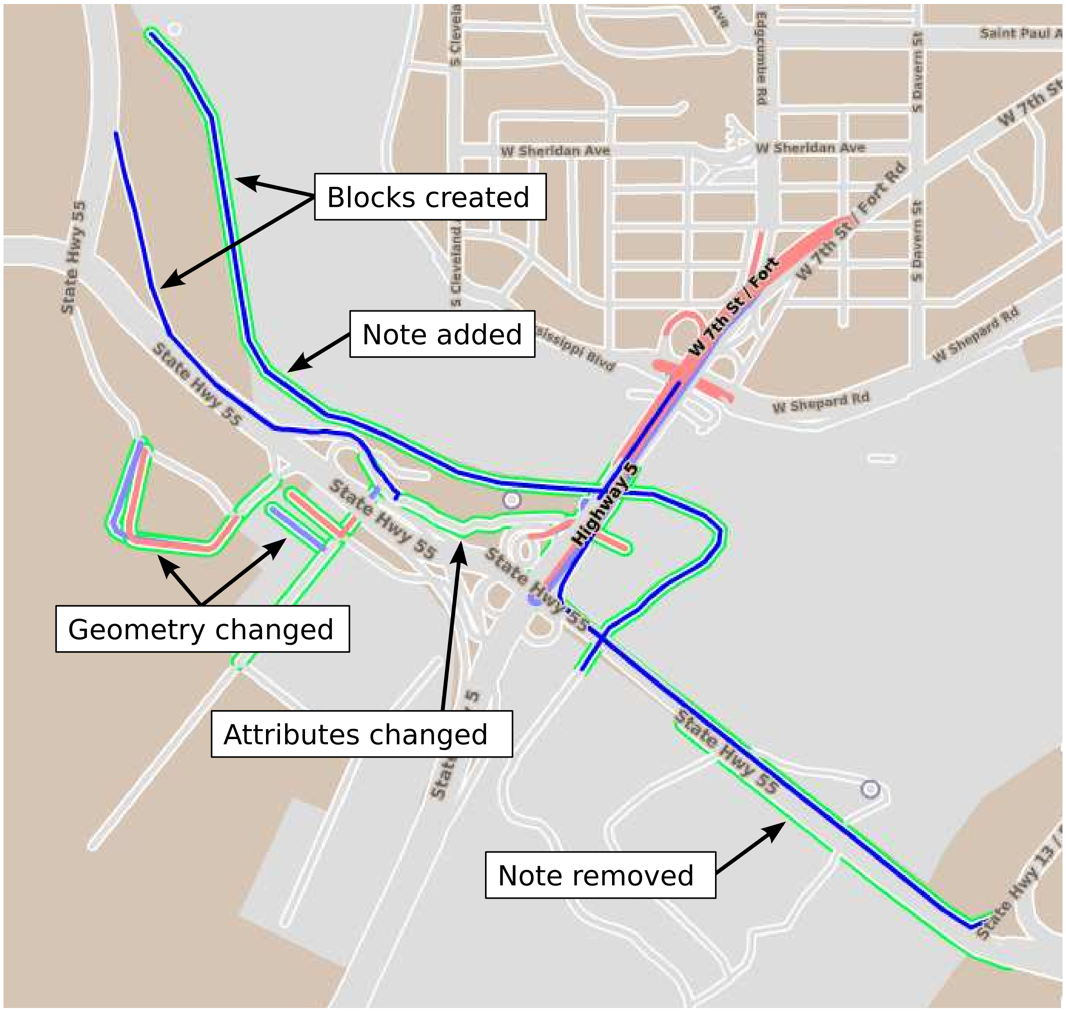

Figure 1.2: Changes in one revision, with some highlights indicated. New blocks are shown in dark blue; changed geometry is indicated with light red (old geometry) and light blue (new geometry); and changes to non-geometric attributes and notes are indicated with a green outline. This revision added 7 blocks and changed 19 (7 geometrically, 6 in non-geometric attributes, and 6 in both), added 5 notes to 5 blocks, and removed 4 notes from 3 blocks. Figure from [83].

- *Blocks or Road Segments* – 156,373 atomic segments of the roads and trails that make up the transportation network, e.g., the block of Union St. between Beacon St. and Washington Ave.

- *Points* – 3,083 points of interest, e.g., TCF Bank Stadium.

- *Regions* –347 cities, neighborhoods, and other defined geographic regions, e.g., Marcy-Holmes, a Minneapolis neighborhood.

- *Notes* – 2,537 text notes attached to 6,690 blocks, e.g. "icy during the winter".

- *Tags* – 826 brief text labels attached to 24,040 blocks, points, and regions, e.g., "bumpy" (attached to a block).

We adopted Cyclopath as our research platform for this dissertation because of two important reasons: First, Cyclopath is not the only wiki-based social production system. Wikipedia is the one of the most well-known deployments of the wiki technology. Similarly, several smaller wikis are based on the Wikia platform and even Q& A sites such as Stack Overflow employ wiki-like editing patterns to help the community create better answers to questions in a collaborative fashion. Second, Cyclopath is not the only geographic social production community either. It is similar to several web sites that combine a map-based interface with an open content model. Community-focused sites such as FixMyStreet[9] and SeeClickFix[10] let locals plot the location of potholes and similar problems on a map. OpenStreetMap[11] is a large ongoing effort to build a worldwide street map using the wiki model, and Google Map Maker lets users directly edit Google Maps data (in some countries) and submit those changes for inclusion in the public map.

[9] www.fixmystreet.com
[10] www.seeclickfix.com
[11] www.openstreetmap.org

1.2.2 Chapter Summaries

The contents of this dissertation are organized into four main chapters. At a high level, Chapter 2 describes research that furthers our understanding on how participants contribute information in an online social production community by analyzing patterns in contributory behavior, Chapter 3 and Chapter 4 cover research that evaluates techniques to increase social production online, and Chapter 5 describes how user participation in a system like Cyclopath can be instrumental in making urban areas smarter and more efficient. Below are brief summaries of these chapters:

- *Chapter 2 – Understanding Contribution and Contributor Specializations* presents research that studies how certain patterns in contribution are more common that others: despite the freedom available, most contribution units contain a single type of work, and most contributors tend to specialize in particular types of work. The chapter concludes by recommending various ways of using this understanding of contribution patterns to potentially increase contribution, such as intelligent task routing.

- *Chapter 3 – Increasing Contributions through Compliance with Requests* builds on the design implications from Chapter 2 and evaluates *foot-in-the-door* and *low-ball*—two of the most well-known techniques from social psychological research that increase compliance with requests, the core of intelligent task routing—through a live field experiment and a follow-up user survey. The chapter ends with the conclusion that while these techniques may work in the short-term, they might cause more long-term harm because they attempt to coerce users into doing tasks that they would otherwise not do.

- *Chapter 4 – Increasing Contributions through Compliance with Requests* reports research attempts to overcome the limitations of Chapter 3 by trying to leverage a natural activity—information consumption in the form of route-finding—to motivate contribution. The approach followed is to have users drag routes to suit their needs when dissatisfied with the computed route, automatically recognize the road segments that caused them to do so using machine learning techniques, and engage them in a dialog that enables them to make simple contributions about their route.

- *Chapter 5 – Towards a Smarter City: Utility in Transportation Planning* the highlights the utility of online social production in a novel context by describing how user contributions to a geographic social production community such as Cyclopath can be instrumental in making urban areas smarter and more efficient. It presents a route analysis extension of Cyclopath that transportation planners can use to make better planning decisions.

Finally, Chapter 6 summarizes the contributions of this dissertation and discusses some future work.

Chapter 2

Understanding Contribution and Contributor Specialization

2.1 Introduction

Contributors to online social production systems often specialize in the work they choose to do, whether by *topic* (e.g., some users answer questions about cats while others address cooking) or by *work type* (e.g., some Wikipedia users prefer to patrol for vandalism while others fix typos). We believe that understanding these patterns is important because it has implications for designing both the user experience and policies of open content communities.

We studied specialization in the context of Cyclopath. The general specialization dimensions noted above have direct analogues in a geographic context. *Topic* translates to *geographic location and extent*: users can specialize in different geographic areas (e.g., one user might edit in the Cedar-Riverside neighborhood, while another might focus on the suburb of Richfield) as well as the shape of the areas they edit (e.g., one user might focus his or her edits in the area near home, while another might edit along a favorite bike path). Similarly, *work type* corresponds to the *type of map feature* edited. In Cyclopath, users can edit roads and trails (the *blocks* that form the transportation network), points, and regions as well as the notes and tags that can be attached to these features. In this chapter, we study the latter, i.e. specialization by work type by framing the following research questions:

RQ1. **Specialization of Contributions.** *Considered in the aggregate, are revisions in Cyclopath – also the units of work in other wikis – biased toward any particular type of work?* Yes. We found that despite the freedom available, a majority (80%) of revisions consist of a single work type, with byway editing being the most popular. In those that do not, we see a definite object+annotation combination.

RQ2. **Specialization of Contributors.** *Do individual Cyclopath users specialize by work type?* Yes. A majority (65%) of users specialized in one specific type of work. Block editing was by far the most common specialization, even though this is the most difficult editing task.

RQ3. **Change in Specialization.** *Does specialization by work type change as users gain experience?* Yes and no. We saw interesting changes as users gained experience: although a majority of users (65%) did not change their specialization, a large minority (35%) did, and they tended to either transition to specializing in the important but difficult task of block editing or to diversify their editing and become generalists.

2.2 Related Work

Much research has analyzed the different roles of users in online communities, and in social production systems specifically. A fundamental finding is that participation is highly unequal: a very low proportion of "power" or "elite" users accounts for a very high proportion of participation [56, 84]. However, users specialize in ways other than simply the amount of work they do, and prior work has addressed the same dimensions of specialization that we do: topic and work type.

The basic organization of online communities reflects the obvious fact that different people are interested in different topics. For example, Usenet groups were defined for particular hobbies, television shows, and rock bands. However, even within a particular community, different users are interested in and knowledgeable about different topics. For example, Demartini observed that Wikipedia editors specialize in certain topics, then developed algorithms that analyze user edits to create topic expertise profiles [24], and Cosley et al. developed algorithms to match users with tasks in topics with which they were familiar [21, 22].

Research in a variety of communities has found that users specialize in their participation. In online discussion forums, Turner et al. [97] and Welser et al. [104] identify different roles that users assume, notably "Question Person" and "Answer Person" and try to build models to predict them using users' patterns of communication. In the context of Wikipedia, Welser et al. [103] mapped out various social roles that contributors can play, such as technical editors, substantive experts, vandal fighters, and social networkers. Bryant, Forte, and Bruckman [10] found that Wikipedia editors shifted concerns as they became more experienced, evolving from a focus on topics about which they had some personal expertise to taking on different types of "community maintenance" tasks, e.g. monitoring for vandalism and enforcing policies like "Neutral Point of View." Participants in open source software development may also be organized into implementation, interaction, group and design-oriented roles, each providing different utility to the project [5]. Further, most developers contribute to a few specific modules within a large open source project [101].

2.3 Study Design

At the beginning of Fall 2010, we carried out a quantitative study to investigate the nature of work type specialization in Cyclopath. We believe this extends prior work in several ways. First, in an attempt to move towards generality, we study the concept of specialization in a context that is qualitatively different and at a scale that is more real-world than an edge case like Wikipedia, which is the platform for most prior work. Second, we extend prior research on task specialization by studying work type specialization in a wiki context and how it varies over users' life-cycles. Third, we map our findings to clear implications for designers of commonly-found social production communities.

We analyzed usage data from the initial release through September 9, 2010. This dataset contains a total of 12,311 revisions. Of these, 10,777 were made by 544 registered users, and the remaining 1,534 were made anonymously. We focused on publicly visible contributions, and hence, do not consider ratings (which are private contributions) or viewing activity and route requests (which are private and not contributions).

Our analyses are organized by our research questions. For each question, we describe the procedures used and the results found.

2.4 RQ1: Specialization of Contributions

Procedure. As described earlier, a Cyclopath revision can consist of edits to multiple map features of different types, such as blocks or road segments, points, regions, tags and notes. In this section, we examine the types of map features that actually are edited together in revisions. We count a revision as a block revision if it edited only blocks. Similar terminology also holds for revisions editing other types of map features, like point, region, note, and tag. We also count revisions of mixed types, e.g., block+note revisions.

For this analysis, we ignore the number of map features edited in a revision. For example, a revision modifying 10 blocks and another modifying 2 blocks both count as block revisions. A revision with 1 block edit and 5 note edits and a revision with 5 block edits and 1 note edits both count as block+note revisions. We ignored the number of edited features in this analysis because we are concerned only with co-occurrence, not frequency.

We also consider revisions by registered and anonymous users separately. While prior work on Cyclopath shows that some revisions done by anonymous users can be attributed to registered users [80], the number of such revisions is low, and thus we do not apply that attribution process here.

Results. Figure 2.1 shows the number of revisions modifying particular map features and combinations of features. The results for registered and anonymous users are presented separately. Tags and regions were introduced

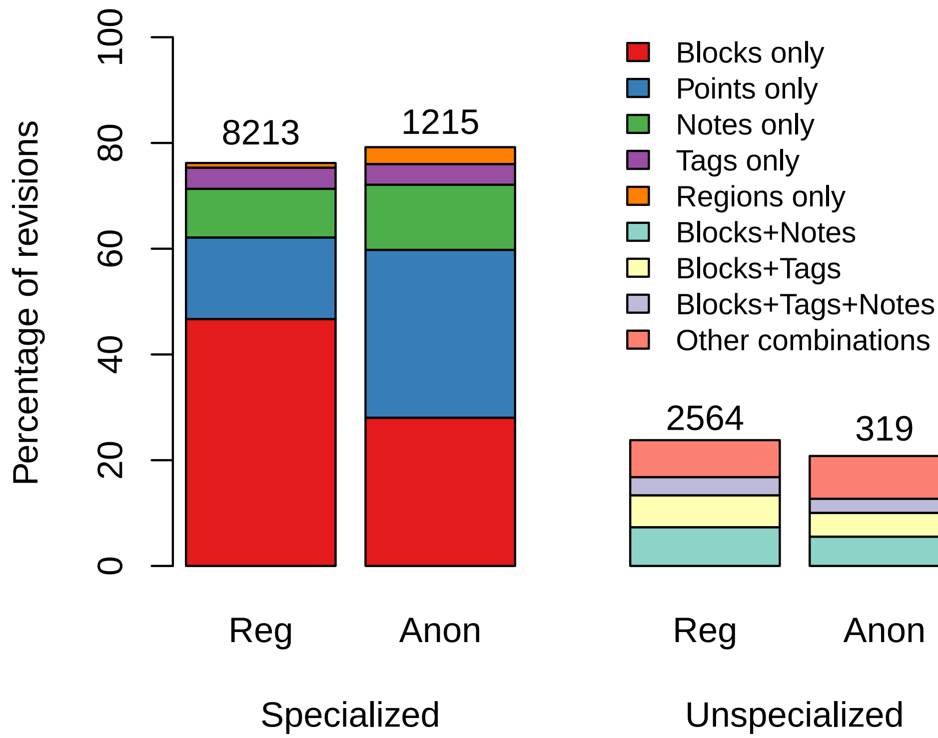

Figure 2.1: Specialization of revisions by registered (*Reg*) and anonymous (*Anon*) users. Note that bar height is normalized between the two classes of users; the actual revision counts are above the bars.

about 9 months and 14 months, respectively, after Cyclopath was released, which explains in part their lower usage. So few revisions modify regions that they are not visible in Figure 2.1, and we therefore exclude regions from further analysis. Our observations concerning the results are as follows:

1. *Most revisions consist of a single type of work.* Although no Cyclopath norm dictates this, 78% of revisions made by registered users, and 80% made by anonymous users ($p < 0.001$ in both cases[1]) consist of edits to a single type of map feature. We call such revisions *specialized* in the relevant map feature.

2. *Block edits are the most common work type.* We believe there are several

[1] In each case, we compared the proportion of specialized revisions to unspecialized ones using a 2-sample test for equality of proportions using continuity correction.

reasons for this. First, there are two orders of magnitude more blocks in Cyclopath than other types of features. Second, blocks are the crucial unit in Cyclopath, as they form the basis of routing. Without blocks – and without accurate connections among blocks – routing would be impossible. The other features add useful information but are not strictly necessary.

3. *There are clear differences in the editing behaviors of registered and anonymous users.* Specifically, block editing accounts for a much lower proportion of revisions made by anonymous users ($\chi^2 = 152.64, df = 1, p < 0.001$). We think that this is because while editing blocks is very important, it also is difficult (due to intrinsic properties of the task – block editing involves checking connectivity to neighboring blocks, shape, alignment etc.) As noted later in this chapter, users apparently need time to learn and understand block editing; indeed, revisions made by registered users early in their careers have a work type distribution quite similar to anonymous users.

4. *Certain combinations of work types are most common.* Blocks+tags and blocks+notes (in general, object+annotation) are the most popular combinations ($\chi^2 = 30.44, df = 1, p < 0.001$). Since tags and notes let users provide additional information about blocks (as well as points), it makes sense that users would add information to explain their edits to blocks.

Our results show a strong degree of specialization in the entire set of revisions, nearly 80%. A natural issue to investigate next is whether individual users specialize.

2.5 RQ2: Specialization of Contributors

2.5.1 Identifying specialists

Procedure. We began by developing exploratory visualizations of the data to help us identify major patterns and guide quantitative analysis. We chose

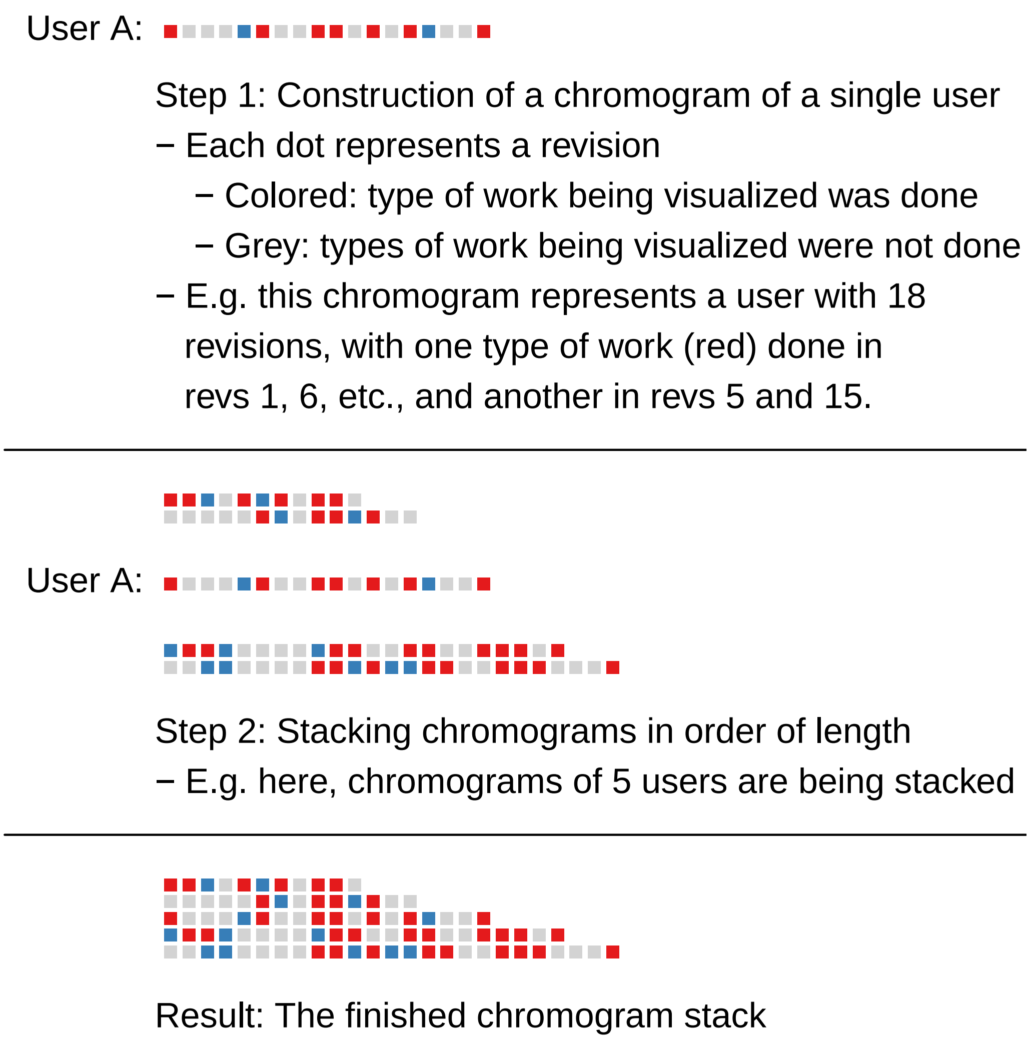

Figure 2.2: Construction of a chromogram stack.

the chromogram technique because it has been used in the past to discover patterns in Wikipedia revisions [102]. Since there is a separate chromogram for each user's revisions, it is hard to identify patterns across a set of users; one has to study the chromograms of different users separately and consolidate observations externally. We overcame this difficulty by "stacking" chromograms of multiple users one above the other; this lets us observe cross-user patterns. We call this extension a *chromogram stack*.

Figure Figure 2.2 explains how to construct a chromogram stack. We use colors to indicate work of a particular type in a revision and grey to indicate that no work type of analytic interest was present. Due to their use of colors, chromogram stacks are best read on a color screen or color printout.

Results. Figure Figure 2.3 shows a part of the chromogram stack for edits to blocks and points. Red indicates a block revision, blue a point revision, black a block+point revision, and gray neither. We limited the chromogram stack to include only users with at least 15 revisions, and we truncated the display at 80 revisions. The figure reveals several patterns, including a few rows that are mostly blue and many more rows that are predominantly red. This suggests the existence of work type specialists and that there are more block specialists than point specialists. To confirm these suggestions, we next formally define what it means for a user to be a specialist and quantify their distribution in the Cyclopath user population.

2.5.2 Counting specialists

Procedure. We define a user as *specialized in a map feature* if more than 60% of the total number of that user's revisions are specialized in that map feature. For example, a user who has made 40 revisions, of which 30 are point specialized, is a point specialist. We chose 60% because it is a common supermajority threshold; further, this definition guarantees that a user specialized in a particular work type makes revisions specialized in that work type at least 50% more often than all other work types combined. As before, we consider only registered users with at least 15 revisions in our analysis.

