Ratings Use in an Online Discussion System: The Slashdot Case

by

Clifford A. Lampe

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2006

Doctoral Committee:

Professor Paul J. Resnick, Chair
Professor W. Russell Neuman
Professor Judith S. Olson
Associate Professor Mark S. Ackerman
Marc A. Smith, Microsoft Corporation

Dedication

For Mary Beth

Acknowledgements

A dissertation depends on the effort and support of many people. Friends, family, colleagues, classmates and advisors all leave their indelible print on a work like this, and I hope to some extent they all feel some ownership of this document. Several people deserve individual recognition for their investment and assistance in this process.

First, I would like to extend my deepest thanks to my advisor, Paul Resnick. Paul's intelligence, dedication, enthusiasm and experience were essential in guiding me through the dissertation process. He blended encouragement and constraint in the right way to help make the dissertation better at each step in the process. Paul invested valuable time and effort into this project at every turn, and was constantly willing to help. He is the type of thoughtful, brilliant researcher that I hope to become someday.

I am also indebted to the other members of my committee. It was a fortuitous meeting with Judy Olson that drew me to the School of Information, a diversion for which I'll always be grateful. Judy is an incredible mentor to graduate students at SI, and her contributions are invariably precise and accurate. Russ Neuman helped me to think about this research in a context outside of the information sciences, and his seminar on communication technology has provided long range direction for future work. Mark Ackerman helped greatly in the early stages of forming this work to choose a specific topic out of what felt like a sea of options. His perspective on research helped me think of new methods to apply to this topic. Marc Smith has been a great resource for thinking about large-scale conversation, both in his work on the subject and in the guidance he provides on the research. A five minute walk with Marc is worth an hour of talking with many other people.

This work would not have been possible without the help and cooperation of the administrators of Slashdot. In particular, Rob “CmdrTaco” Malda has not only been generous as a provider of data, but is a natural scientist who helped me interpret many of the results from that data. Jamie “Jamie” McCarthy has also been willing to provide data, and help provide context to the work presented here. Jamie has been especially invaluable in thinking about design consequences of this work. Jon “CowboyNeal” Pater had the unenviable task of providing much of the data for this dissertation, often with me pressuring and cajoling from over his shoulder. His patience and willingness have been greatly appreciated. Nate Oostendorp and KurtDeMaagd also deserve special thanks for their help in creating programs that helped analyze data, teaching me how to clean data, and their general good nature in doing so.

I very much appreciate my colleague from the School of Information, Erik Johnston, who was a great help in thinking about and producing this work. He has a great career ahead of him. Thanks also to my dissertation group at the School of Information: Kelly Garrett, Cal Lee, Matt Bietz, Denise Anthony and Kate Williams. Their support and advice helped me avoid many potential pitfalls in the process, and I look forward to a long career of collaboration with them. Thanks to the School of Information, the Alliance for Community Technology and the National Science Foundation for funding and other support through the dissertation process.

Finally, thanks to my wife, Mary Beth. It’s hard to describe how tough the dissertation process must be for a spouse, and how essential their support is to its successful completion. Mary Beth has been my constant source of comfort, strength, perspective, and resolve through conducting this project, as well as in life in general.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

The growth of the Web has allowed for increased numbers and variety of participants to interact online. Many-to-many interactions enabled by large, heterogeneous, public online forums potentially create relationships that have value to participants. Online interactions may be used to support traditionally offline activities, as is found in work on software to support group-based work activities (Ackerman and McDonald 1996; Olson and Olson 2001), but also may support novel types of interactions (Hollan and Stornetta 1992; Resnick 2001).

However, as affordances of large-scale online conversations create new forms of benefit, they also create new problems for participants. Information overload occurs when the number of messages a participant receives exceeds their capacity to consume those messages. Information overload burdens not only the individual, but can also be unhealthy for the larger online social system as well (Butler 2001; Jones, Ravid et al. 2004). The anonymous nature of the online environment can also lead to anti-social behavior (Sproull and Kiesler 1991), which has had harmful consequences in earlier forms of online, many-to-many interactions (Curtis 1992; Turkle 1995; Rheingold 2000).

Many methods have been used to manage online conversations, though these methods frequently do not scale to large-scale, many-to-many environments. One method that has not been addressed in depth is the use of technologically-enabled rating systems to provide feedback about the content and participants of online conversations. These rating systems, commonly referred to as recommender and reputation systems, are widely used on the Internet for sharing experiences with books[1], consumer electronics[2], and participants in online auctions[3].

While users in small online conversation forums might also benefit from the explicit rating of content, scale remains a factor in this analysis. First, massive participation is often necessary to fuel sufficient ratings. Second, readers in online forums likely use many cues to determine what to pay attention to. Whether a comment includes links or bullet points, or whether it is long or short, might be cues that users employ to determine quality in an online conversation. In small forums, it is possible to consider these cues, whereas in large forums the cues may become overwhelming and explicit ratings provide clarity.