This metric assumes that the "value" of each revision, as a unit of work,

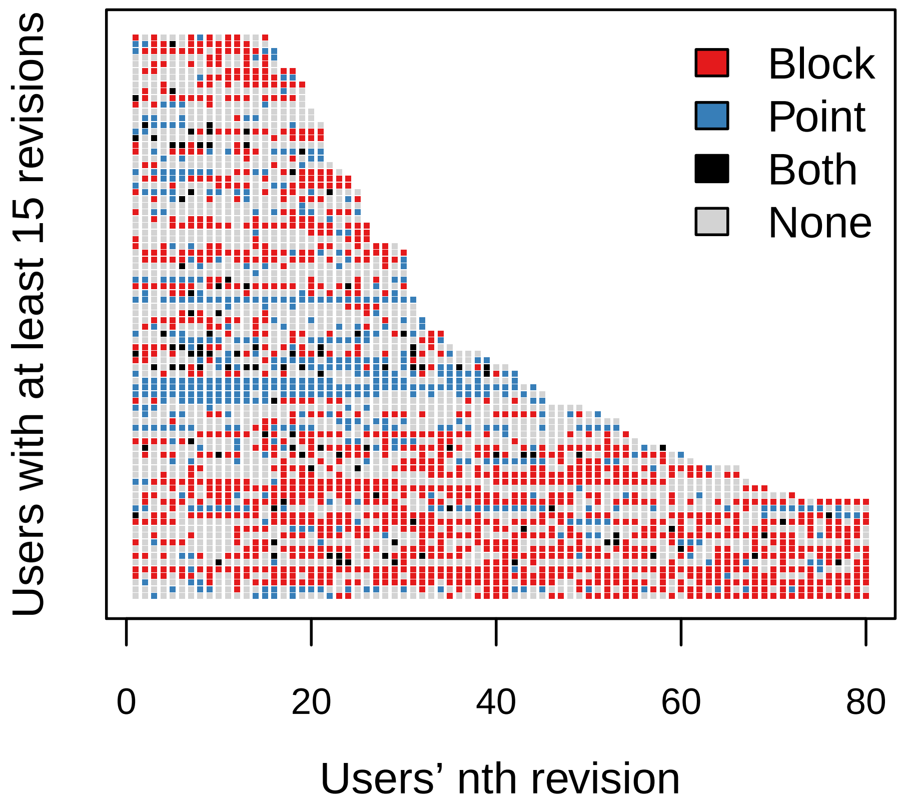

Figure 2.3: A section of the chromogram stack for edits to points and blocks: There are some blue rows, indicating specialization in points, and many red rows, indicating specialization in blocks.

is the same. Clearly, this will not be true for any given pair of revisions; for example, a revision with edits to 5 blocks does not represent the same amount of work as a revision with edits to 2 points. However, in the aggregate, such differences balance out, because the probabilistic expected value of a revision equals the mean value of all revisions.[2]

[2]We compared this metric to one using *an atomic map feature edit* (e.g. changing the geometry of one block, or creating a new point) as the unit of work. Using this alternative, a user would be defined as specialized in a map feature if more that 60% of the total number of map features the user has edited are of one type (e.g., block). We found no differences in the trends, and the overall pattern of results was the same. We chose revisions as they are

Work type	Number of specialists among	
	Study users	Experienced users
Block	38	20
Point	10	2
Note	5	0
Tag	2	0
none	29	5
Total	84	27

Table 2.1: Specialists by work type, using our super-majority specialization metric: There are many more block specialists than other types of specialists.

Results. Table 2.1 shows the results of computing our user specialization metric. We count the number of specialists by work type among all our study users (users with 15 revisions or more), as well as within experienced users only (top 5% of all contributors, as defined by Panciera et al. [80]). Using the Fisher-Exact test for equality of proportions, we found that the proportions of the different types of specialists were not all identical ($p < 0.001$). Post-hoc pairwise tests showed that the proportion of block specialists is significantly more than those of point, note and tag specialists ($p < 0.001$ in all cases). The results confirm the patterns suggested in the chromogram stack: 55 out of 84 (65%) or about two thirds of the users in our study are specialists, with the greatest number (nearly half) specializing in block edits. Further, we see that experienced users are predominantly block specialists (also using Fisher-Exact, $p = 0.005$).

2.5.3 Summary of results

Bringing together the results noted above, we make three key observations:

1. *Most users specialize in editing one type of map feature.* Various factors might lead users to specialize, including different knowledge, different perceptions of what feature types are most important to the community, and different preferences for and understanding of the various Cyclopath

more clearly parallel with other wiki work.

editing tools.

2. *There are many more block specialists than any other type of specialist.* In the previous section, we saw that more revisions are specialized in blocks than in any other type of map feature. The same reasons we offered to explain specialization at the revision level apply here: more opportunities and greater importance. In addition, users may also find performing block edits most interesting, since this is a rare and thus potentially appealing feature in map-based interfaces.

3. *Experienced users devote a higher proportion of their effort to editing blocks than do anonymous and less experienced users.* Since blocks play a central role in the key public Cyclopath service – route finding – this finding is consistent with prior research showing that experienced users are more committed to their communities [10, 78]. Further, block editing is one of the most difficult types of work in Cyclopath, and thus takes time to learn and commitment to master.

Thus far, we have quantified the nature of work type specialization among users. However, we also have uncovered hints that users may change specializations as they gain experience. We explore this question next.

2.6 RQ3: Change in Specialization

2.6.1 Specialists at different experience levels

Procedure. To investigate the distribution of user specialization at different user experience levels, we segmented revisions into buckets constituting progressively larger portions of users' revision histories (the first 15, 30, 60, 120, 240, and 480 revisions each user made). We then counted the specialists of each type of work (using the procedure in the previous section) after each interval.

Results. Figure Figure 2.4 shows our results. The clearest pattern is that the proportion of block specialists increases for more experienced users. If we

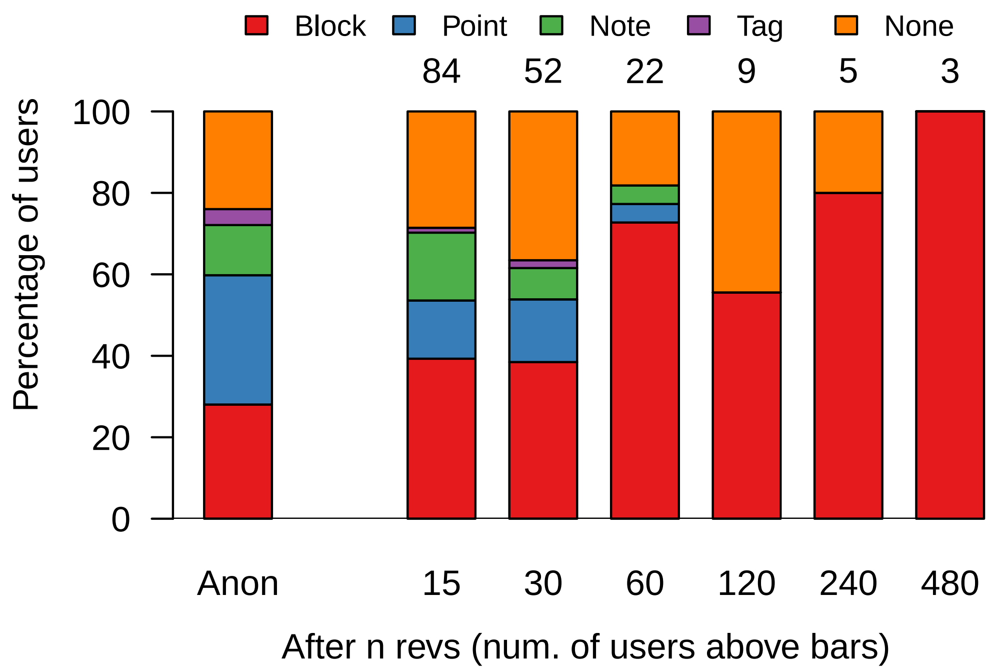

Figure 2.4: Distribution of specialization in different types of map features at various stages in users' editing careers. At low experience levels, the distribution of specialists resembles the distribution of specializations in anonymous revisions (*Anon*). At higher experience, only block specialists remain.

consider only the first 15 revisions of registered users (including those who went on to make many more revisions), the proportion of block specialists resembles the proportion of block specialized revisions made by anonymous users. This supports the conjecture that block specialization evolves over time, due (we speculate) to increased commitment to the community and increased mastery of complicated editing tools. This gives rise to the following questions: What happens to the users who do not specialize in editing blocks early in their careers? Do they drop out at a higher rate than block specialists? Or do they change their specialization to block editing as they gain experience? We address these questions next.

Initial	Now				
	Blocks	Points	Notes	Tags	*none*
Blocks	**18**	0	0	0	2
Points	0	**6**	0	0	3
Notes	4	0	**2**	0	2
Tags	0	0	0	**0**	0
none	4	2	0	1	**8**

Table 2.2: Change in specialization. *Initial* is the users' first 15 revisions, while *now* is the full revision history. 34 out of the 52 users (in bold) examined did not change specializations, whereas 18 did.

2.6.2 Change in user specialization

Procedure. To find out whether users change specializations and become block specialists, we took the subset of users who had the opportunity to change specializations – the 52 users who had made at least 30 revisions – and compared each's specialization at two points: (a) their first 15 revisions to (b) their entire revision history.

Results. The results of this analysis are tabulated in Table 2.2. First, a solid majority (65%) of users do not change specialization over their editing careers ($\chi^2 = 64.14, df = 1, p < 0.001$). However, this leaves a large minority (35%) who do change. The two most notable patterns are: eight (15%) users develop into block specialists, and seven (13%) other users turn into generalists.

2.6.3 Effect of habituation

Procedure. To measure the extent to which users became habituated in their editing patterns, we did an analysis based on one by Sen et al. concerning tagging behavior [91]. For each user, we computed the cosine similarity of his or her n^{th} revision to the previous $n - 1$ revisions, for all values of n. Cosine similarity values range from 0 (complete dissimilarity) to 1 (identity).

At the computational level, a single revision is represented as a vector with one component for each type of map feature: block, point, region, note, and tag (in this order). The value for a given component represents the proportion of features in the revision that were of that type. To clarify how to interpret

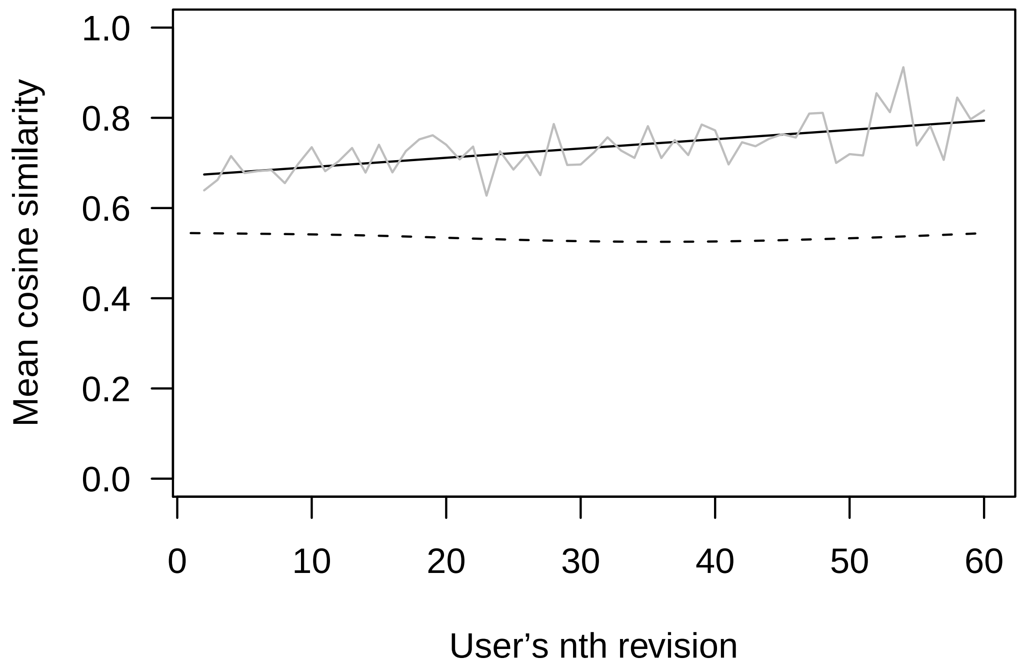

Figure 2.5: Mean cosine similarity of a user's n^{th} revision with his or her prior revisions (solid line). Similarities to a uniform distribution is also shown (dotted line). Users increasingly "become set in their own ways" as they contribute more.

differences between two cosine similarity values, we again follow Sen et al. and provide a frame of reference by computing the similarity between any revision and a vector [0.2, 0.2, 0.2, 0.2, 0.2] representing a uniform distribution of work.[3]

Results. The results of the habituation analysis are shown in Figure 2.5. Notice that there is a clear increase in similarity of a user's current revision to previous revisions over time, beginning at about 0.6 and rising to about 0.8 after 60 revisions.

Unlike Sen et al., we did not study the effect of community influence on user's editing behavior. In their case, community influence was direct (and

[3]As tagging and regions were added during the analytic time frame we cover, we used a uniform vector of [0.33, 0.33, 0.33] prior to the release of tagging and regions, and [0.25, 0.25, 0.25, 0.25] after tagging was released but before regions were added.

experimentally manipulated): namely, it consisted of what community-applied tags were displayed to a user in the tag application interface. In Cyclopath, there is no correspondingly direct exposure to the nature of revisions made by the community as a whole: the Recent Changes list and Geographic Diff mode (to view what exactly changed in one or more revisions) interfaces only allow scanning of individual revisions.

2.6.4 Summary of results

Bringing together the results noted above, we make the following key observations:

1. *Most users did not change their specialization over their life-cycles; however, those who did generally turned into block specialists or generalists.* Based on prior work [10], taking on more core and complicated tasks, like block editing, could be interpreted as an indication of higher commitment to being a Cyclopath contributor. One interesting example is that one of the most prolific Cyclopath users began as a note specialist, then transitioned into a block specialist at their 83rd revision. By the end of the data analyzed, 72% of the user's revisions were block specialized. There are several possible reasons for the transition to becoming a generalist, including users wanting to diversify and try out new types of work and being in the process of transitioning to block specialization. Qualitative analyses can help us answer these questions.

2. *Users tended to get habituated to their editing patterns.* Is this degree of habituation ("becoming set in one's ways") inevitable? Or could we intervene to try to change these patterns? Prior work suggests that interventions may nudge users to new patterns [85]. However, once the intervention was discontinued, users went back to their old patterns of work.

2.7 Implications and Future Work

2.7.1 Is specialization desirable?

Recall that users can specialize by topic (geographic location and extent) and work type. Topic specialization certainly is desirable. Just like Yahoo! Answers needs people who know about cats, cars, and quarks, Cyclopath needs users who know about Minneapolis, St. Paul, and the Minnesota River Bottoms trail.

Whether work type specialization is desirable is a more subtle issue. Do we really need about half of all users and a majority of all revisions to be specialized in blocks? Our hypothesis is that the distribution of specializations reflects the relative importance of the different types of work. Further, when we consider user life-cycles, we see that casual and new users of the system contribute a healthy number of non-critical work types—points and annotations (notes and tags)—while experienced and power users account for a large proportion of the critical task of block editing. There is a steady stream of new and casual users, for whom doing simpler tasks like editing points, notes, and tags is an easy entry to the system.

2.7.2 Modal interfaces

Most Cyclopath users view the map and request routes, but do not edit. Therefore, segregating "view mode" from "edit mode" (like Wikipedia) would almost certainly make the user interface significantly easier to use. Further, because most Cyclopath revisions consist of edits to a single type of map feature, redesigning the interface tools to support single-feature editing also has potential to ease users' editing work. This doesn't require strict modal separation; instead, modifications that make it easier to make multiple edits of the same type (for example) is one promising idea. This has possible extensions in other social production communities as well; for example, Wikipedia could have a separate "wikifying" mode.

2.7.3 Composition of coherent work units

Prior work on Cyclopath showed that presenting work opportunities to users in a visually comprehensible format elicited a significant increase in participation [85]. Our current findings can help designers of social production communities construct these work opportunities more intelligently: offer tasks that involve common, meaningful combinations of work types, like editing objects+annotations, e.g., in Cyclopath, adding points only, or editing some blocks and then adding notes about them.

2.7.4 Intelligent task routing

Intelligent task routing is the process of automatically recommending tasks to users who are likely to have the interest and ability to perform them [21]. It has been tested and shown effective in Cyclopath [85], MovieLens [21], and Wikipedia [22]. Our findings here suggest that work type also would be effective for task routing. For example, in Cyclopath, we could recommend tasks involving fixing connectivity at intersections to block specialists, and work in Wikipedia [22] can be refined to recommend tasks involving wikifying, and ensuring neutral point of view to corresponding specialists. If the community needs specialists for a new type of task (e.g. for monitoring map edits), we could also devise methods to cultivate new specializations (intelligent recruiting).

Development campaigns are a particularly interesting application of intelligent task routing. Wikipedia has “WikiProjects”, domain-specific collaborative efforts to organize volunteer work. Contributors to WikiProjects specialize, e.g., some people add new content, whereas some others fix links and typos. In geographic crowd-sourced communities, there could be map-wide development campaigns organized by map feature work, like fixing intersections and bridges or adding all sports-related points of interest. Of course, a campaign might recruit on both dimensions: “the Uptown neighborhood needs a tag specialist!”

2.7.5 Specialization by topic

Along with work type, topic is an important dimension of tasks. In a geographic context, we can operationalize topic in two ways:

- **Geographic shape.** There are two obvious types of *geographic shapes* that could delimit a cyclist's knowledge: *area* – editing is focused on areas (e.g., the neighborhood surrounding one's home) – and *route* – editing is focused in a "linear" way, e.g., along portions of a work commute or favorite recreational trails.

  Do users specialize by geographic shape? Preliminary studies indicate that there are users whose edits are mostly area-shaped and others whose edits are mostly route-shaped (see Figure 2.6 for an example of a route-shaped collection of edits). However, more exploration and analyses (using spatial statistical tools like the SANET [75]) are necessary to confirm and extend these observations.

- **Geographic extent.** Research on Wikipedia has shown that people diversify their edits as they become more experienced [10]. This leads to the question whether the *geographic extent* of users' contributions grew as they gained experience: did they edit across a broader portion of the map as they contributed more? Preliminary analyses using Ripley's K [87]—a metric from the field of spatial statistics— provide supporting evidence.

  Ripley's K measures the degree of radial spatial compactness of a set of points based on inter-point distances: the more spatially compact the set of points, the higher is the K value. For example, Figure 2.7 shows Ripley's K for two quite different set of edits. The use of Ripley's K requires us to define two parameters:

  1. *Study area* defines the baseline for measuring how clustered or dispersed a set of points is. For our analysis, we have 72 "sets of points", consisting of the geographic centroids of the revisions for each user having more than 15 revisions. We chose a study area

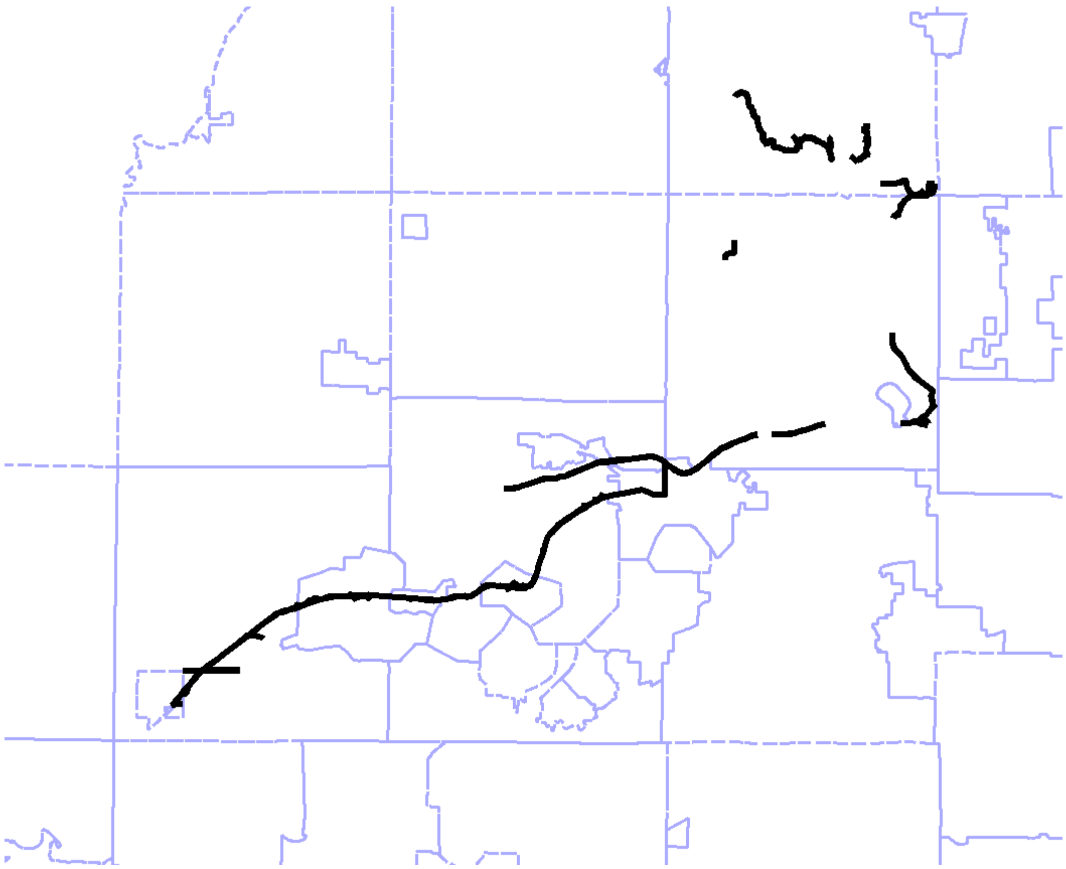

Figure 2.6: Example of topic specialization. Bold black indicates locations of contributions and the light blue lines are city boundaries. This user has made most of their edits along a route-shaped pattern.

equal to the extent of the entire Cyclopath map. This is analogous to using the entire space of Wikipedia articles to study trends in specialization in specific topics. A different choice of study area would have simply scaled all the numbers uniformly, leaving the trends among them unchanged.

2. *Clustering distance* defines the maximum distance between any two points that are considered to be clustered. Essentially, it is a way of indicating our understanding of compactness to the Ripley's K algorithm. We tried several values between 50 and 5,000 meters, with no change in the nature of our results, so we settled on 50

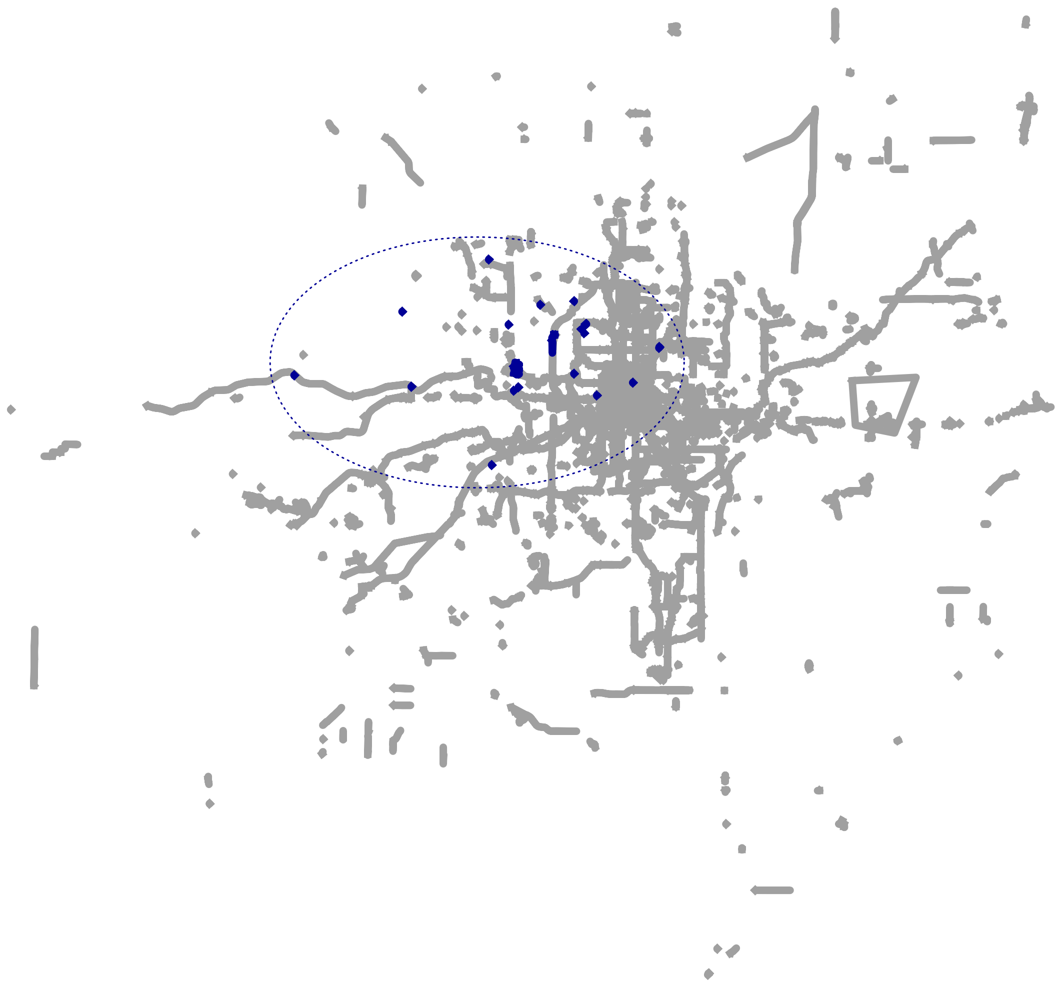

Figure 2.7: Objects edited by two users. The edit history shown in gray has a Ripley's K of 3050 and covers nearly the entire metro area, while the edit history shown in blue and circled has K of 13800.

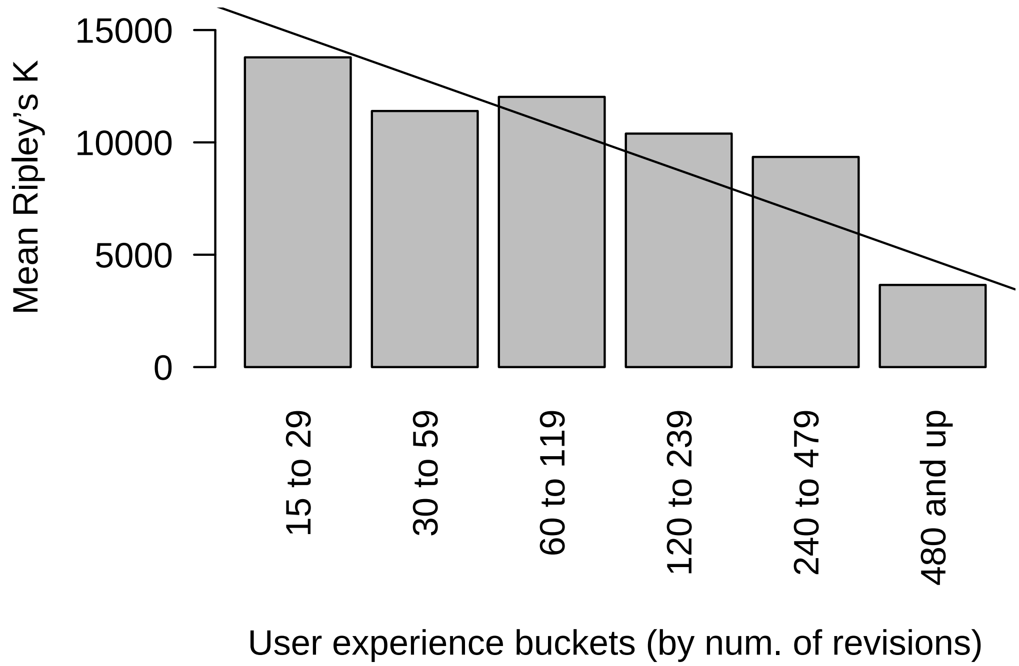

Figure 2.8: Clustering at different experience levels. At higher experience levels, spatial dispersion of revisions increases (slope $= -1669$, $R^2 = 0.75$). Note that the buckets are exponential.

meters.

Figure 2.8 summarizes the results of our analysis and lets us make several interesting observations. First, as we look at users who contribute more, we see that their contributions are less spatially compact. In other words, "power editors" work has broader geographical extent (i.e., power editors work in a larger variety of places). Further, in some cases this extent encompasses nearly the entire 8,000-square-kilometer metropolitan area. This can be confirmed by comparing the rightmost bar in Figure 2.8 with the edit history shown in gray in Figure 2.7; both have a Ripley's K value of about 3,000. We believe these results suggest that users start by editing in areas they are familiar with, which typically are smaller. As they gain experience, they diversify into other areas that are less familiar to them. Consistent with Bryant et al.'s study of Wikipedia editors [10], we see this as a transition from *my neighborhood* to *my community.*

2.8 Summary

We studied specialization of users and their contributions in a geographic wiki, Cyclopath, exploring the well-known dimension of work type. We found clear specialization by work type, with block specialists most numerous. We also saw user life-cycle effects: Experienced users either took on more difficult tasks (becoming block specialists) or more diverse tasks (becoming generalists). These shifts suggest an evolution from *my neighborhood* to *my community.*

By creating modular interfaces, composing smarter work units, and routing them to people who are more likely to fulfill them, we may be able reduce under- contribution in online, social production communities to an extent. However, techniques like intelligent task routing are, at their core, dependent on users complying with requests to contribute. Most participants in an online social production community consume, and very few contribute. Thus, having users comply with requests for contribution may involve persuading participants to carry out activities they would not ordinarily do. In the next chapter, we evaluate a few social psychological techniques that may make this possible. These techniques have the potential to enhance the effects of intelligent task routing by eliciting contribution above and beyond what is possible by merely routing tasks to the appropriate people.

Chapter 3

Increasing Contributions through Compliance with Requests

3.1 Introduction

In the previous chapter, we furthered our understanding of how participants in a social production system contribute to its shared, central resource. This knowledge has the potential to help us design systems and techniques that encourage contribution and growth. Several of these techniques, such as those based on intelligent task routing and goal-setting, assign contributory tasks to participants, and are only effective to the extent that these requests to contribute are complied with.

Fortunately, a large body of literature from social psychology suggests techniques that may be used to motivate people to contribute to social production communities. One set of these techniques is known as *compliance without pressure* [32]: these techniques are designed to increase compliance to requests in the absence of any obvious sources of pressure. Two of the most popular techniques are:

- *Foot-in-the-door* (FITD): Once a person performs a small request, he/she is more likely to perform a subsequent, larger demand [32].

- *Low-ball* (LB): Once a person commits to a request, he/she is more likely to perform it at a later stage, even if its cost is increased [18].

While these techniques have been evaluated extensively in several offline contexts, there has been very little evaluation in the world of computer- mediated communication [38, 39], and none in online social production communities.

We evaluated FITD and LB techniques in the context of Cyclopath. We carried out a work campaign in which we requested Cyclopath users to contribute information to the map in the form of tags with the goal of improving the map for the next cycling season. We conducted a field study and a follow-up survey and found that LB succeeded in eliciting more contributions than the control condition, while FITD had limited success.

This chapter is organized as follows. We first survey relevant related work and explain the motivation to carry out empirical evaluation in the online context. Second, we describe our field experiment, its results and their interpretation using theories from social psychology. Third, we describe our follow-up survey, and conclude with implications for design and future research.

3.2 Related Work

Compliance without pressure techniques are designed to lead people to comply with requests without any obvious source of external pressure. Researchers from social psychology and marketing have developed several techniques that fall under this category, e.g. FITD, LB, *door-in-the-face* [17], *that's-not-all* [11] and *bait-and-switch*. FITD is one of the oldest and the most popular of them with well over a hundred studies devoted to it over the last 50 years [12]. LB has also been found to be one of the most effective of these techniques [49, 50]. Because of their effectiveness, FITD and LB are interesting candidate solutions to the problem of under-contribution in online communities, and we thus chose to investigate them.

Several theories have tried to explain why and how FITD and LB work (or not). The most popular explanation for the success of the FITD technique is Bem's self-perception theory [6], which postulates that when a person is induced to comply with a smaller request, he/she is more likely to comply with a subsequent, larger request because of perceiving himself/herself as the type of person that does such tasks. Alternative theories including those citing psychological reactance, conformity to existing social norms including reciprocity and cognitive dissonance [30] have also been used to explain both successes and failures of the FITD technique. The LB effect has been generally explained using the theory of commitment [18], which says that people generally tend to stick to their commitments when acting in public view. We elaborate on these techniques in a later section in this paper while interpreting the results of our field experiment.

While compliance-without-pressure techniques may have potential to increase participation in online communities, if users perceive these techniques as manipulative, their use could negatively impact long-term member satisfaction and commitment. Accordingly, in this work, along with evaluating how successful FITD and LB are in an online social production context, we also evaluate the extent to which they are harmful and discuss the pragmatic issues designers of online communities might face when employing some of them.

3.3 Need for Empirical Evaluation

This research focuses on empirically evaluating the techniques of FITD and LB in the context of an online social production community. Is this necessary? Could we not merely assume findings from the offline settings of social psychology? Online communities have several key differences that makes them unique, and justify the need for an empirical evaluation of proven offline techniques:

1. **Anonymity on the Internet.** Psychological phenomena like cognitive dissonance, considered one of the factors behind the FITD and LB effects, as well as the LB effect itself have been shown to be more effective in public situations [13, 95], where visibility is higher (and consequently,

anonymity is lower). However, since obtaining an identity on the Internet is cheap [33], online communications have a higher degree of anonymity than offline situations. Hence, it is not clear if online social production communities are public (contributions are shared) or private spaces (people can work from the confines of their homes), thereby raising questions over the effectiveness of the FITD and LB techniques.

2. **Nature of online communication.** Email and similar modes of communication are used most often by online social production communities to interact with their members, as it is the norm on the Internet. These modes are fundamentally not as personal as face-to-face or telephone communication – modes used by most studies of compliance in offline situations. Further, due to problems of email overload, it easy to ignore incoming messages. These reasons affect any interventions that use email as the underlying mode of communication.

Although online communities are different from their offline counterparts in some respects, they are similar in many others. Indeed, the social sciences have provided design inspiration and specific techniques that Computer Supported Cooperative Work (CSCW) designers and researchers have used to create more effective online communities [81]. Researchers have successfully applied theories and models like the Collective Effort Model [51, 62], goal-setting [62] and social comparison [47] to address the problem of under-contribution. Closer to this work, the FITD technique has been shown to be effective over email [38, 39].

Drawing inspiration from these successes, our research extends prior work in significant ways:

- We empirically evaluate the FITD and LB techniques in the context of an online social production community using a live field experiment and a follow-up survey.
- We demonstrate that the LB technique is largely successful in eliciting a higher quantity of contributions, while the FITD technique receives mixed results.

- Based on what we learned from our studies, we discuss the advantages and limitations of employing these (and similar) techniques in real-world online social production communities.

3.4 Experiment Design

3.4.1 General Structure

We conducted a field study in Cyclopath during winter 2010-11. Like bike-riding (at least, in colder climates), Cyclopath use is seasonal—it sees over 150 route requests per day and over 100 revisions per week during the summer and only about a third of those numbers during the winter. This made winter a good time for our field study—any effects we observe would primarily be due to our manipulation and not the general motivations that drive contributions during summer.

Prior work in Cyclopath suggests that notes on blocks and points often contain potential tags [96]. If extracted, these tags could provide more options for users to customize their routes. Further, this task could be performed by any person, regardless of his/her familiarity with the note or the area on the map where the note was applied (just like the task of fixing intersections [85]). Hence we chose to ask Cyclopath's users to do the task of extracting tags from notes. Prior to the start of our experiment, users had added 6,373 tags (applied to 4,312 blocks and points) and 8,405 notes (applied to 6,084 blocks).

3.4.2 Hypotheses and Variables

Our experiment design was guided by two hypotheses:

H1. The FITD technique will result in higher compliance for the contribution request than the request being presented alone.

H2. The LB technique will result in higher compliance for the contribution request than the request being presented alone.

We measured compliance using two outcome (dependent) variables:

- **Response:** Whether the user responded at all and the number of responses per user (we allowed users to respond multiple times, as described later) to the target request.
- **Quantity of Work:** The amount of work done by the user in response to our request, measured in terms of number of tag-applications to blocks and points, and number of new tags introduced into the tag vocabulary.
- **Quality of Work:** The usefulness of work done by the user in response to our request, measured in terms of user ratings of usefulness of tags.

3.4.3 Method

3.4.3.1 Partitioning Users

We partitioned users into three groups—FITD (953 subjects), LB (953 subjects), and a control group (951 subjects). However, since participation on Cyclopath is highly unequal (like most voluntary online activities), we took care to ensure that user experience levels were more-or-less evenly distributed among the three groups, i.e. not all highly experienced users (measured by the number of prior revisions made) were assigned to the same group. Specifically, we ordered users in descending order by the number of revisions they saved and stepped through this list assigning users to control, FITD and LB groups in a cyclic fashion.

3.4.3.2 Procedure

All communication with users was conducted via email. We presented both experimental groups (FITD and LB) with their *initial contacts* on December 9, 2010 at 5 pm. We then followed this up with the *target request* (the real request for which we wanted compliance) to all groups, via email 6 days later[1], on December 15, 2010 at 5 pm (see Figure 3.1).

[1] Prior research suggests that a short (near-zero) delay between the initial and the target requests when both originate from the same requester can tend to counteract the FITD effect [12, 46].

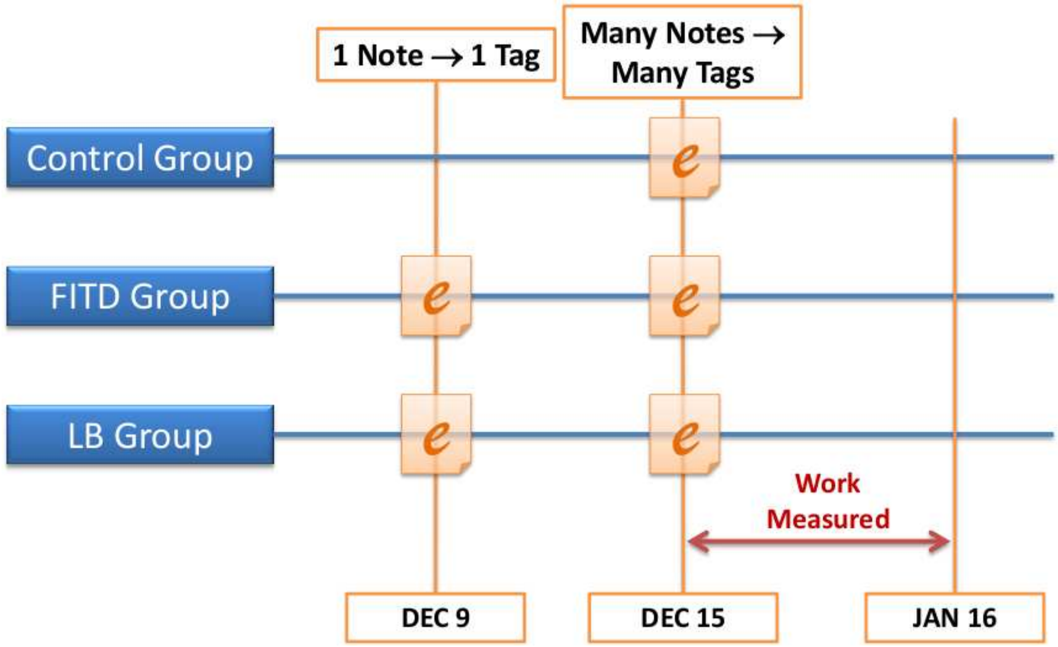

Figure 3.1: Our experimental procedure. The horizontal blue lines represent timelines for each of the subject groups. The initial contact and the target requests were made via e-mail on Dec 9 and Dec 15, respectively. We took a snapshot of the data on Jan 16.

It is important to note here that essentially, we are comparing 3 techniques of eliciting work: "simply asking users to do work" (control) and two "enhanced ways of asking users to do work" (FITD, LB). We could have evaluated other, intermediate techniques like Freedman and Fraser's *familiarity-only* (related to FITD) [32] and *commitment-but-no-increase-in-cost* (related to LB) as well. However, in this chapter, we have explicitly chosen to evaluate FITD and LB in their entirety and not part-by-part.

The following paragraphs detail the initial contact and the target request for the various groups (the control group received no initial contact).

Initial contact – FITD. We sent email to the subjects in the FITD group asking them to do the initial (small) task of extracting just one tag from a single note, with the subject line "A tiny favor: Cyclopath needs your help!". The key passage was: *Would you spare a minute to help this campaign? All we are asking you to do is to read the following note and extract only one* *meaningful tag (one or two word label) from it.* The email also contained the note itself (we thought that once the subject eyeballed it, the task of extracting one tag would seem trivial), a sample note-tag pair, and a link to an interface to submit the tag. This link launched a simple HTML-based interface where the subject could re-read the note and enter the tag he/she extracted (see Figure 3.2(a)). For example, a subject extracted the tags *"Dirt path"* from the note *"dirt path connects to Arcade beware of curb"*. Since the initial FITD contact had to be short and easy, we limited subjects to just one attempt at it: if a subject clicked the link again, the system would display a message saying he/she has already completed the task. After the subject submitted the tag from the HTML interface, he/she was thanked.

Initial contact – LB. We asked the subjects in the LB group (with the same subject line) to agree to extract one tag from a single note, explicitly asserting that this task should take only a minute to complete. The key passages were: *Would you spare a minute to help this campaign? All we are asking you to do is to read a note and extract only one tag (single word label) from it (this should just take you a minute)* and *If you agree now, we'll contact you with specific instructions in near future, after we have heard from more people.* The email also contained a sample note-tag pair and a link to an interface to express commitment. This link launched a simple HTML-based interface (similar to the FITD interface) where the subject was thanked for agreeing (see Figure 3.2(b)).

Target request – all groups. Our target request to all three groups was to extract as many tags as possible from a given set of notes. To avoid any biases resulting from subjects' self-selection, this email request was sent to every subject in the FITD and LB groups, regardless of their response to the initial contact. Crucially, for the FITD and LB conditions, we needed to explain why we were asking for more work than in the original request. Our explanation was that Cyclopath had miscalculated the amount of work required of each user and that more work per user was needed in order to analyze all notes by spring time[2]. The key passage in this email was: *You*

[2]This argument was also included in the email sent to the control group for uniformity,

(a) FITD Interface

(b) LB Interface

Figure 3.2: Initial contact. We designed simple HTML-based interfaces for the FITD subjects to perform their small task, and for the LB subjects to express their commitment.

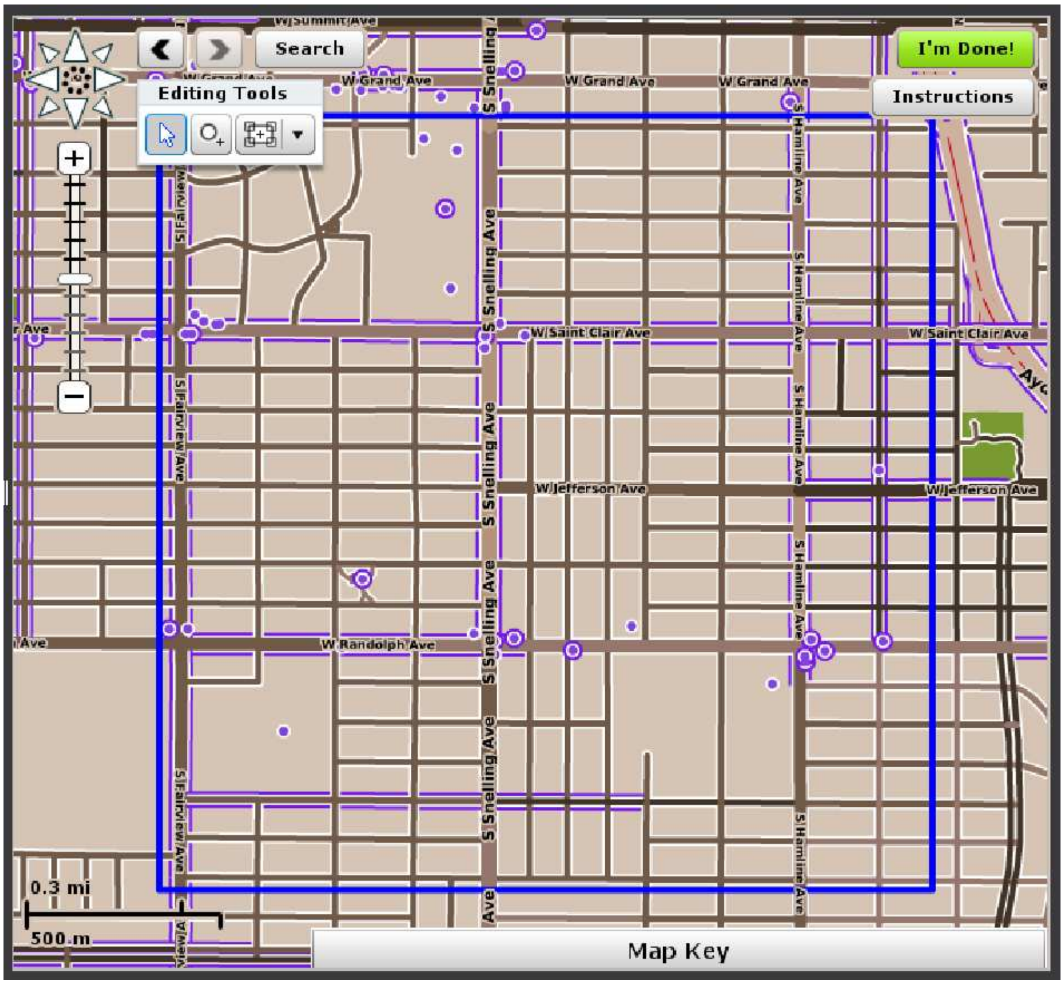

Figure 3.3: Target request. Each subject could perform as many experimental trials as he/she wishes. The blue box showing the area the subject was asked to work in, with the "I'm Done" button to end the trial at the top right.

will be presented with a area on the map. Read any notes you find on blocks in this area and extract around 2-3 potentially useful tags (one or two word labels) from each note. This should take you 15-20 minutes. The link in the email launched an experimental trial (the procedure as well as the code for conducting trials was adapted from [85]):

1. *Begin trial.* The subject begins the trial by clicking on the link provided in the email.

2. *Select viewport.* The system chooses a viewport (100 km^2 area) randomly on the map, ensuring that there are at least 5 notes available to extract tags from.[3]

3. *Display viewport.* The map pans and zooms to the selected viewport, and highlights it with a blue box. An instruction dialog box pops up that details the task to be done, with brief instructions on how to do it and links to help (see Figure 3.3). All blocks and points with notes were highlighted with a purple halo.

4. *Subject does work.* The subject is now free to use the system. Although we explicitly asked the subject to work in the assigned area, the system did not restrict him/her in any way – the subject was free to make edits anywhere on the map.

5. *End trial.* The subject clicks a button labeled "I'm Done", which results in the completion of one trial. The subject can now either leave Cyclopath, use it for other purposes, or start another trial.

We recorded when the subjects started and finished the trials, the number of trials requested, and everything they did as a part of those trials.

though not framed as an "excuse".

[3]This was to ensure that there was *some* work available to do. Further, at the outset of our analyses, we also verified that there were no statistically significant differences in the number of notes available per user and per trial among the three groups.

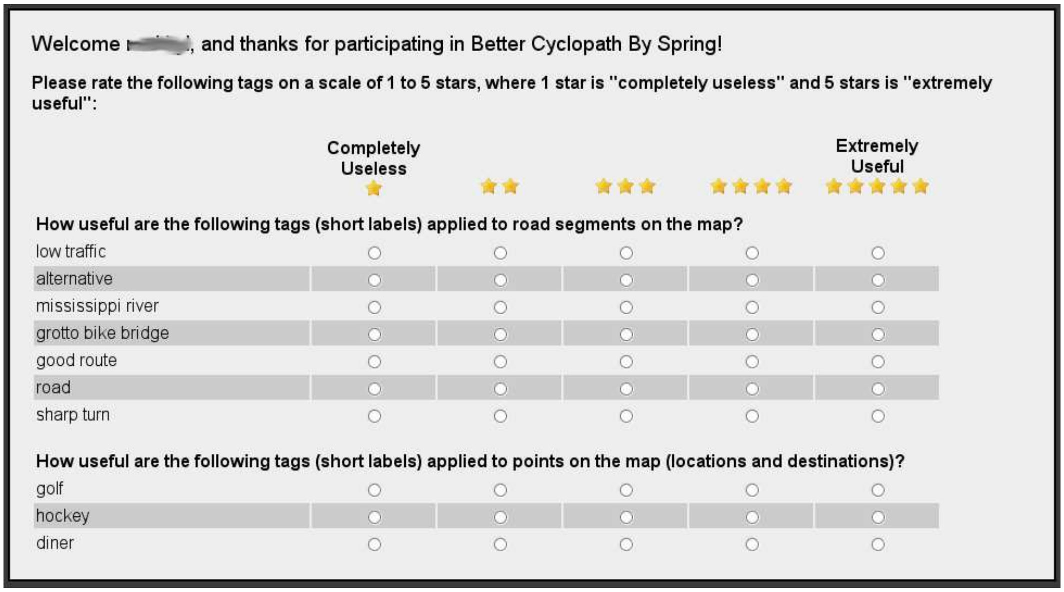

Figure 3.4: The interface used to gather user ratings on tag quality. Subjects were asked to rate 10 tags on a scale from 1 to 5 stars based on perceived usefulness.

Measurement of tag quality To measure tag quality, we designed a rating interface as shown in Figure 3.4. On April 13, 2011, we sent an email to every registered user with a link to this interface.[4] Each person was asked to rate 10 tags based on perceived usefulness on a scale of 1-star to 5-stars (1-star being "completely useless" and 5-stars being "extremely useful"). These tags were randomly chosen from among the pool of tags added by subjects during their experimental trials. To assist raters in their judgments, tags applied to blocks and points were presented separately. However, we did not provide any information about the location the lags were applied. We did this so that subjects familiar with a location would not have an advantage.

3.4.4 Results

We took a snapshot of the database on January 16, 2011. By this date, 190 users completed 597 trials. We removed 28 trials from our analyses – 15 were

[4] Since the timing of this measurement coincided with our follow-up survey (described later), we bundled this interface together with the survey.

Group	Response
FITD	11.4% (109)
LB	12.7% (121)

Table 3.1: Response to the initial contact. Response rates are low, but typical of contributory behavior in social production communities.

Group	Response	Trials per subject
Control	6.14% (58)	0.22 (205)
FITD	5.14% (49)	0.10 (92)
LB	7.35% (70)	0.29 (272)

Table 3.2: Response to the target request. The numbers in parentheses are the respective absolute counts. There are no statistically significant differences between the three groups of subjects.

unusually long (over 2 hours[5]) and 13 were done by 6 users who were not in any of our groups (these users joined Cyclopath after the experiment started). This left us with 569 trials by 177 subjects. We present our results organized by our outcome variables.

3.4.4.1 Response

Initial contact response. Table 3.1 shows the response to the initial contact. Out of the 953 subjects in the FITD condition to whom we sent the initial contact emails, 109 (11.4%) completed the short task. Out of the 953 subjects in the LB condition to whom we sent the initial contact emails, 121 (12.7%) responded.

Target request response. Table 3.2 shows how subjects responded to the target request. We see no significant differences among groups in the response rates, or the number of trials per subject. However, we found interesting differences in the quantity of work done, as described in the following subsection.

[5]The median length of a trial was 1.5 minute. We hypothesize that the unusually long trials represent cases where the subject forgot to click the "I'm done" button, remembering only much later. Consequently, we could not identify, with reasonable confidence, the parts of the trial that were done as a response to the target request.

(a) Tags created per user.

Group	Total	Mean	SD	p
Control	70	0.07	0.65	
FITD	99	0.10	1.05	
LB	193	0.20	1.64	$*$

(b) Tags applied per user.

Group	Total	Mean	SD	p
Control	542	0.57	5.07	
FITD	571	0.60	6.05	
LB	1702	1.79	18.07	$*$

Table 3.3: Tagging work done per user in response to the target request. $*$ indicates statistical significance at the 0.1 level when compared with control.

3.4.4.2 Quantity of Work

Metrics. When a tag that does not already exist in the system is applied to a new map feature, a new tag is introduced into the tagging vocabulary. We call such an action "creating a tag." On the other hand, if an existing tag is applied, it is merely re-used, i.e., a new connection is added between the existing tag and the map feature to which it is applied. We call such an action "applying a tag." Naturally, when a tag is created, it is also applied to some map feature (e.g., a block), resulting in a tag application. We counted the number of tag applications (split into tag applications on blocks and points) and the number of new tags added, per user and per trial.

Work done per user. LB subjects applied more tags on average (about 3 times) than FITD or control subjects. These results were statistically significant. We performed statistical tests (for p-value computations) on $log(1+x)$-transformed variables to account for the non-normality (long-tail nature) of the data and compensate for counts of zero. We first did a one-way analysis of variance (ANOVA) on the number of tag applications ($F(2, 2854) = 2.78, p = .06$), then did post-hoc pairwise comparisons using Tukey's HSD test (the difference between LB and control groups was marginally significant, $p = .06$). On an average, LB subjects also created nearly thrice as many tags as control subjects (ANOVA: $F(2, 2854) = 2.70, p = .07$, then Tukey's HSD: LB vs. control:

(a) Tags created per trial.

Group	Total	Mean	SD
Control	70	0.34	1.08
FITD	99	1.08	2.43
LB	193	0.71	2.54

(b) Tags applied per trial.

Group	Total	Mean	SD
Control	542	2.66	7.20
FITD	571	6.23	13.11
LB	1702	6.25	13.79

Table 3.4: Tagging work done per trial in response to the target request.

$p = .08$). Table 3.3 summarizes the number of tag applications and new tags added per user by subjects from the three groups.

Work done per trial. We see evidence of the success of the LB technique at the per-trial level too. The average LB trial produced more than twice as many tags (6.25) as the average control trial (2.66). Table 3.4 summarizes these comparisons. Looking back at Table 3.2, we see that the average FITD subject did fewer trials than the average control subject. This explains why there is a significant difference between the FITD and control groups in terms of work done per trial, despite there being none in terms of work done per user.

3.4.4.3 Quality of Work

To compare the quality of tag applications added by each group during the field study, we computed the average tag rating weighted by the number of times the tag was applied. For greater reliability, we only used tags that had been rated by at least two distinct raters. For example, suppose that only two tags with at least two ratings were added by the control group – *bumpy* (mean rating = 4) was applied 10 times and *scenic* (mean rating = 3) was applied 5 times – the weighted mean rating for the entire group would be computed as $\frac{4\times10+3\times5}{10+5} = 3.67$.

Figure 3.5 shows the results of our comparison. We found that there were

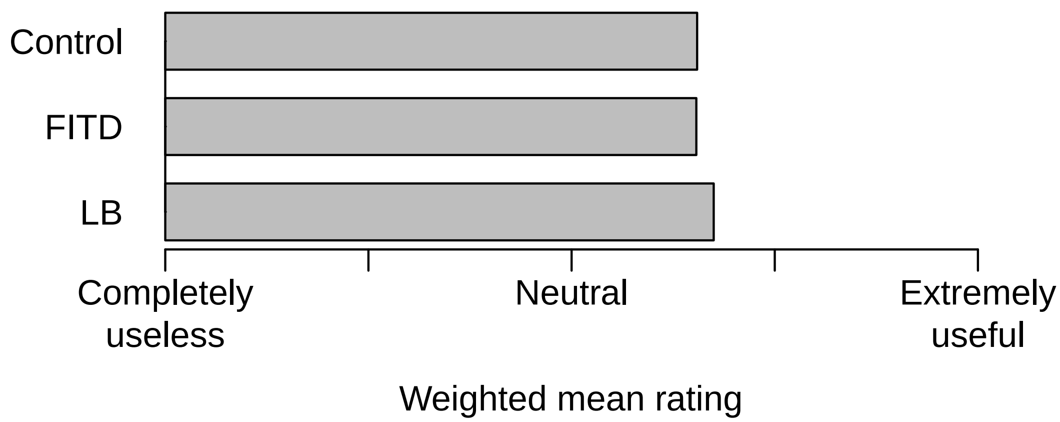

Figure 3.5: Quality of tagging. There were no statistically significant differences between the three groups.

no statistically significant differences in the tag quality among the three experimental groups. This means that the FITD and the LB groups produced contributions that were at least as good as those of the control group. While a positive change would have been more desirable, the fact that despite being subjected to a social psychological manipulation, users' contribution quality did not drop below that of the control group is an encouraging sign. That said, for each group, tag quality was higher than the "neutral" (midpoint) rating, indicating that the techniques we evaluated produced useful contribution.

Summary of results. Our results suggest that the LB technique significantly increased the quantity of work subjects put in (hypothesis H2). The FITD technique, on the other hand, is not as successful – an average FITD trial produced more work than a control trial, but on an average, FITD subjects did no more work than control subjects ($0.57 \approx 0.60$) (hypothesis H1). There also was evidence that the LB technique fared better than the FITD technique.

3.4.5 Discussion – Why these results?

We have just described an empirical investigation of the FITD and LB techniques of compliance-without-pressure in the context of an online social production community, Cyclopath. Why did we obtain the results we saw? What aspects of our procedure caused these results, and what aspects produced opposing forces? Answering these questions is important for crafting recommendations concerning the use of these techniques for designers of online communities. We now refer back to relevant social psychological theories to make sense of our results.

3.4.5.1 Why did LB succeed?

The LB effect is theoretically simpler: it is generally explained by the theory of commitment. The seminal work on the LB effect [18] attributed its success to the development of commitment to the task in the minds of the person as a consequence of the initial contact. Further research [14] has shown that commitment to the requester, not the task, is actually responsible for the increased likelihood of compliance to the target request. In our case, either is possible: task commitment, since we explicitly asked LB subjects to agree to the task of extracting a tag from a note, or requester commitment, as Cyclopath is a community resource, and people tend to care for the communities they are a part of.

3.4.5.2 Why did FITD meet with limited success?

There have been several attempts at explaining the FITD effect (or the lack of it) using a variety of psychological processes and theories [12]. Based on these theories, we speculate on the following reasons behind our results.[6]

Self-perception not strong enough. The self-perception theory [6] is the most popular explanation for the success of the FITD effect. The reasoning is that, when a person is induced to comply with the smaller request, he/she infers from this behavior that he/she is the kind of person that does such tasks and supports the cause behind the request; this self-perception later

[6]Burger's review [14] of about 100 FITD studies clearly points out that the degree to which each of the psychological processes known to enhance/reduce the FITD effect operate depends on the specific procedures used in the intervention. Hence, one must exercise caution while generalizing any speculations we make regarding how strongly they operated in our case.

causes him/her to comply with the larger request. However, if a weak FITD effect is observed (as our results indicate), then we can hypothesize that the inferential process of self-perception triggered by the smaller request did not take place or was not strong enough.

Reactance as a result of excessive requesting. Psychological reactance may also undermine the FITD effect – the subject may simply feel badgered and pressured by the repeated requests and refuse the second, larger request. Our results showed a significant difference in the work done per trial, but no difference in the work done per user between the FITD and control groups. This is because although the number of subjects from the FITD group who responded to the target request is about the same as that from the control group, the FITD group completed fewer trials than the control group (see Table 3.2). A possible hypothesis for this behavior is that some FITD subjects might have either felt badgered by the repeated requesting or felt that they have "already done their part" when they complied with the smaller request.

Reciprocity: Cyclopath did not return the favors asked. Expectation of reciprocity may actually work against the processes driving the FITD effect. If the subject, after complying with the smaller request, expects the requester to give something in return, then the fact that he/she instead comes back with a larger request has the potential of inviting a refusal on the part of the subject. On the other hand, had the subject turned down the initial contact, then the norm of reciprocity would drive him/her to comply with the target request even if it is larger. Indeed, 21 subjects who did not respond to our initial request did respond to the target request. Whether reciprocity was the reason for this behavior is worthy of investigation.

Conformity to norms: Was the target request not to be done? Another factor that might affect compliance is the knowledge of existing social norms. For example, if the subject learned that the norm is that the larger request is not complied with, then he/she also might refuse. However, in this study, Cyclopath saw nearly 500 revisions to the map after the target request was sent out: so, if any norm was present, it was the norm of contributing.

Consistency needs. The cognitive dissonance theory [30] stipulates that people have a need to appear consistent and resolve any dissonances in their selves as soon as possible, even if it means making irrational decisions. In the FITD context, when a subject is presented with the larger request, he/she behaves consistently with his/her prior behavior in a similar context, i.e., the compliance to the initial smaller request. Although social psychologists have developed a scale to measure the preference for consistency [37], any conclusion we draw here would be about the people who use Cyclopath, and not the design of the system itself.

We conducted a follow-up study to investigate which of these processes may have affected our results.

3.5 Follow-up Study

3.5.1 Research Questions

We designed a follow-up study to address the following research questions that we constructed from analyzing the results of our field study:

3.5.1.1 RQ-FITD: Why did the FITD technique meet with limited success?

Possible reasons include:

- FITD subjects did not perceive themselves to be the type of people who respond to requests from community sites like Cyclopath (self-perception).
- FITD subjects did not feel motivated enough to comply with the target request because they felt that they had already done their part when they responded to the initial contact (reactance).
- FITD subjects felt that the repeating requesting behavior from Cyclopath was badgering them (reactance).
- FITD subjects who turned down the initial contact felt obligated to comply with the target request on the grounds of reciprocity (reciprocity).

3.5.1.2 RQ-LB: Why did the LB technique succeed in eliciting more contributions?

Possible reasons include:

- LB subjects complied with the target request because they felt a sense of commitment towards the task of extracting tags from notes (task commitment).
- LB subjects complied with the target request because they felt a sense of commitment towards Cyclopath (requester commitment).

3.5.2 Method

We sent out a survey on April 13, 2011 to subjects from all the groups. In the survey, subjects were asked to imagine scenarios similar to the experimental procedure of our field study; e.g., the FITD group was presented with the scenario: *Imagine this: You receive an email from Cyclopath asking you to do a task. The task is an easy one: you're asked to read a 20-word note, and create a single meaningful tag for it (a 1-2 minute task). What would your response be?* We took this approach (of asking people to imagine) instead of relying on subjects' memories of their experimental experience for two reasons: First, the interval between the start of the experiment and the survey was about 4 months, and it was unlikely that subjects would accurately remember their experiences over such a large interval. Second, this approach made it possible for us to administer the survey even to those who had not participated in the experiment.

Note that for subjects who participated in the field study, we assigned them a survey consistent with their experience in the field study, i.e. FITD-relevant items be presented only to the FITD group, and so on. Users who registered after the start of the experiment (and hence were not assigned to any group) were randomly assigned to one of the three groups for the purpose of the survey.

Subjects were asked to rate several statements on a 7-point Likert scale (1-completely disagree to 7-completely agree) with regards to each scenario.

Group	Survey Item	M	SD
FITD	Initial contact: How would you respond?		
	I would do the task because it was requested by Cyclopath.	5.53	1.23
	I would do the task because it is small and easy.	5.42	1.54
	I would do the task regardless of who requested it.	3.13	1.73
FITD	Initial contact: How would you feel after you did it?		
	I have done my part towards helping Cyclopath.	5.47	1.16
	I am the type of person who typically responds to requests from community sites.	4.70	1.65
FITD	Target request: How would you respond?		
	Asking me to do something else after only a week has gone by is too often.	4.62	1.32
	I'm the type of person that helps Cyclopath whenever called upon.	4.43	1.39
LB	Initial contact: How would you respond?		
	I would commit to such a request because it was requested by Cyclopath.	5.57	1.35
	I would commit to such a request regardless of who requested it.	2.65	1.58
	I would commit to such a request because it is small and easy.	4.98	1.50
LB	Target request: How would you respond?		
	This is unfair, Cyclopath just increased the amount of work it asked me to do.	4.47	1.50
	I'm a person of my word: I agreed to help, so I'll do it now.	4.63	1.50
Control	Target request: How would you respond?		
	I would do the task because it was requested by Cyclopath.	5.13	1.39
	I would do the task regardless of who requested it.	3.00	1.63
	It is too much of a favor to ask on the part of Cyclopath.	2.64	1.15

Table 3.5: Results of our follow-up survey.

3.5.3 Results and Discussion

We received 148 responses to our survey by April 20, 2011, a week after it was launched. Out of these, 43 (29%) of the users were among those who had participated in the previously described field study, whereas 105 (71%) were new. There were no statistically significant differences in the responses of these two sets of users. Table 3.5 presents the most important results. Our survey results showed some condition-specific patterns and some general population-wide patterns.

We observed general ideological patterns across all groups: respondents said they would do the requests because they came from Cyclopath ($M \geq 5.13, 1.23 \leq SD \leq 1.38$ in all cases) and would not do so in general ($M \leq 3.13, 1.58 \leq SD \leq 1.73$ in all cases). Moreover, respondents from the control group disagreed with the statement that the task requested of them (the target request) was too much of a favor ($M = 2.64, SD = 1.15$)! This shows users' general commitment to the Cyclopath community and can be a powerful motivation to contribute whenever requested.

FITD. Respondents from the FITD condition agreed with the statement that they would perform the task requested in the initial contact because it was small and easy ($M = 5.42, SD = 1.54$) – a verification of the FITD design. Further, we saw that FITD respondents agreed with the fact that they felt that they had *done their part towards helping Cyclopath* ($M = 5.47, SD = 1.16$) and that asking them to do something more after just a week had passed by was too often ($M = 4.62, SD = 1.32$). This is an indication that a likely reason for the lower-than-expected contributions to the FITD target request was that people felt that they were badgered.

The results also seem to indicate that this reason outweighs the self- perception effect: contributions from the FITD group in the field study were lower despite FITD respondents in the survey agreeing that after doing the task requested during the initial contact, they would feel like they are the type of people that would respond to requests for help from community sites ($M = 4.70, SD = 1.65$).

We did not find any statistically significant evidence for reciprocity; however, one respondent said, *"yeah if you help us back, this ain't a one-way street buddy"* as a response to how he/she would feel after responding to the initial contact.

LB. Respondents from the LB group verified the LB design by agreeing to the statement that they would commit to the request in the initial contact because it was small and easy ($M = 4.98, SD = 1.50$). Further, they also agreed to the statement that they would contribute to the target request because they were *persons of their word* ($M = 4.98, SD = 1.50$); in other words, committed to doing the task requested.

3.6 Implications

We believe this work is useful to both researchers and practitioners designing interventions to increase community participation. The success of the LB technique showed that an intervention as subtle as requesting prior commitment produced large, significant increases in participation. This low-cost-high-returns mechanism is potentially of great use to large scale campaigns such as WikiProjects and crowd-sourcing initiatives. However, care should be taken while applying theories from social psychology to online communities: there are subtleties and points of caution that must be considered.

3.6.1 Compliance-without-pressure in the online world

Earlier in the paper, we outlined how features such as anonymity and the loose, detached nature of online communication (as opposed to the more involved face-to-face) motivated the need for an empirical evaluation of the FITD and LB techniques. Looking at our results in the light of these premises, we see that the LB technique succeeded due to a heightened feeling of commitment despite the relative anonymity associated with the online community. However, FITD only produced limited success because it was very easy for subjects to simply close their browsers if they felt badgered by the repeated requesting. Thus, differences between online and offline interactions must be carefully considered before planning any intervention using a compliance-without-pressure

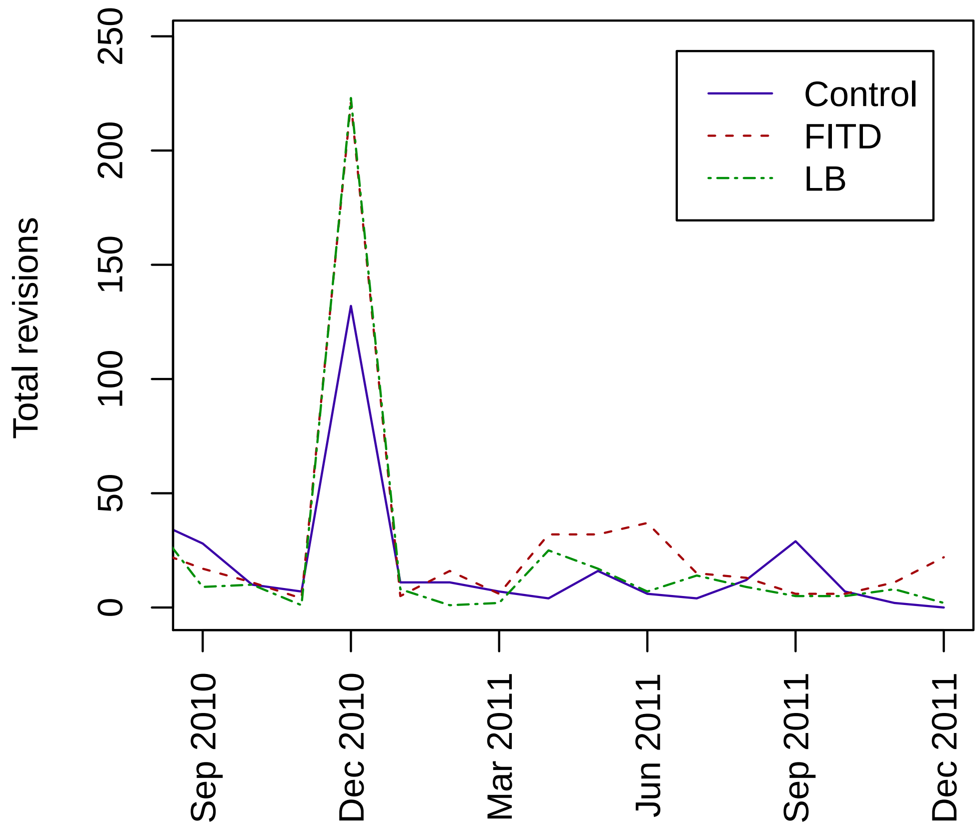

Figure 3.6: One-shot Effect. Our intervention saw a burst of activity just after the experimental emails were sent out (Dec, 2010), which dropped off steeply soon after. Note that the activity of Cyclopath's most prolific contributor has been excluded from this figure.

technique.

3.6.2 Short versus long-term effects

We have seen in the past that interventions that try to boost member participation succeed in doing so only while they are active, and that participation drops to pre-intervention levels once it is withdrawn [85]. Our FITD and LB interventions fell within the same category. We saw a burst of activity (measured as number revisions saved) immediately following the initial contact and the target request, which dropped off steeply soon after (Figure 3.6).

The self-perception effect, one of the explanations for the success of the FITD technique, is one of the most likely processes that could have caused long-term effects, whereas psychological reactance would work against any. In our case, we did not see any evidence for self-perception, which is consistent with our understanding of why our work campaign produced a one-shot effect.

While the failure to cause a long-term increase in participation levels is a limitation of the techniques we evaluated, we believe that they have utility in one-shot interventions. Online social production communities do require one-shot bursts of member participation at times – e.g, if a WikiProject in Wikipedia wishes to categorize several hundred articles into a hierarchical structure, or if a geographic open content community like FixMyStreet[7] needs to identify all potholes in a locality as a part of a neighborhood campaign.

That said, research efforts should try to focus on producing long-term increases in participation. In other words, researchers must try to devise techniques that enable the community to support itself. One way of tackling this issue is to employ the powerful and ubiquitous norm of reciprocity, an approach we are currently exploring.

3.6.3 Member satisfaction and retention

From past research, we know that although users can be manipulated into behaving in a particular manner, they tend to detect the manipulation and become less satisfied with the system [23]. This reduces the value of perhaps the greatest asset of an online community: the trust and commitment of its members.

Similarly, in the context of our study, we must consider the possible effects on member satisfaction of implementing the type of compliance-without-pressure techniques we studied. Our users certainly felt that the repeated emailing was excessive (in case of FITD), and that increasing the cost of the requested task was unfair (LB) (Table 3.5). From an online community designer's perspective, users might be viewed as *workers* – in which case they can

[7] www.fixmystreet.com

be manipulated to achieve maximum throughput – or as *community members* – in which case they must be carefully preserved and nurtured. Recognizing the strengths of the community (e.g., existing levels of user commitment) is critical: if apparently effective in the short term, techniques like FITD and LB might do more long term damage to users' perceptions of and commitment to a community. After all, the tangible product of a group activity is not the only outcome that matters; in the long run, users will continue to participate only if their needs are met [40, 58]. Finally, since there are so many online communities, users who are not satisfied with one are likely to leave it in favor of an alternative, more attractive one that serves the same purpose.

3.6.4 Limitations

An alternative explanation for the LB effect is the role of time: people tend to discount how busy they are or how much effort is needed when they make a commitment in the future [89]. LB studies that vary the amount of time between the initial contact and the target request are required to test this explanation.

Our study used a simple survey to conduct the qualitative follow-up to the field experiment. A limitation of post-hoc surveys are that they are presented out-of-context. Although we tried to mitigate this problem by making the survey as standalone as possible, better qualitative information could be collected if the survey was interwoven into the field experiment itself.

3.7 Summary

We described an evaluation of the compliance-without-pressure techniques—foot-in-the-door (FITD) and low-ball (LB)—in the context of an online social production community, using a field experiment and a follow-up survey. We found that despite the subtlety in its manipulation, the LB technique elicited nearly 3 times more participation. However, while both techniques managed to coerce users into performing actions that they would have otherwise not and produce an increase in contribution, this was short-lived. As a result, although techniques such as FITD and LB could be effective in certain short-term situations, we need to explore means for increasing and maintaining healthy levels of contributory participation in a long-term and sustainable manner.

One way of overcoming the short-term nature of the techniques evaluated in this chapter is to explore how information consumption—the most common type of activity—can be used as a source of contribution. By leveraging consumption which is a more natural process than contribution, we have the potential of increasing contribution without any of the harmful effects of techniques such as FITD and LB. Eliciting contribution in the context of consumption may invoke the powerful principle of reciprocity and may be a step towards establishing a form of a feedback loop, thereby helping a social production community sustain itself in the long term.

Chapter 4

Extracting Contributions from Consumption Activities

4.1 Introduction

Consumption is much more common than contribution in online, social production communities. Consumers have been estimated to outnumber contributers in Wikipedia by a ratio of 10,000:1 [43]. Other research on online groups reports that about 90% of the total membership of the online group never makes a contribution [52, 67]. We believe that harnessing consumption activities to drive contribution gives us two distinct advantages over techniques studied in the previous chapter. First, consumption is a much more typical activity than contribution. Recollect that our motivation for evaluating compliance-without-pressure techniques was to strengthen the effects of mechanisms such as intelligent task routing by persuading participants to perform actions (i.e. contribution) they would naturally not do. Hence, one way of overcoming the negative effects of foot-in-the-door and low-ball is to concentrate on actions that participants naturally do perform, e.g. consumption. Second, eliciting contributions in the context of consumption may be more effective due to the powerful principle of reciprocity: *Cyclopath and its users' contributions gave me this route; hence, I am obliged to give back in the form of contributions of my own.* Thus, we may be able to reduce the problem of under-contribution in a sustainable and long-term way, and reduce some of the harmful side-effects of techniques like those studied in the previous chapter.

To that end, this chapter presents an idea to leverage the process of evaluation and feedback that occurs naturally during information consumption and provide a pathway for the user to translate some of the feedback content into simple forms of contribution. Specifically, in Cyclopath, the idea translates to enabling consumers to submit specific information about road segments just after they have obtained a route, while their evaluation of the route is still fresh in their minds. We argue that this idea is likely to work through two steps: First, we establish that route evaluations contain contributory potential—information that, through appropriate techniques, may be assimilated into Cyclopath's database—by analyzing their textual representations that users shared with us (Study 1). Second, we study one such technique—specifying an alternative route by dragging the route line—and show that using machine learning algorithms, a part of the contributory potential can be automatically identified from a route-drag action thereby making it easier for the consumer to translate their route evaluation into concrete contributions such as ratings, tags and notes (Study 2).

4.2 Related Work

4.2.1 Nature of Online Social Production

Most contribution in online social production communities is done by a few contributors [84]. Most participants are information consumers. Many will never contribute, while others may be learning about the community [72] and may transition into contribution and organization-related activities as they gain more experience [82].

An important motivation to contribute is to fill gaps or fix problems; in other words, to improve information consumption. For example, Bryant et al. found that new users primarily use Wikipedia for information gathering, and identify problems and mistakes in passing and fix them [10]. In the context of open source software development, Hertel et al. discovered that one reason people contributed code was to make the software meet their own needs [48]. Prior research on Cyclopath revealed that a major reason for users to edit the map and rate road segments is to get Cyclopath to compute the route they desire [79]. However, there are barriers to contributing information, including: (a) a perception that one has no information to contribute, and (b) the combination that contribution is not required and that it is perceived to be a high-cost activity [74].

Prior research has studied ways to motivate users to contribute information to public resources. Intelligent Task Routing systems match users with tasks they are likely to be willing and able to perform; these systems have been shown to boost contributions [22, 85]. However, matching tasks to users accurately is challenging. *Gamification* is a popular technique that attempts to make activities fun and game-like; the ESP Game is the seminal example [100]. Finally, techniques that manipulate users into participating and contributing information may succeed in the short-term but might cause longer-term harm, because as we noted in the previous chapter, users tend to recognize the manipulations and may consider them unfair [23, 66].

4.2.2 Nature of User Feedback

Several instances of prior research and experience have highlighted the utility of user feedback. First, user feedback can contain useful content. It is well known in the field of user interface design, that users are better able to communicate interface requirements when presented with prototypes, even if they are low-fidelity. Various research studies and practical use cases have demonstrated the utility of employing the *recognition-over-recall* principle for effectively capturing user requirements and reactions to interfaces [71]. Also, within the information retrieval domain, implicitly captured user feedback has been shown to be useful for building a user model and inferring user preferences [53, 92].

Second, user feedback on the output of a system can be used to improve the system itself. Chen and Pu [16] suggest that a combination of system-generated and user-driven critiquing is the best way for a recommender system to incorporate user feedback and improve its recommendations. An evaluation of a simple tool on Wikipedia demonstrated that consumers can be persuaded to make small contributions in a low-cost, low-risk way [43] through the medium of providing feedback on the article they are reading. Several artificial intelligence and machine learning techniques rely on user feedback to improve their internal models and performance, including the well-known Winston Learning Algorithm which does so through examples and counter-examples [106].

Third, the cognitive processes behind generation of feedback are often automatic and hence, low-cost for people. When presented with a stimulus, natural human tendency is to develop an evaluation about it. To some extent, this construction of evaluation is automatic and happens without one's knowledge. This is due to the use of associative/heuristic-based cognitive information processing [26, 93]—one of the two channels described by dual-processing models of cognition. Evaluations that are more "intuitive" or affective, involving how one subjectively feels about the stimulus, appear to be more associatively driven, compared to more analytic, rational judgments such as those about causation [27]. Further, it seems likely that more richly detailed, specific stimuli are better cues for responses from the associative system [26].

Prior work suggests our opportunity: a pathway from naturally-occurring evaluation and feedback processes to contributions of information has the potential to overcome barriers to contribution and shortcomings of previous work elicitation techniques.

4.3 This Research: Context and Outline

We instantiate our idea of a pathway from naturally occurring feedback to contribution in Cyclopath as follows. The user can submit feedback about the route, including specific segments, immediately after obtaining the route on Cyclopath, while their evaluation of the route is fresh in their minds. We present two studies as evidence for this method.

Study 1. We analyzed naturally occurring textual route feedback, finding that it contains contributory potential—information that, through

Question (Does this feedback...)	% Yes	N
...include positive evaluations about any roads/areas?	24%	394
...include negative evaluations about any roads/areas?	48%	392
...include any objective facts about any roads, their surroundings and/or vehicles on them?	51%	399
...suggest any alternative roads or areas to take?	39%	440

Table 4.1: Findings from our textual route feedback analysis: the majority of naturally occurring route feedback contains contributory potential. Combining the first two rows, 57% of comments contained either a positive or negative evaluation. The N varies across questions and is less than 488 because we ignored responses that did not meet our 67% agreement threshold for that question.

appropriate techniques, may be assimilated into Cyclopath's underlying database.

Study 2. We studied one such technique, user modification of a computed route by "dragging" the route. We showed that machine learning algorithms can identify automatically part of the contributory potential; this in turn makes it easier for users to translate their route evaluation into concrete contributions such as ratings, tags and notes.

4.4 Study 1: Analyzing Textual Feedback

Aim: To determine whether naturally occurring route feedback contains contributory potential – information that can be potentially assimilated by Cyclopath

4.4.1 Data

From May 2009 to January 2012, Cyclopath included a simple route feedback tool. When a user obtained a route from Cyclopath, a button was displayed next to the route details, inviting feedback about the route. When the user clicked this button, a simple form popped up that asked them to tell us how satisfied they were with route on a 5-point Likert scale (very dissatisfied to very satisfied) and optionally, express in text, what they thought of the route.

While this feature was active, 688 instances of route feedback were submitted, 488 of which had textual comments. The median comment length was 137 characters ($mean = 169.020, sd = 132.256$). 120 comments were submitted by 90 registered users, and 368 comments were submitted by anonymous users.

4.4.2 Method

We developed a coding scheme to evaluate the contributory potential contained in the route feedback comments. For each comment we asked: *Does this feedback...*

1. ...include positive evaluations about any roads/areas?
2. ...include negative evaluations about any roads/areas?
3. ...include any objective facts about any roads, their surroundings and/or vehicles on them?
4. ...suggest any alternative roads or areas to take?

These questions let us identify specific types of information useful for Cyclopath: positive/negative evaluations correspond to bikeability ratings of road and trail segments, and objective information corresponds translate to tags and notes. We also looked for the occurrence of alternative route suggestions since this is an intuitive way of expressing route feedback used on popular mapping sites like Google Maps.

We coded the route feedback via crowdsourcing, using CrowdFlower[1] to deploy the coding task on Amazon Mechanical Turk. CrowdFlower provides a service that augments Mechanical Turk's platform with survey-building interfaces, quality-control mechanisms and reporting tools. As a quality-control measure, we created 32 artificial route route feedback comments with known

[1] crowdflower.com

answers ("gold standard items") and added them to the set of 488 actual comments to be coded, resulting in a dataset of 520 total items[2]. We omitted responses submitted by workers whose accuracy on the gold standard items (their "trust score") was less than 70%.

We received an average of 3.14 responses per item. We only considered responses where there was at least 67% agreement between workers, where responses were weighted by the workers' trust scores. For example, if three workers had the same trust scores, then 67% agreement would mean a simple 2-out-of-3 majority. See the CrowdFlower documentation[3] for more detail on their quality control mechanisms.

4.4.3 Findings

Naturally occurring route feedback contains evaluations of roads or areas. 24% of comments contained positive evaluations about roads/areas, 48% contained negative evaluations, and 57% contained one or the other or both.

> "Totally smooth, pleasant route, not much interaction with cars, felt very safe." (positive evaluation)
>
> "Biking straight down Ramsey Hill from Summit Ave to Grand Ave is an awful idea and I would never bike that way. It is a very steep hill with a stop light at the bottom. No way." (negative evaluation)

Naturally occurring route feedback contains objective facts about roads, etc. 51% of comments contained such information.

> "The part of the route from 3.65 to 5.29 is wretched. The bicycle path (in the directions, it's called "unnamed bicycle path") is downright dangerous, too. The path crosses two sets of railroad tracks and the seams are dangerously deep and wide. In addition, the path sweeps down onto a driveway to an industrial site, and that driveway is filled with grit and rubble. I take my chances on Country Road C, which

[2]The recommended amount of quality control items is about 5% of the data set size.
[3]crowdflower.com/solutions/self-service/faqs

> is not great, either, but at least I avoid the wheel-swallowing railroad tracks."

Naturally occurring route feedback contains descriptions of alternative routes. 39% of comments contained such information.

> "A better route is getting on E River Road to Franklin Ave Bridge, cross the River, get on West River Road, take that to Greenway Trail and follow all the way to The Depot. Too much traffic on Cedar Ave."

Non-contributors offer feedback. We wondered whether most route feedback came from people who already contributed to Cyclopath; this would be disappointing, as it would decrease the chance of attracting new contributors through a route-feedback-based mechanism. However, this was not the case. First, as already noted, 368 of 488 route feedback comments came from anonymous users. And anonymous users account for a small proportion of contribution to Cyclopath (3,667 out of 17,559 revisions, or 21%). Therefore, we can infer that most of the anonymous route feedback came from non-contributors. Second, route feedback comments from anonymous users were more likely to include contributory potential than comments from registered users: negative evaluations (52% vs 37%, p = .012), objective information (55% vs 39%, p = .007), alternative routes (43% vs 28%, p = .010)[4]. Third, only half of the 90 registered users who submitted comments had contributed to Cyclopath. Therefore, route feedback offers a significant opportunity to obtain contributions from new sources.

Not only dissatisfied users provide useful feedback. Most useful content came from users who said they were dissatisfied with their routes. However, users who were satisfied or very satisfied with their routes accounted for 22% of comments with positive or negative evaluations of roads, 23% of comments with alternative route suggestions, and 24% of comments with objective information (see Figure 4.1).

[4]To compute statistical significance, we used the 2-sample test for equality of proportions with continuity correction.

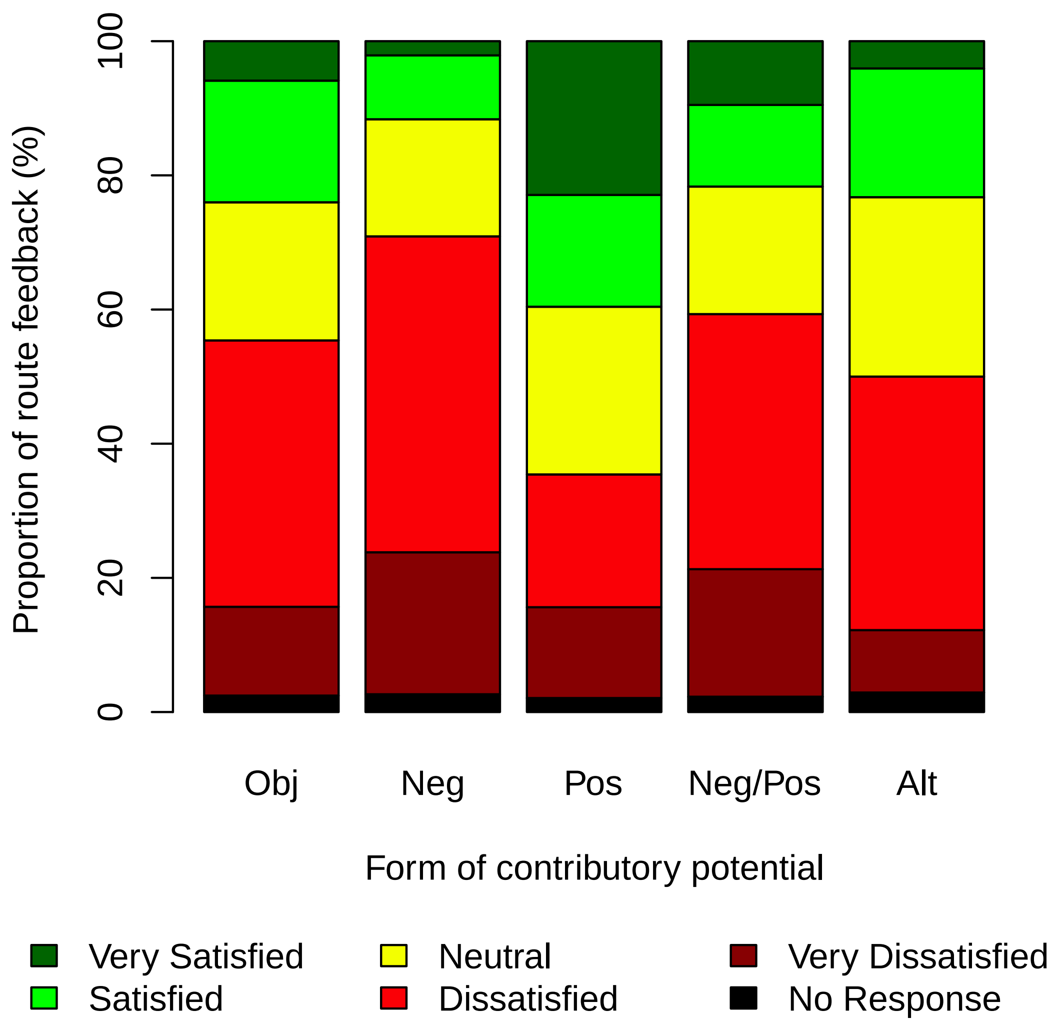

Figure 4.1: Relationship between users' satisfaction with routes and the contributory potential of their route feedback. Obj = Objective Information, Neg = Negative Evaluation, Pos = Positive Evaluation, Alt = Alternative Route Suggestion. It is not only the dissatisfied who submit useful feedback.

4.4.4 Discussion and Implications

Through an analysis of their textual representations, we found that naturally occurring route feedback is rich in several kinds of useful information. How can that information be mapped on to Cyclopath constructs? We believe that there are techniques that are more structured than plain text that can better capture this contributory potential:

Evaluations → Ratings. Evaluations of roads can be expressed as bikeability ratings. A simple ratings widget could be displayed along with a route, enabling users to quickly identify segments they want to rate higher or lower.

Objective Facts → Tags, Notes. Cyclopath represents factual information as tags (short phrases) and notes (longer, free-form text). As with ratings, an appropriate widget could be displayed to enable easy annotation of road segments. Further, techniques like Tag Expression [99] have been shown to be effective in eliciting tags and tag preferences.

Alternative Routes → *???* When a user specifies an alternative route, there is no direct correspondence to specific Cyclopath constructs. Indeed, various things could be changed in order for Cyclopath to produce a different route: bikeability ratings, tags, or the details of the routing algorithm (e.g., how various parameters are weighted). Instead, an alternative route suggestion – along with the rejection of the original route – constitutes an opportunity for Cyclopath to learn. *How* that learning might be done is the topic of the next study.

4.5 Study 2: Analyzing Visual Feedback

Aim: To investigate a semi-automated technique for learning from visual route feedback

We have shown that naturally occurring route feedback often includes a suggestion of an alternative route. Further, an intuitive means for users to specify an alternative route is by "dragging" the route, as on Google Maps and Cyclopath. There are two ways a system could respond to this learning opportunity: (1) Automatically modify the route planner's internal models and/or algorithms, thus producing better routes in the future [109, 110, 69]; (2) Engage the user in a dialogue to elicit reasons for preferring the new route over the old route [88]. These reasons then can be analyzed to produce new system knowledge.

We explored the second approach for the following reasons: (1) Richer types of information can be gathered, e.g., notes and tags in addition to ratings. Indeed, in the context of recommender systems, researchers have recommended a hybrid system-and-user-generated critiquing approach to improving personalization and system performance [16]. (2) The consumer is introduced to and nudged towards contribution through simple, easy steps. This is consistent with theories such as Legitimate Peripheral Participation [61] and the Reader-to-Leader framework [82] that model the transition from consumption to contribution. These approaches have been demonstrated to have value in the context of Wikipedia [43]. Thus, the result of a semi-automated, dialogue approach may be not just a *contribution*, but also a new *contributor*.

The following use case illustrates the type of dialogue we have in mind.

1. A user requests a route.

2. Cyclopath computes the best path based on available information.

3. The user is not satisfied with the route, and corrects it by dragging the route line to follow their preferred path.

4. Cyclopath automatically identifies segments along the original and new routes that it thinks caused the user to modify the route, highlights the segments, and offers the user the opportunity to provide ratings, tags, and notes for the segments. The user responds as he/she desires.

This use case is plausible under three conditions: (1) The user's route preference is based on preferences for individual segments of the route. (2) The system can identify the appropriate route segments with sufficient accuracy. (3) The user will provide information about these segments. We consider each of these in turn.

Users might prefer one route over another because of properties of individual segments of the routes: for example, they might want to avoid hilly or high traffic segments and instead ride on quiet, scenic off-road facilities. Alternatively, users might prefer one route over another for more holistic reasons, for example, to ride near certain resources (say, places to stop for a drink). In this research, we chose to focus on the former type of preference because it directly benefits the A* graph search algorithm—a popular route-finding algorithm and one that is also used by Cyclopath—which, at each step in the search process, considers only the properties of the possible next segments that could be added to the route (and not global properties like "make sure some part of the route passes within a kilometer of a place to get a drink"). Thus, we conducted a study to explore the reasons why users prefer one route over another, specifically, the extent to which they do so because of properties of specific segments (Study 2A).

If users do tend to prefer one route over another because of the properties of specific segments, then the next challenge is for the system to identify these segments. We therefore developed a machine learning classifier to predict such segments and evaluated its accuracy (Study 2B).

But even if the system can identify these segments – segments in an old route that a user is apt to dislike, segments in the user's suggested route that the user is apt to like – why would a user go to the trouble of providing knowledge (ratings, tags, notes) about these segments? Besides the likelihood that the user will have this information fresh in their mind, prior work offers compelling answers. As we noted above, fixing problems and improving one's own experience are common reasons people begin contributing to a social production system; Panciera et al. found that fixing problems was the most common reason for Cyclopath users to begin editing [79]. Moreover, prior work on Cyclopath has shown that focusing users' attention on specific geographic areas makes them more likely to contribute information [85]. Given this prior work, we did not study user motivation further in this research.

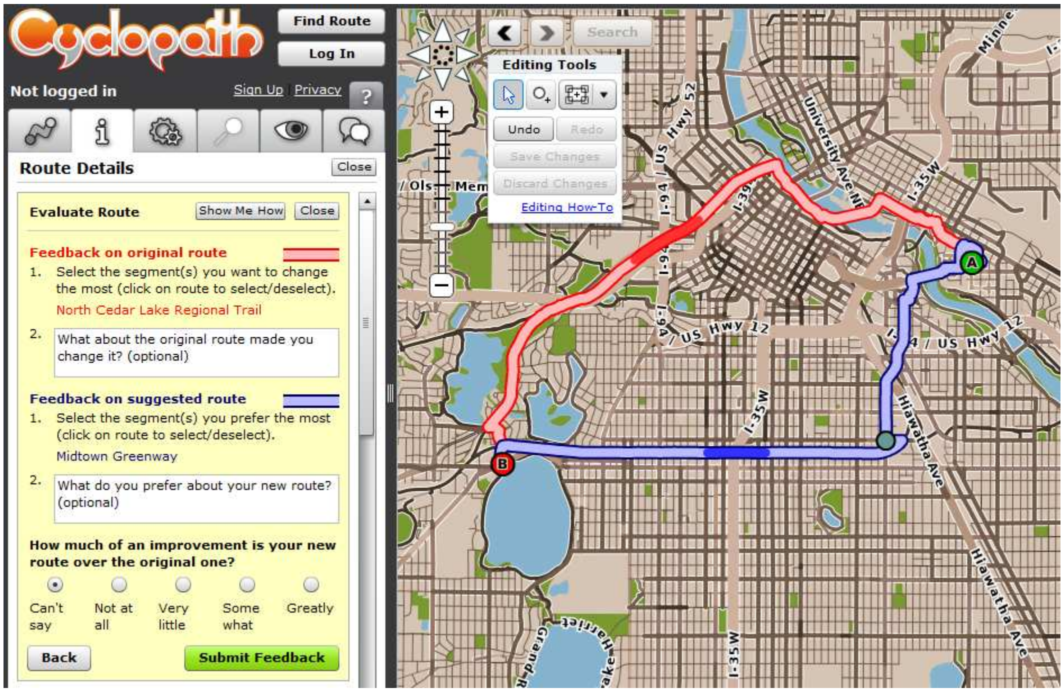

Figure 4.2: The Cyclopath interface to collect visual route feedback through dragging. This images is showing step 2 of the 2-step process where the user is being asked to mark road segment(s) that caused them to modify their route. The "Show Me How" button led users to a tutorial video describing the process.

4.5.1 Data and Method

We modified the Cyclopath interface to elicit visual route feedback. When a user generated a route using Cyclopath's route planner, a button was displayed in the route details panel labeled "Suggest Different Route". Clicking this button led to the user being guided through a two-step procedure.

In step 1, users were asked to drag the route to meet their preferences. (Route dragging was an existing feature of Cyclopath, which works similarly to other mapping applications such as Google Maps: clicking on the route line introduces an intermediate waypoint at the location of the click and dragging it changes the route line in such a way that it is forced to pass through the waypoint.)

In step 2, users were asked to provide specific feedback on the original route and their suggested route, including identifying specific segments of either route that most influenced their preference. These segments constituted a data set that we used to train and evaluate a machine-learning classifier. Users also were asked to estimate how much of an improvement their suggested route was over the route Cyclopath computed. Note that we implemented step 2 solely for the purpose of data gathering – in the wild, users would only be required to drag their routes for Cyclopath to suggest road segments to annotate.

We received 85 instances of route feedback between October 25, 2012 and April 16, 2013. Each feedback instance consisted of the original route, the suggested route, one or more waypoints, zero or more marked segments, optional reasons for changing the original route and suggesting the alternative and an evaluation of how much of an improvement the suggested route was over the original.

In 38 of these 85 cases, users had marked specific segment(s) as reasons for modifying the route. In most cases (36 of 38) where users marked segments, they also provided an explanation for modifying the route. An additional 17 feedback instances carried explanations (for changing the original or for suggesting the alternative) but no marked segments, making a total of 53 feedback instances with explanations.

As in Study 1, most of the feedback was submitted anonymously: 23 instances were submitted by 20 registered users and the remaining 62 anonymously. Anonymous users also provided more feedback than registered ones, accounting for 24 of 36 instances where liked/disliked segments, and 33 of 53 instances where comments were submitted. However, 18 of the 20 registered users had made at least 1 contribution.

4.5.2 Study 2A: Why did the user modify the route?

4.5.2.1 Aim

In this study, we explore two main questions:

1. Why did users modify their original route and suggest the alternative route?

(a) Entire section

Attribute	Original	Suggested	p	
Avg. annual daily motor traffic	6240	5250	$< .001$	***
Shoulder width (m)	0.831	1.16	$< .001$	***
Outer lane width (ft)	9.91	12.1	$< .001$	***
Speed limit (mph)	31.6	31.6	.79	
Avg. user bikeability rating	4.09	4.73	$< .001$	***

(b) Marked segments only

Attribute	Original	Suggested	p	
Avg. annual daily motor traffic	8030	4300	$< .001$	***
Shoulder width (m)	0.856	1.12	.309	
Outer lane width (ft)	12.0	13.5	.0502	.
Speed limit (mph)	32.0	31.0	.011	*
Avg. user bikeability rating	2.97	2.96	.943	

Table 4.2: Differences between the original and suggested route sections. All values are means. Suggested sections had lower automotive traffic, wider shoulders and outer lanes, and higher user-submitted bikeability ratings. All road attributes were loaded into Cyclopath when map data was initially imported from the Minnesota Department of Transportation in 2008 and except average annual daily traffic, the others can be modified and kept up to date by users. Statistical significance codes: .: < 0.1, *: < 0.05, **: < 0.01, ***: < 0.001 (using Welch two-sample t-test).

2. Specifically, did the users modify the route because of the road segment(s) they marked?

4.5.2.2 Method

We manually coded the comments users specified along with their visual route feedback to to understand why users chose to modify their routes. For the first question, coders assigned one or more of the following sets of codes to the reasons users gave for rejecting the original route and for suggesting the alternative:

- Longer vs. Shorter
- Unsafe vs. Safe

- No dedicated bicycle facilities vs. Dedicated bicycle facilities
- Hills vs. No hills
- Bad surface vs. Good surface
- High traffic vs. Low traffic
- Noisy vs. Quiet
- Many stoppages vs. Few stoppages
- Bad intersection vs. Good intersection

The author of this dissertation developed these codes by identifying prominent patterns within the feedback comments. These patterns are consistent with those used by prior research on bicycling route choice [94].

For the second question, we used the following ordered set of rules to code the comments. That is, if Rule 1 let us conclude that the user modified route because of the segment(s) they marked, then rules 2 and 3 were not considered. *We conclude that the user modified the route because of the segments they marked if...*

Rule 1. ...the user mentions a road segment by name in the feedback and has also marked a road segment with the same name. For example, one comment said *"Big hill on North St, too much traffic on University, freeway on/off traffic on 12th St"* and the user marked segments of North St. and 12th St.

Rule 2. ...the feedback content is highly synonymous with the road type and the contents of the tags and notes attached to the marked segments. To help make this determination, we augmented each feedback instance with road types, tags and notes attached to the corresponding marked segments. For example, a comment said *"This is a long, steep hill."* and the "hill" tag was applied to one of the segments the user marked.

Rule 3. ...we find factual, objective information with the feedback that we can verify applies to the marked segments using aerial imagery or by browsing around the area. For example, a comment said *"I don't wanrt [sic] to ride through the U [University of Minnesota]. Don't like the traffic and vibe."* and browsing the map around the marked segment reveals that the segments were within the University of Minnesota neighborhood.

Thus, for this coding task, coders assigned one of two codes to user comments: "Yes" (by Rules 1, 2 or 3) or "No evidence". Three independent coders coded the feedback instances for the both the coding tasks.

4.5.2.3 Findings

Why did users modify their routes? We found evidence for several important reasons. The most common reasons for rejecting the original route were *high traffic* (17/53; inter-rater agreement for this category using Fleiss' Kappa, $\kappa = 0.85$) and *hilliness* (10/53; $\kappa = 0.91$), whereas those for suggesting the alternative route were *presence of dedicated bicycle facilities* (12/53; $\kappa = 0.79$) and *low traffic* (11/53; $\kappa = 0.79$).

These findings are consistent with prior research finding that for commuter bicyclists, the most important factors in choosing a route are: travel time, the presence of a bicycle facility (especially a bike lane or separate path and especially on a bridge), the level of automobile traffic, quality of the pavement or riding surface [94].

This is also consistent with the data we collected. Table 4.2 shows quantitative differences between the original and suggested sections of routes. We see that on an average, the suggested section had much lower automotive traffic, wider shoulders and outer lanes, and a higher average user-submitted bikeability rating.

Did users modify their routes because of the segment(s) they marked? Yes, in many cases. Marking segments was an optional step of the data collection process—users could successfully submit route feedback by simply dragging the route and doing nothing else. Despite this, in 29 out of

85 cases, users marked segment(s) as a response to the prompt "Select the segment(s) you want to change the most" and in 26 out of 85 cases, users marked segment(s) as a response to the prompt "Select the segment(s) you prefer the most". In sum, in 38 out of 85 cases (45%), users indicated at least one segment as a reason why they modified their route.

Further, combining "Yes (by Rule 1)", "Yes (by Rule 2)" and "Yes (by Rule 3)" into a single code "Yes", we found evidence that the marked segments were a part of the reason why users modified the route in 23 of the 36 instances (64%) where users had marked segments ($\kappa = 0.50$). Note that this is a lower-bound: in the remaining 34% cases, we could not reliably conclude from our manual coding whether the feedback was about the segments marked.

Finally, as a response to the question, "How much of an improvement is your new route over the original one?", we obtained the following results: *can't say* (34 cases), *very little* (3 cases), *somewhat* (26 cases) and *greatly* (22 cases). This suggests that obtaining information about why users preferred one route over another, and then mapping it into a form usable by the route finder will result in demonstrably better routes.

4.5.3 Study 2B: Predicting liked/disliked segments

4.5.3.1 Aim

Given a user's route modification, can the system accurately identify the road segments that most influenced the user's preference for one route over another? In other words, given the original and suggested sections, can we predict the segments that the users marked?

4.5.3.2 Method

We adopted a machine learning approach for this task. For our choice of classifier to use, we chose the Random Forest classifier due to its sensitivity towards handling data sets with imbalanced classes. We trained and evaluated this classifier using the Weka data mining software package [107]. We evaluated the classifier using 10-fold cross-validation.

We used several attributes of the road segments in question, the routes (original and suggested versions) to which they belonged, and the user that requested the routes (and submitted the feedback, whenever available):

- **Road attributes:** Type of the road (local road, highway, bike trail, etc.), lane width, lane count, shoulder width, speed limit, average annual daily automotive traffic, number of user-recorded bikeability ratings, average user bikeability rating, Cyclopath-estimated bikeability rating (calculated based on road properties), presence of bike lane (or similar tags such as bike trail), presence of hills, presence of roughness (or similar tags such as unpaved), presence of traffic (or similar tags such as busy).
- **Route attributes:** Length (in meters), number of road segments, average bikeability rating.
- **User properties (when available):** Number of revisions made, number of routes requested, number of map views done, number of days since registering for Cyclopath.

We used three primary metrics in reporting the performance of our machine learning classifiers: sensitivity, specificity, and area under the Receiver-Operator-Characteristic (ROC) curve (also known as Area-Under-Curve or AUC) [29]. Our metrics may be interpreted as:

- **Sensitivity:** proportion of marked road segments that are correctly classified
- **Specificity:** proportion of unmarked road segments that are correctly classified
- **AUC:** a single scalar value representing the overall performance of the classifier.

We used the 0-R classifier to provide a baseline model as a frame of reference for interpreting our results. A 0-R algorithm always predicts the most commonly occurring class. Due to the design of our data gathering procedure

(a) Given the original section, predict the road segment(s) the user marked.

Attributes Used	Sensitivity	Specificity	AUC
Road	0.346	0.978	0.814
Road + User	0.365	0.987	0.857
Road + Route	0.738	0.984	0.951
Road + User + Route	0.757	0.985	0.962

(b) Given the suggested section, predict the road segment(s) the user marked.

Attributes Used	Sensitivity	Specificity	AUC
Road	0.339	0.983	0.870
Road + User	0.347	0.984	0.883
Road + Route	0.771	0.989	0.964
Road + User + Route	0.792	0.989	0.974

Table 4.3: Classifier performance. Using all three sets of attributes (road, user and route), we can detect about 76% of marked segment(s) for the original section of the route, and about 79% for the suggested section. The values are average of 10-fold cross-validation. User attributes had a small effect on sensitivity because a majority of the feedback was submitted anonymously.

(the instructions we gave to the users), there will be many fewer marked than unmarked road segments in any given old/new route pair. So, given an old (or new) route section, the 0-R algorithm will always predict that none of its road segments were marked: sensitivity = 0, specificity = 1, AUC = 0.5. Note that our baseline outperforms random guessing, which would converge to an overall performance of sensitivity = 0.5 and specificity = 0.5.

4.5.3.3 Findings

Our classifier outperformed the baseline 0-R in all our cross-validation tests. On an average, given the original section of the route, the marked road segments within it were correctly identified in about 76% (sensitivity = 0.757) of the cases when road, route and user attributes were all utilized for prediction (see Table 4.3(a)).

Similarly, given the suggested section of the route, the marked segments were correctly identified in about 79% (sensitivity = 0.792) of the cases on an average by using all three sets of attributes (see Table 4.3(b)).

4.5.4 Discussion, Implications and Limitations

In this study, we have presented a technique to leverage the intuitive action of correcting a route when dissatisfied with it to capture the content of the route evaluation. We now address the generality, utility and limitations of this study.

4.5.4.1 Generality of Results and Ideas

Social production communities such as Wikipedia and Cyclopath offer contributors tasks across a wide spectrum of expertise and difficulty: some tasks are quite easy and do not require topic or work-type expertise (e.g. slightly modifying a route in Cyclopath, correcting a typo in Wikipedia) whereas other tasks can require both (e.g. fixing road segment connectivity around a complex intersection, or composing and editing an entire article in Wikipedia). Prior research suggests that consumers take their first steps into becoming a contributor by performing the simpler tasks [61, 82], a pattern that is observed in both Wikipedia [10] and Cyclopath [65]. Fixing or tweaking the output that the system computes and presents is one such simple task that consumers have a propensity to do: after all, it brings them direct benefit. Consequently, in Study 2, we explored how route dragging can be used to engage the consumer in a dialogue that may automatically read a part of the user's natural route evaluation and by focusing attention on specific segments, lower the cost for translating the rest into concrete contributions such as ratings, tags and notes by focusing user attention on specific segments.

This basic idea is applicable in a wide spectrum of social production systems. For example, when a recommender system presents the consumer with a list of recommended items, the consumer could correct it by rating the items listed or specifying different items. To some extent, Amazon already does this with the "Fix this recommendation" feature. Similarly, another domain where this idea is applicable is the collaborative construction of the semantic web.

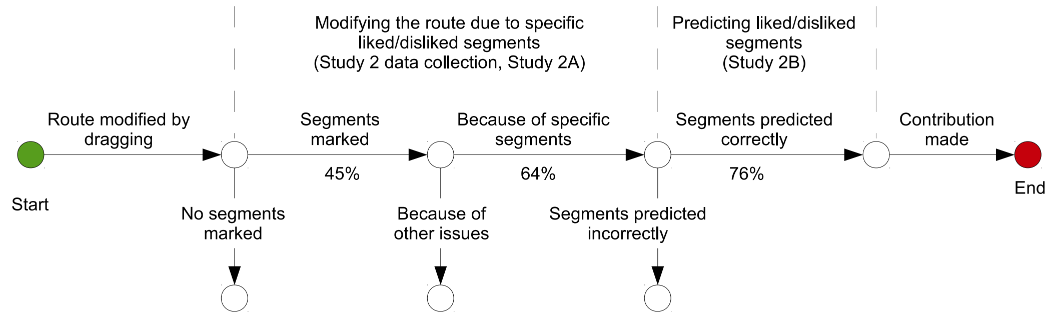

Figure 4.3: Flowchart illustrating the back-of-the-envelope calculation used to estimate the probability that a user would make a contribution after dragging a route through the technique explored in Study 2. Note that in a real-life use-case, the user would drag their route and then directly contribute information about the predicted segments – the intermediate steps would happen internally (either within the user's mind or within the system).

It may be easier to identify the inaccuracy and correct the output of a information query on a structured knowledge graph, than to identify the holes by browsing through the graph itself.

4.5.4.2 Estimation of Utility

How useful might the contributions resulting from the system-user dialog explored in Study 2 be? Study 2A highlighted several reasons users mentioned for modifying their routes. Many of the reasons can be expressed as tags: e.g., hilliness can be represented by the *hill* tag and dedicated bicycle facilities can be recorded using the *bikelane* tag. By automatically predicting a part of users' feedback in the form of especially liked/disliked segments, we can reduce the cost of contribution to just that involved in submitting ratings, tags, and notes.

A more interesting approach is to estimate the potential "conversion rate", i.e., the proportion of route-modification actions that might be leveraged to obtain a contribution. It is precisely those times when users modify computed routes (by dragging) when they are dissatisfied with the route; thus, these constitute opportunities for Cyclopath to learn. Our results let us estimate the conversion rate. To do so, must enumerate and analyze the various steps involved in the process. See Figure Figure 4.3 for a back-of-the-envelope version of this analysis. Recall that for a route-modification action to be "converted" to a contribution, two things must happen:

1. *The modification must be due to the user liking/disliking specific segments.* Study 2A showed that in 45% of cases when users described an alternative route, they marked segments they liked/disliked; further, we have evidence that at least 64% of the time, segments were marked because of issues with the segments themselves, as opposed to issues with their surroundings. Thus, we that when a user modifies a computed route, the probability that specific road or trail segments caused the modification is $45\% \times 64\% = 29\%$.

2. *Cyclopath should be able to successfully predict these segments.* In Study 2B, we built a predictor that could identify these segments with a sensitivity of about 76%. In other words, when users modified a route because of specific segments, the classifier had a 76% probability of predicting the segments correctly.

Thus, if each route-modification action with correctly-predicted segments yielded one contribution, we would get a route-modification-to-contribution conversion rate of about $29\% \times 76\% = 22\%$ (22 contributions per 100 route-modification actions). This may be an underestimate, as prior research has shown that users may contribute beyond what they are asked directly [85]. A controlled field study is required to provide more concrete evidence in this regard.

How sustainable is this process for eliciting contributions? Since users modify routes for self-benefit, we think the first part of the process will continue to be "triggered" until the route finder is perfect for all users... which likely never will happen. And we already have presented an argument why users are likely to provide simple explanations (in the form of ratings, tags, and notes) for why they modified a route.

Moreover, the utility extends beyond getting more *contributions*: we may be able to get more *contributors* too. Indeed, research has shown that having consumers do simple contribution tasks, such as providing feedback, may convert them into regular contributors [43]. This is valuable in and of itself, because an increased inflow of contributors into the community is an important force to combat attrition and increase the diversity and quality of the community resource [3].

4.5.4.3 Limitations

An important limitation of the process we have outlined for eliciting contributions from a route modification is that it captures only one possible type of preference, i.e., preference for (or against) individual route segments. However, as we noted, it is possible that users may have more holistic preferences. For example, one user gave the following reason for modifying their route: "It brings me by Freewheel Cycle."

Another limitation of this process is that dragging-based feedback depends on the road network being connected correctly. For example, intersection faults (road segments that visually cross but are not connected in the database) can cause routing to work incorrectly. While collecting visual route feedback, we observed several cases where users who had the motivation and skills to make geographic edits, made the required connections in the road network in order to enable them to drag their route. On the other hand, some users who had the motivation but not the skills wrote to us complaining that they were not able to drag the route on paths they wanted. After investigating why, we found out that some roads were not connected appropriately. We connected these roads and the users were then able to record their feedback correctly.

Some users mentioned this limitation as a part of their feedback comments.

> "Shorter distance, wide shoulders, and new path on east side of Flying Cloud between ValleyView and Technology. Couldn't make the route follow Flying Cloud there..."
>
> "System doesn't seem to know that both Sibley and Jackson connect to Shepard Rd"

With a modification to the route-dragging mechanism, we believe we can convert this limitation to an advantage. For example, if we made the route-dragging permissive enough to allow the route to treat intersection faults like regular connections, we may be able to either automatically create connections and fix the faults, or at least flag them for user attention.

4.6 Summary

In this research, we have explored the possibility of harnessing a commonly-found information consumption behavior to drive contribution into a social production system. We have done so in the context of Cyclopath and route-planning, shown that route feedback contains contributory potential, and outlined a process for focusing it on specific road segments.

Chapter 5

The Utility of Online Social Production: The Case of Urban Planning

5.1 Introduction

Having studied how online social production happens in Cyclopath and a few techniques that may be able to increase it, let us step back and examine the utility of this phenomenon. Cyclopath is driven by the efforts of many citizens of the Minneapolis-St. Paul metropolitan area. Do all these efforts and findings make it a better resource the population that it was originally designed and built for? Does it make the Twin Cities a "smarter" and a more efficient urban habitat? Prior research on Cyclopath has provided evidence that citizens have benefited from the use of Cyclopath. For example, Figure 5.1 shows modifications made to the map by citizens resulted in routes becoming shorter due to missing road connections being identified [85]. In this chapter, we shall extend our understanding of the benefits imparted by the social production on the Cyclopath platform, by looking at a different dimension: transportation planning.

Any urban area involves an interplay between the citizens that define it and the governments that manage it. Therefore, any solution that aims to

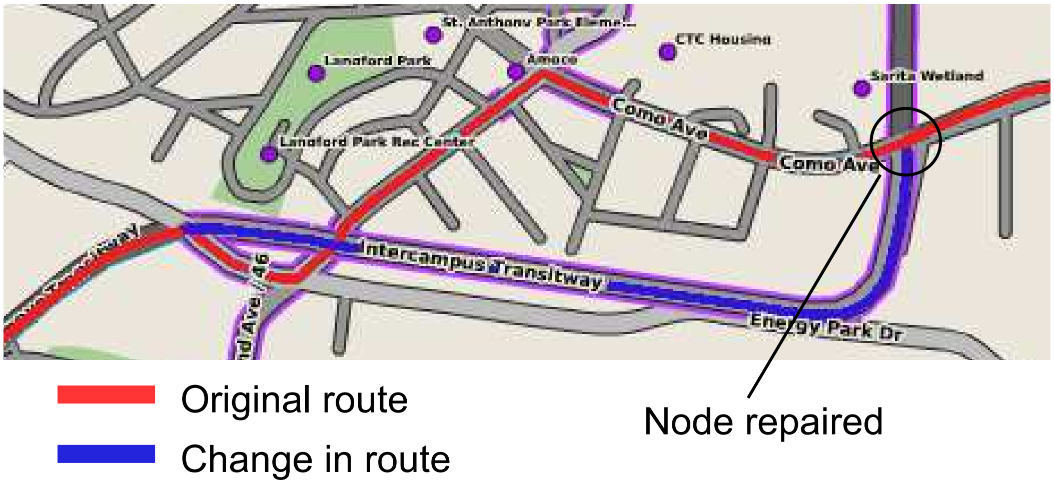

Figure 5.1: An example of route improved from 15.6 km to 15.0 km due to information contributed by Cyclopath users (Figure borrowed from [85].)

make a city smarter must address the problems of, and cater to both these stakeholders. The same holds true for Cyclopath. In this chapter, we describe how Cyclopath helps the bicycle transportation planners of the Twin Cities metropolitan area, by enabling them to harness the citizen knowledge gathered by Cyclopath and perform novel analyses that can strengthen the planning process.

The advent of the Internet and the Social Web in the recent years is producing an increasing use of volunteered geographic information (VGI) as a mechanism to inform and complement traditional urban planning processes. Here, we describe the design of a novel bicycling route analysis tool built on top of Cyclopath that empowers planners to analyze cycling routes and connectivity in the transportation network by tapping into the local knowledge of citizens through an easy-to-use web interface.

5.2 Related Work

Goodchild introduced the term *Volunteered Geographic Information* (VGI) as the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic informa-

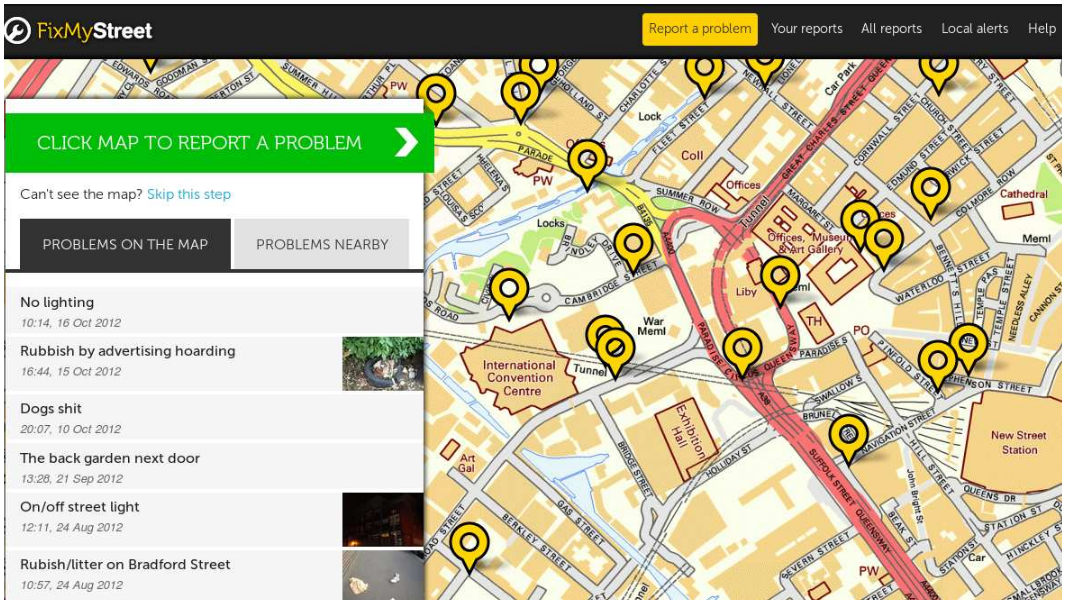

Figure 5.2: FixMyStreet.com.

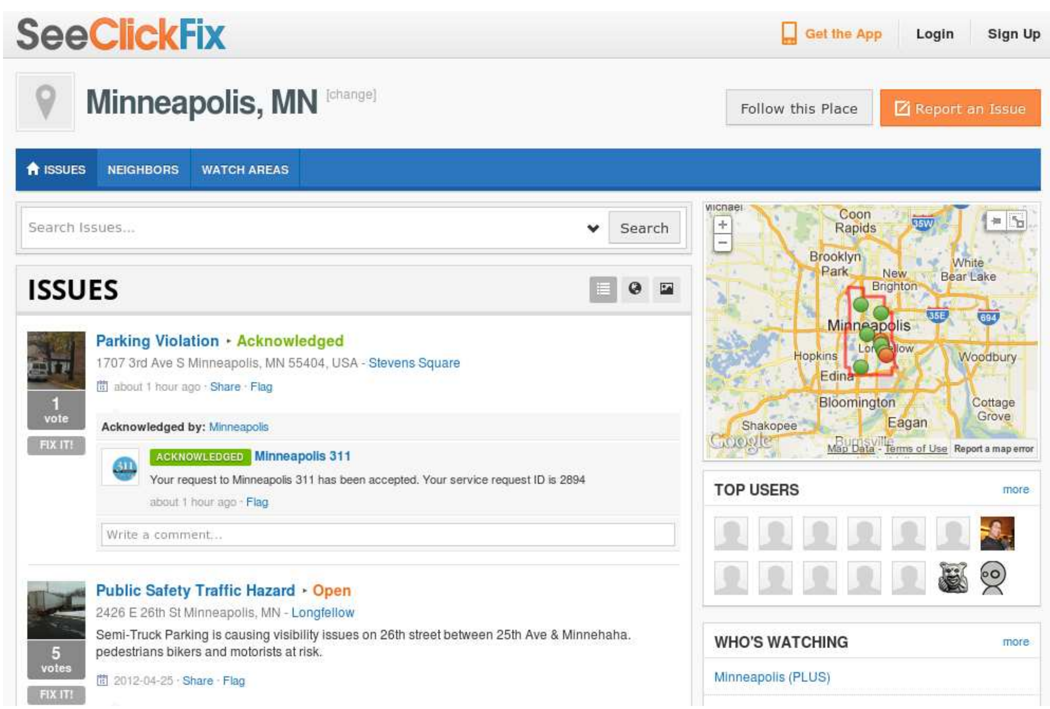

Figure 5.3: SeeClickFix.com.

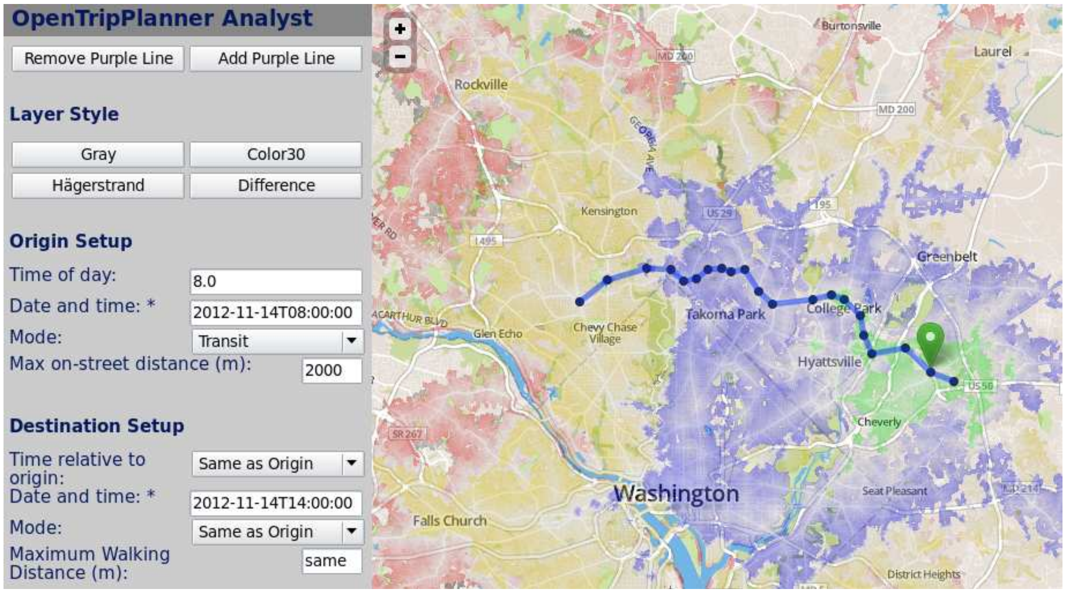

Figure 5.4: OpenTripPlanner Analyst.

tion [36]. Crossing paths with the increased amount of citizen involvement in urban decision-making, we have seen VGI being leveraged for urban planning purposes [90].

Accordingly, the recent years have seen the development of several VGI-based citizen science applications that have been useful for planning purposes. FixMyStreet [55] (Figure 5.2) and SeeClickFix [70] (Figure 5.3) empower citizens to report potholes and local issues from their neighborhood on a map so that the attention of the city authorities can be directed towards them. Portland Metro's MetroMap [9] enabled residents to access city maps, and perform their own analysis, interpret the results, and make policy suggestions as a part of the Metro's Region 2040 planning effort. OpenTripPlanner Analyst[1] (Figure 5.4) is a route analysis tool developed by OpenPlans to analyze transit trip times, demand for service, and perform system-modification analyses.

[1] openplans.org/work/opentripplanner-analyst/

5.3 System Design

We built a route analysis system by extending Cyclopath. Cyclopath records all routes requested by its users. Additionally, being a wiki, it has collected over 80,000 bikeability ratings, 5,400 tag applications and 7,500 notes about road segments in its transportation network. We use this data to power our route analysis tool.

5.3.1 Inputs

- **From, To:** Our route analysis tool analyses routes between two sets of regions. Each region set can be specified by name (e.g. "Minneapolis, St. Paul") or by tag (e.g. all regions tagged "park").

- **Source of Destinations:** End-points or destinations to compute routes between can be chosen in three ways. *User-generated* destinations are extracted from actual routes requested by Cyclopath users. These are the closest to reality. *Synthetic* destinations are randomly chosen intersections in the given geographic region. These are useful to analyze routes from/to regions that have not seen sufficient route requests. *Past-job* destinations are those which were used in a previous route analysis task. These are useful when comparing the results of two different analyses with each other.

- **Rider Profile:** Our tool analyses routes using a given set of riding preferences organized into profiles. Each profile contains a preference for shorter vs. more bikeable routes and a set of liked and disliked road properties expressed using tags (e.g., prefer "scenic" roads, and avoid "bumpy" roads).

- **Number of Routes:** This field represents the number of routes to be analyzed.

- **Map Revision:** Since Cyclopath is a wiki, analysts can choose to analyze routes using the current (up-to-date) version of the map, or a historic version. The map revision field represents this choice.

5.3.2 Output

The route analysis tool output consists of a set of Esri shapefiles and a text file with a few descriptive statistics. The most important shapefile is the one with the entire transportation network with each road segment carrying a count representing the number of routes analyzed that pass through it, its length, average rating, and number of user ratings.

5.4 Usage Scenario: A "What-If" Analysis

5.4.1 Aim

The Metropolitan Council, the regional planning agency serving the 7-county Minneapolis-St. Paul metro area, is currently building a light-rail transit line connecting the downtowns of Minneapolis and St. Paul, dubbed the "Green Line". As a part of this planning and construction, they want to analyze bicycle connectivity between Green Line train stations and nearby cities and neighborhoods. For example, one of the analyses they are interested in is to examine connectivity between the city of Roseville and the Green Line Hamline Avenue station (see Figure 5.5(a)). They are interested in identifying connectivity bottlenecks and evaluating potential solutions to reduce them.

5.4.2 Study 1: Bikeability Gap Analysis

Amy is a fictional transportation planner from the Metropolitan Council and her organization has received funding to improve bikeability and connectivity in this area, and several alternatives have been identified. Amy's task is to evaluate a design that would create a new, bike-friendly bridge over the railroad tracks. She uses Cyclopath's geo-editing tools to create a new region (polygon) surrounding the intersections around the Green Line Hamline Avenue station. Then, she starts a new route analysis task, analyzing 500 routes between previously user-generated destinations between the city of Roseville and the newly-created region above, using the default rider profile, which evenly balances distance with bikeability and avoids prohibited and closed roads (These are the default settings for Cyclopath's route-finder).

When the task completes, Amy downloads the generated shapefile and using the popular GIS tool, ArcGIS, she creates a visualization showing roads with thickness based on the number of routes that pass through them and colored based on their average rating (see Figure 5.5(b)). Looking at the result, she clearly discovers that despite being considered bike-unfriendly, the Snelling Avenue bridge tends to be a recommended route to cyclists.

This is because that bridge is the only segment providing a north-south connection across the railroad tracks in its immediate neighborhood. Further, it is rated low for cycling because it is a state highway with a lot of heavy traffic. This indicates that an alternative bike-friendly connection would improve the connectivity between Roseville and the Green Line Hamline Avenue station.

5.4.3 Study 2: Alternatives Analysis

To evaluate whether a new bike-friendly connection across the railroad tracks would improve the connectivity for Roseville, Amy uses Cyclopath's map editing tools again to create a new segment extending Hamline Avenue across the tracks. She rates this new, proposed segment high on bikeability and repeats the above route analysis process ensuring that the same set of route destinations are analyzed, and obtains a new shapefile.

Creating a similar visualization using ArcGIS, she sees that all routes from Roseville to the Green Line Hamline Avenue station now employ the newly-added segment instead of the Snelling Avenue bridge. This gives her confidence to recommend to her department that constructing a new bike-friendly bridge across the railroad tracks near Hamline Avenue would improve the connectivity significantly for residents of Roseville.

(a) Study 1: Bikeability Gap Analysis

(b) Study 2: Alternatives Analysis

Figure 5.5: Evaluating bike connectivity: Thickness of the lines are proportional to the number of routes that pass through those road segments; colors represent bikeability (green: more bikeable, red: less bikeable).

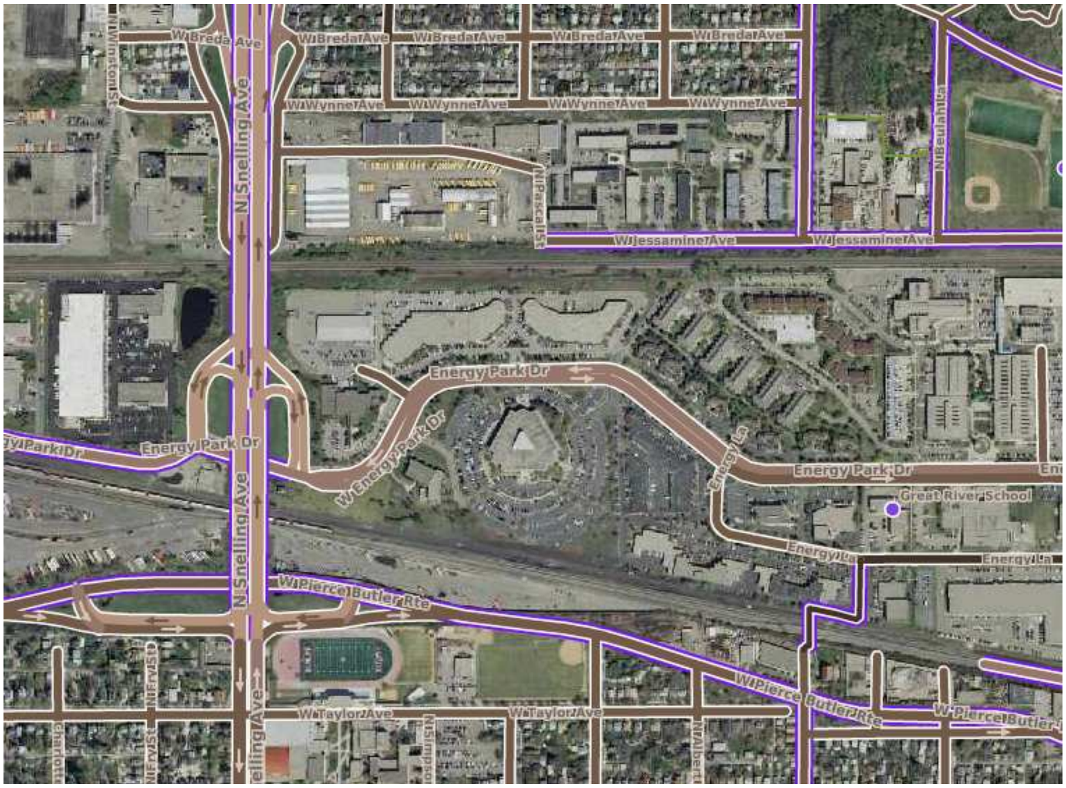

Figure 5.6: Purple highlights on road segments indicate presence of user-submitted notes and discussions.

5.5 Discussion: The Geowiki Factor

Route analysis on Cyclopath gets its distinctive advantage over other, similar, planning tools due to the underlying geowiki architecture. First, Cyclopath's transportation network is maintained by citizen cyclists who often add shortcuts such as parking lots (segments valuable to cyclists, but not found on traditional maps that are designed for motor vehicles) to the map. Second, user-gathered bikeability ratings make for more personalized route-finding [86]. Together, these factors make route analysis more accurate and closer to the ground reality.

Moreover, user-gathered notes help in understanding the results of route analysis tasks (Figure 5.6). For example, the following notes about the bike-unfriendly Snelling Avenue bridge help understand why users dislike it:

> Due to the freeway nature of the road, I prefer to ride on the sidewalk and slow down for the (very) occasional pedestrian.

> On and off ramps to/from Como make this area a danger for unexperienced bikers. I also use the sidewalk.

On the other hand, a route analysis system based on crowd-sourced geographic information also has its challenges. Firstly, in the wiki architecture, access to the data is not controlled: anyone with an Internet connection can make edits to Cyclopath. While quality control tools such as the Recent Changes list and the ability to revert a potentially malicious edit easily help mitigate this issue, there is still a chance that incorrect information seeped into the analysis process. Secondly, the issue of trust in citizen-reported information is an important one. Traditionally, geographic information has been collected, curated and managed by experts in Geographic Information Systems (GIS) and cartography. The recent developments in VGI is moving some of this control to the citizens; and a geowiki like Cyclopath is going a step further by making the map publicly-editable. Lastly, the work processes behind transportation planning activities would need appropriate modifications to accommodate citizen-reporting and validation.

5.6 Summary

In this chapter, we have described an example of the utility of online social production through a novel transportation planning tool based on Cyclopath, that aims to leverage the power of VGI for planning purposes more effectively.

Chapter 6

Summary and Conclusion

6.1 Summary

In this dissertation, we have studied several aspects of social production in the context of Cyclopath, a geographic wiki. In Chapter 2, we studied how Cyclopath users contributed to the resource. We discovered that despite the freedom available, most users contributed information in specialized units containing either only one type of work, or certain specific object + annotation combinations. We also discovered that contributors also went on to specialize in specific work types over their life-cycle with most experienced ones choosing to specialize in editing blocks. Based on these findings, we recommended designing coherent work units while carrying out campaigns to elicit contributions, and using techniques such as intelligent task routing to match users with the appropriate type of work based on their interests.

In Chapter 3, we attempted to improve the core of intelligent task routing and most simple work campaigns by investigating how to increase users' compliance with requests to do work. We evaluated foot-in-the-door and low-ball —two of the most popular social psychological techniques—and discovered that while such techniques might succeed in the short term (indeed, our experiments showed that the low-ball technique produced a remarkable one-shot boost in contributions), they may cause some long-term harm because of their nature of trying to get people to do something that would not ordinarily do.

Chapter 4 attempted to overcome this limitation by trying to harness an activity that people naturally do perform – the consumption activity of route-planning. In this chapter, we first established that when a user obtains a route on Cyclopath, the evaluation of it that often automatically develops in their mind, contains information that can be assimilated into the resource. Then, we described how the intuitive process of modifying one's route when dissatisfied with it can lead to an opportunity for the system to elicit useful information from the consumer, possibly introducing them to contribution.

While the above chapters described academic research that studied how user participation happens in online social production communities and ways and means for increasing it, Chapter 5 stepped back and looked its one particular utility: how do user contributions on Cyclopath make it a better tool for transportation planners? By enabling better planning decisions, Cyclopath has the potential to make Minneapolis and St. Paul a smarter urban habitat where citizens not only help each other, but also their governing bodies.

6.2 Generality of Research

Most of the research described in this dissertation was conducted in the context of Cyclopath, a geographic wiki. Although showing trends of increasing in popularity, geographic wikis are still relatively uncommon on the Web. The natural question, then, is how does this research generalize? How does a researcher or a practitioner take the lessons learned from the research described in this thesis and apply them in the real world?

We think there are several ways in which this may be done. Systems like Cyclopath might be uncommon, but most of the research described in this thesis is not dependent on Cyclopath being a spatial or geographic system. Indeed, most of the findings and implications can be directly translated to other, non-spatial and non-geographic communities with relative ease. For example, the concept of work-type specialization as studied in Chapter 2 in the context of Cyclopath, is not unique to it: Editors on Wikipedia contribute in various different ways – some make simple edits such as correcting spelling and grammatical errors and ensuring that the article follows the style guidelines set by the community, whereas others make much bigger and complex edits, such as composing an entire article or organizing a collection of articles. Q&A sites such as StackOverflow even recognize different types of work participants do by awarding them different badges and reputation scores.

Similarly, the compliance-without-pressure techniques studied in Chapter 3 can be implemented in any online social community that wishes to generate a burst of increased participation to meet a short-term demand. For example, a WikiProject may employ such a technique to organize a set of articles into a particular hierarchy.

Finally, the success of the process to extract contributions from consumption actions as described in Chapter 4 is dependent on a fundamental design artifact: the ability to consume information by computation, as opposed to direct-access. For example, in Cyclopath, using the route planner to compute a route would be consumption through computation, whereas inspecting the map and visually charting out a route would be direct-access. Similarly, in a recommender system such as Amazon, using the recommendation engine to identify a list of products to purchase during the holiday season is computation, whereas building such a list by manually navigating through the product directory would be direct-access. The core idea of this paper is that if the output of a computation could be easily and naturally "corrected" by the consumer, then with some simple help from the consumer, the system can utilize this "corrective action" to improve itself.

When viewed in this perspective, the idea goes beyond Cyclopath and becomes applicable to a wider spectrum of online social production systems. For example, if the consumer could correct Amazon and instead, specify the list of products that they would like to see, Amazon could compare the two lists and automatically deduce the consumer's preferences. To some extent, Amazon already does this with the "Fix this recommendation" feature. Similarly, another domain where the idea of this paper is applicable is the collaborative construction of the semantic web. It may be easier to identify the inaccuracy and correct the output of a information query on a structured knowledge graph, than it is to identify the holes by browsing through the graph itself.

6.3 Future Work

Every piece of research not only provides answers to unanswered research questions, but also poses new questions for future work. This dissertation is no exception. While we did further our understanding of how contributors participated in Cyclopath and found some success in evaluating various techniques to increase contribution, there are a few unanswered questions.

The quantitative methods we used in Chapter 2 do not provide the "whys" behind specialization. Do people intend to specialize? Do people care if there are specialists available in the community? Do people notice when they change their specialization? Can they explain why they change? Surveys, interviews, and other qualitative methods are needed to answer these questions. Similarly, in Chapter 4, an important future work for us is to investigate further into how users choose to express their feedback in the context of a modified route. We might do this by inviting the user to modify their route (just like our Study 2) and if the user accepts the invitation, employ our new classifier to automatically predict road segments that might be the reason the user modified the route, and highlight them on the map. Then, for each highlighted segment, we could ask the user to express feedback about the segment in one or more of three different ways: ratings, tags and notes. We may then be able to study if there is any association between the route modification and the existence and nature of the information elicited. Finally, our future work on the just described route analysis project includes going beyond anecdotal evidence by working with transportation planners and other stakeholders to understand how such a tool fits in their workflow through an extensive user study.

Stepping back a bit, we would like to emphasize that online social production has a tremendous potential to support and shape the societies in which we live. Hence, to the extent possible, research and engineering should try to design systems that connect people in novel ways, and just like Cyclopath, help citizens help each other better. Many recent communities are already doing this. Online social production has been used to support societies in need of community support such as communities in crisis situations, people with disabilities and populations living in developing countries. For example, Ushahidi[1] is helping people in crisis situations such as earthquakes, famines and civil wars exchange information with each other and with relief organizations through a combined mobile + web approach. Similarly, the Humanitarian OpenStreetMap group[2] is establishing a bridge between humanitarian responders on the ground and the volunteer mappers of the popular open-content web map, OpenStreetMap to assist in crisis response and economic development of under-served populations. Online social networks such as Inclusive Planet[3] are giving people with visual impairments a home on the web, a place where they can meet new friends and make new career-related connections. Online social production has also been used to shape societies by helping organize community activism [68]. For example, using story-telling, Hollaback[4] is helping societies recognize and stand up to the problem of street harassment and eve-teasing. The White House hosts a service where citizens themselves can initiate the change their society needs by organizing petitions and gathering popular response[5]. In short, as future work, we would like to urge researchers, designers and engineers (ourselves included) to examine our innovation efforts through the lens, "how is this shaping and supporting the society in which we live?"

[1] www.ushahidi.com

[2] hot.openstreetmap.org

[3] www.inclusiveplanet.com

[4] www.ihollaback.com

[5] petitions.whitehouse.gov

Chapter 7

References

This thesis includes much or most of [64, 65, 66]. Also, Chapter 4 is a working version of a paper in preparation with Loren Terveen.

[1] Ban Al-Ani, Gloria Mark, Justin Chung, and Jennifer Jones. The egyptian blogosphere: a counter-narrative of the revolution. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 17–26. ACM, 2012.

[2] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Steering user behavior with badges. In *Proceedings of the 22nd international conference on World Wide Web*, pages 95–106. International World Wide Web Conferences Steering Committee, 2013.

[3] Ofer Arazy, Wayne Morgan, and Raymond Patterson. Wisdom of the crowds: Decentralized knowledge construction in wikipedia. In *16th Annual Workshop on Information Technologies & Systems (WITS) Paper*, 2006.

[4] Alexander Ardichvili, Vaughn Page, and Tim Wentling. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of knowledge management*, 7(1):64–77, 2003.

[5] Flore Barcellini, Françoise Détienne, and Jean-Marie Burkhardt. A situated approach of roles and participation in open source software communities. *Human–Computer Interaction*, (just-accepted), 2013.

[6] D.J. Bem. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological review*, 74(3):183–200, 1967.

[7] Yochai Benkler. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press, 2006.

[8] Jessica Bernstein-Wax. What motivates participation in online communities, 2007.

[9] M. Bosworth, J. Donovan, and P. Couey. Portland metro's dream for public involvement. *Community participation and geographic information systems*, 2002.

[10] Susan L Bryant, Andrea Forte, and Amy Bruckman. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 1–10. ACM, 2005.

[11] J.M. Burger. Increasing compliance by improving the deal: The that's-not-all technique. *Journal of Personality and Social Psychology*, 51(2):277, 1986.

[12] J.M. Burger. The foot-in-the-door compliance procedure: A multiple-process analysis and review. *Personality and Social Psychology Review*, 3(4):303, 1999.

[13] J.M. Burger and T. Cornelius. Raising the Price of Agreement: Public Commitment and the Lowball Compliance Procedure. *Journal of Applied Social Psychology*, 33(5):923–934, 2003.

[14] J.M. Burger and R.E. Petty. The low-ball compliance technique: Task or person commitment. *Journal of Personality and Social Psychology*, 40(3):492–500, 1981.

[15] Brian Butler, Lee Sproull, Sara Kiesler, and Robert Kraut. Community effort in online groups: Who does the work and why? *Leadership at a distance: Research in technologically supported work*, pages 171–194, 2002.

[16] Li Chen and Pearl Pu. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction*, 22(1-2):125–150, 2012.

[17] R.B. Cialdini et al. Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology*, 31(2):206, 1975.

[18] R.B. Cialdini et al. Low-ball procedure for producing compliance: Commitment then cost. *Journal of Personality and Social Psychology*, 36(5):463–476, 1978.

[19] E Gil Clary, Mark Snyder, and Robert Ridge. Volunteers' motivations: A functional strategy for the recruitment, placement, and retention of volunteers. *Nonprofit Management and Leadership*, 2(4):333–350, 1992.

[20] John Coate. Cyberspace innkeeping: Building online community. *Reinventing technology, rediscovering community: Critical explorations of computing as a social practice*, pages 170–172, 1997.

[21] D. Cosley et al. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proc. CHI*, 2006.

[22] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, pages 32–41. ACM, 2007.

[23] Dan Cosley, Shyong K Lam, Istvan Albert, Joseph A Konstan, and John Riedl. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 585–592. ACM, 2003.

[24] G. Demartini. Finding experts using Wikipedia. In *Proc. Workshop on Finding Experts on the Web with Semantics at ISWC/ASWC*, 2007.

[25] Paul Dourish and Sara Bly. Portholes: supporting awareness in a distributed work group. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 541–547. ACM, 1992.

[26] Seymour Epstein. Cognitive-experiential self-theory of personality. *Handbook of psychology*, 2003.

[27] Seymour Epstein, Abigail Lipson, Carolyn Holstein, and Eileen Huh. Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology*, 62(2):328, 1992.

[28] Thomas Erickson and Wendy A Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)*, 7(1):59–83, 2000.

[29] Tom Fawcett. ROC graphs: Notes and practical considerations for researchers. *Machine Learning*, 31:1–38, 2004.

[30] L. Festinger. *A theory of cognitive dissonance*. Stanford University Press, 1957.

[31] Andrew J Flanagin, Kristin Page Hocevar, and Siriphan Nancy Samahito. Connecting with the user-generated web: how group identification impacts online information sharing and evaluation. *Information, Communication & Society*, (ahead-of-print):1–12, 2013.

[32] J.L. Freedman and S.C. Fraser. Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4(2):195–202, 1966.

[33] E.J. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Ann Arbor*, 1001:48109–1092, 1999.

[34] Paul Fugelstad, Patrick Dwyer, Jennifer Filson Moses, John Kim, Cleila Anna Mannino, Loren Terveen, and Mark Snyder. What makes users rate (share, tag, edit...)?: predicting patterns of participation in online communities. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 969–978. ACM, 2012.

[35] E. Gilbert and K. Karahalios. Using social visualization to motivate social production. *IEEE Transactions on Multimedia*, 11(3):413, 2009.

[36] M. F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 2007.

[37] R.E. Guadagno et al. When saying yes leads to saying no: Preference for consistency and the reverse foot-in-the-door effect. *Personality and Social Psychology Bulletin*, 27(7):859, 2001.

[38] N. Guéguen. Foot-in-the-door technique and computer-mediated communication. *Computers in Human Behavior*, 18(1):11–15, 2002.

[39] N. Guéguen and C. Jacob. Fund-raising on the web: the effect of an electronic foot-in-the-door on donation. *CyberPsychology & Behavior*, 4(6):705–709, 2001.

[40] J.R. Hackman. The design of work teams. *Handbook of organizational behavior*, 315:342, 1987.

[41] Mordechai Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *Pervasive Computing, IEEE*, 7(4):12–18, 2008.

[42] Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. The rise and decline of an open collaboration system: How wikipedias reaction to popularity is causing its decline. *American Behavioral Scientist*, 2012.

[43] Aaron Halfaker, Oliver Keyes, and Dario Taraborelli. Making peripheral participation legitimate: reader engagement experiments in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 849–860. ACM, 2013.

[44] Aaron Halfaker, Aniket Kittur, and John Riedl. Don't bite the newbies: how reverts affect the quantity and quality of wikipedia work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pages 163–172. ACM, 2011.

[45] Il-Horn Hann, Jeff Roberts, Sandra Slaughter, and Roy Fielding. Why do developers contribute to open source projects? first evidence of economic incentives. In *2nd Workshop on Open Source Software Engineering, Orlando, FL*, 2002.

[46] H. Harari et al. Faculty Helpfulness to Students. *Personality and Social Psychology Bulletin*, 6(3):373, 1980.

[47] F. Harper et al. Social comparisons to motivate contributions to an online community. *Persuasive Technology*, pages 148–159, 2007.

[48] Guido Hertel, Sven Niedner, and Stefanie Herrmann. Motivation of software developers in open source projects: an internet-based survey of contributors to the linux kernel. *Research policy*, 32(7):1159–1177, 2003.

[49] J. Hornik et al. Comparison of three inducement techniques to improve compliance in a health survey conducted by telephone. *Public Health Reports*, 105(5):524, 1990.

[50] R.V. Joule. Tabacco deprivation: The foot-in-the-door technique versus the low-ball technique. *European Journal of Social Psychology*, 17(3):361–365, 1987.

[51] Steven J Karau and Kipling D Williams. Social loafing: A meta-analytic review and theoretical integration. *Journal of personality and social psychology*, 65(4):681, 1993.

[52] Jon Katz. Luring the lurkers, slashdot, 1998.

[53] Diane Kelly and Jaime Teevan. Implicit feedback for inferring user preference: a bibliography. In *ACM SIGIR Forum*, volume 37, pages 18–28. ACM, 2003.

[54] Amy Jo Kim. *Community building on the web: Secret strategies for successful online communities.* Peachpit Press, 2006.

[55] Stephen F. King and Paul Brown. Fix my street or else: Using the Internet to voice local public service concerns. In *Proc. ICEGOV '07*, pages 72–80, 2007.

[56] A. Kittur, E. Chi, B.A. Pendleton, B. Suh, and T. Mytkowicz. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proc. alt.CHI*, 2007.

[57] Joon Koh, Young-Gul Kim, Brian Butler, and Gee-Woo Bock. Encouraging participation in virtual communities. *Communications of the ACM*, 50(2):68–73, 2007.

[58] R.E. Kraut. Applying social psychological theory to the problems of group work. *HCI models, theories and frameworks: Toward a multidisciplinary science*, pages 325–356, 2003.

[59] Karim Lakhani and Robert Wolf. Why hackers do what they do: Understanding motivation and effort in free/open source software projects. 2003.

[60] Shyong Tony K Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R Musicant, Loren Terveen, and John Riedl. Wp: clubhouse?: an exploration of wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pages 1–10. ACM, 2011.

[61] Jean Lave and Etienne Wenger. *Situated learning: Legitimate peripheral participation.* Cambridge university press, 1991.

[62] K. Ling et al. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4):00–00, 2005.

[63] Gloria J Mark, Ban Al-Ani, and Bryan Semaan. Resilience through technology adoption: merging the old and the new in iraq. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 689–698. ACM, 2009.

[64] Mikhil Masli, Landon Bouma, Andrew Owen, and Loren Terveen. Geowiki + route analysis = improved transportation planning. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 213–218. ACM, 2013.

[65] Mikhil Masli, Reid Priedhorsky, and Loren Terveen. Task specialization in social production communities: The case of geographic volunteer work. *Proc. ICWSM*, 2011.

[66] Mikhil Masli and Loren Terveen. Evaluating compliance-without-pressure techniques for increasing participation in online communities. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 2915–2924. ACM, 2012.

[67] Bruce Mason. Issues in virtual ethnography. *Ethnographic studies in real and virtual environments: Inhabited information spaces and connected communities*, pages 61–69, 1999.

[68] Elaine Massung, David Coyle, Kirsten F Cater, Marc Jay, and Chris Preist. Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 371–380. ACM, 2013.

[69] Lorraine McGinty and Barry Smyth. Personalised route planning: A case-based approach. In *Advances in Case-Based Reasoning*, pages 431–443. Springer, 2000.

[70] I. Mergel. Distributed democracy: Seeclickfix.com for crowdsourced issue reporting. *Social Science Research Network*, 2012.

[71] Jakob Nielsen and JoAnn T Hackos. *Usability engineering*, volume 125184069. Academic press Boston, 1993.

[72] Blair Nonnecke and Jenny Preece. Why lurkers lurk. In *Americas Conference on Information Systems*, pages 1–10, 2001.

[73] Blair Nonnecke and Jenny Preece. Silent participants: Getting to know lurkers better. In *From usenet to CoWebs*, pages 110–132. Springer, 2003.

[74] Blair Nonnecke, Jenny Preece, and Dorine Andrews. What lurkers and posters think of each other [online community]. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 9–pp. IEEE, 2004.

[75] A. Okabe, K.I. Okunuki, and S. Shiode. The SANET toolbox: New methods for network spatial analysis. *Transactions in GIS*, 10(4):535–550, 2006.

[76] O. Okolloh. Ushahidi, or'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action*, 59, 2009.

[77] Leysia Palen, Kenneth M Anderson, Gloria Mark, James Martin, Douglas Sicker, Martha Palmer, and Dirk Grunwald. A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. In *Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference*, page 8. British Computer Society, 2010.

[78] K. Panciera, A. Halfaker, and L. Terveen. Wikipedians are born, not made: A study of power editors on Wikipedia. In *Proc. GROUP*, 2009.

[79] Katherine Panciera, Mikhil Masli, and Loren Terveen. How should I go from _ to _ without getting killed?: motivation and benefits in open collaboration. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, pages 183–192. ACM, 2011.

[80] Katherine Panciera, Reid Priedhorsky, Thomas Erickson, and Loren Terveen. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1917–1926. ACM, 2010.

[81] J. Preece. *Online communities.* Wiley & Sons, 2001.

[82] Jennifer Preece and Ben Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32, 2009.

[83] Reid Priedhorsky. *The Value of Geographic Wikis*. PhD thesis, University of Minnesota, August 2010. http://reidster.net/pubs/thesis.pdf.

[84] Reid Priedhorsky, Jilin Chen, Shyong Tony K Lam, Katherine Panciera, Loren Terveen, and John Riedl. Creating, destroying, and restoring value in wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*, pages 259–268. ACM, 2007.

[85] Reid Priedhorsky, Mikhil Masli, and Loren Terveen. Eliciting and focusing geographic volunteer work. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 61–70. ACM, 2010.

[86] Reid Priedhorsky, David Pitchford, Shilad Sen, and Loren Terveen. Recommending routes in the context of bicycling: algorithms, evaluation, and the value of personalization. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 979–988. ACM, 2012.

[87] B.D. Ripley. Tests of randomness for spatial point patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 368–374, 1979.

[88] Seth Rogers, Claude-Nicolas Fiechter, and Pat Langley. An adaptive interactive agent for route advice. In *Proceedings of the third annual conference on Autonomous Agents*, pages 198–205. ACM, 1999.

[89] M.M. Roy et al. Underestimating the duration of future events: Memory incorrectly used or memory bias?. *Psychological Bulletin*, 131(5):738, 2005.

[90] C. J. Seeger. The role of facilitated volunteered geographic information in the landscape planning and site design process. *GeoJournal*, 72(3), 2008.

[91] S. Sen et al. tagging, communities, vocabulary, evolution. In *Proc. CSCW*, page 190, 2006.

[92] Xuehua Shen, Bin Tan, and ChengXiang Zhai. Implicit user modeling for personalized search. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 824–831. ACM, 2005.

[93] Eliot R Smith and Jamie DeCoster. Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and social psychology review*, 4(2):108–131, 2000.

[94] Monique A Stinson and Chandra R Bhat. Commuter bicyclist route choice: analysis using a stated preference survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1828(1):107–115, 2003.

[95] J.T. Tedeschi et al. Cognitive dissonance: Private ratiocination or public spectacle?. *American Psychologist*, 26(8):685, 1971.

[96] Fernando Torre, S Andrew Sheppard, Reid Priedhorsky, and Loren Terveen. bumpy, caution with merging: An exploration of tagging in a geowiki. In *Proceedings of the 16th ACM international conference on Supporting group work*, pages 155–164. ACM, 2010.

[97] T.C. Turner, M.A. Smith, D. Fisher, and H.T. Welser. Picturing Usenet: Mapping computer-mediated collective action. *Computer-Mediated Communication*, 10(4), 2005.

[98] Julita Vassileva and Lingling Sun. Using community visualization to stimulate participation in online communities. *E-Service Journal*, 6(1):3–39, 2007.

[99] Jesse Vig, Matthew Soukup, Shilad Sen, and John Riedl. Tag expression: tagging with feeling. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 323–332. ACM, 2010.

[100] Luis Von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.

[101] Georg Von Krogh, Sebastian Spaeth, and Karim R Lakhani. Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7):1217–1241, 2003.

[102] M. Wattenberg, F. Viégas, and K. Hollenbach. Visualizing activity on Wikipedia with chromograms. *INTERACT*, 2007.

[103] H. Welser, G. Kossinets, S. Marc, and D. Cosley. Finding Social Roles in Wikipedia. In *Proc. Annual Meeting of the American Sociological Association*, 2008.

[104] H.T. Welser, E. Gleave, D. Fisher, and M. Smith. Visualizing the signatures of social roles in online discussion groups. *Social Structure*, 8(2), 2007.

[105] Ruth L Williams and Joseph Cothrel. Four smart ways to run online communities. *Sloan Management Review*, 41(4):81–92, 2000.

[106] Patrick H Winston. Learning structural descriptions from examples. Technical report, DTIC Document, 1970.

[107] Ian H Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2005.

[108] Donghee Wohn, Alcides Velasquez, Tor Bjornrud, and Cliff Lampe. Habit as an explanation of participation in an online peer-production community. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 2905–2914. ACM, 2012.

[109] Brian D Ziebart, Andrew L Maas, Anind K Dey, and J Andrew Bagnell. Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 322–331. ACM, 2008.

[110] Karen Zita, Haigh Jonathan, Richard Shewchuk, and Manuela M Veloso. Exploiting domain geometry in analogical route planning. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(4):509–541, 1997.