The central, organizing thesis of the dissertation is that rating systems may be usefully applied to large-scale online conversation systems, but that details of the technical features and patterns of use will affect their usefulness. To address this theme, the dissertation examines three aspects of ratings use and provision on Slashdot.org[4], a popular online news and discussion site with a large membership.

This dissertation follows a form alternatively described as the "article" or "sandwich" method. In this format, the three separate projects are individual inquiries, connected through the common of theme of ratings use in online communities. The

---

[1] http://www.amazon.com
[2] http://www.epinions.com
[3] http://ebay.com
[4] http://slashdot.org

content of the dissertation, consequently, isn't intended to address a single thesis, but rather the larger topic of how ratings may be applied to an online discussion: how they can be used to structure the reading experience; how they affect writers, and how the ratings themselves may be procured.

The first project uses server log data to examine how ratings affect readers of comments on Slashdot. Rating systems have to meet many conditions in order to operate effectively, such as sufficient number of raters, and agreement on the criteria of ratings (Terveen and Hill 2002; Dellarocas 2003). Analyzing Slashdot's system may show how rating systems applied to online discussions meet known requirements for recommender and reputation systems.

The second project focuses on how new users are affected by feedback from ratings of their initial participation. New users entering a persistent online discussion forum benefit by learning the standards of participation for the site. Combining server information and survey data, this project displays the pattern of new user participation on the site, and the role of feedback from ratings, previous experience and observation in shaping that participation.

The third project uses server and survey information to reveal how Slashdot users employ feedback provided through the rating system to customize how comments are displayed. This inquiry examines in what ways users employ ratings, and makes recommendations for new types of view settings based on comment ratings.

Dissertation Outline

In the context of the use of ratings to provide structure in persistent online conversations, this dissertation uses three studies that contribute to our understanding of how users in an online community use and provide ratings. In the remainder of Chapter 1, I describe Slashdot to provide context for later findings. Chapter 2 reviews the history and salient features of previous online conversation systems, previous mechanisms for structuring online many-to-many systems, and the relevant research in the theory and use of recommender systems in other contexts. In this review, I claim that other methods that have been used to manage online discussion systems have not been as scalable as the use of rating systems on Slashdot. However, online recommender systems depend on sufficient and fairly applied ratings to be effective, which means that ratings on Slashdot should also be sufficiently and fairly applied to be an appropriate tool for use in online discussion systems. Also, ratings on Slashdot should be useful to readers of content, and provide feedback to new users who are trying to learn the standards for participation on the site. Chapters 3, 4 and 5 present studies that address these issues of ratings use and provision on Slashdot.

Chapter 3 shows that most readers are using ratings to change how comments are viewed, but that a small portion of users make interface changes that decrease the impact of the ratings. Although users are employing ratings to change how comments are displayed, there is some amount of friction that prevents them from doing so readily, as evinced by the tendency to make large changes in how comments are displayed rather than incremental changes. By analyzing "lead users", or users who seem more likely to make changes to viewing interfaces, there are some ways that this friction against using ratings to change how comments are viewed may be overcome. Some changes that users make indicate schemas for reader-types, which may be generated from the pattern of changes people make to the labels attached to comment ratings. Additionally, this chapter describes how you can predict whether a change is warranted in the score at which comments are displayed based on the behavior of lead users who have already viewed the set of comments.

Chapter 4 describes how previous experience, observation of the site, and feedback received from comment ratings and replies to comments may affect participation by new users. Previous experience as measured by self-rated online forum expertise, computer expertise and education level had little association with either initial or continued participation on the site. Observation, as measured by number of page views a user makes, was associated with the probability that a user would post a second comment, the score of that second comment, and the overall number of comments the new user made. It was not associated with the final score of the first comment the user posted achieved through the moderation process, or the amount of time that passed between the posting of the first and the second comment. Feedback, as measured by whether a user's first comment was rated up, rated down or ignored, was associated with the amount of time that passed between posting the first and the second comment, and the score of the second comment. It was not associated with the likelihood of posting a second comment, or the total number of comments made.

Chapter 5 describes how although only 28% of comments on Slashdot receive any rating through moderation, the relatively wide range of starting scores for comments allows for a wide dispersion of final comment scores. A conclusion from this finding is that comment rating is used to separate particularly good and bad content from the average. This chapter also describes that moderators generally agree on the disposition of a comment, as measured by variability in multiple ratings of a comment and scores achieved through a meta-moderation process. Ratings that are scored as “unfair” by this meta-moderation process are usually reversed by subsequent moderations. Chapter 5 also shows that comments that occur late in the life of a discussion or below the top thread level are less likely to be rated at all, and less likely to have an “unfair” moderation reversed.

Chapter 6 synthesizes these findings and ties them back to the core argument of ratings use in persistent online conversations. Two general conclusions drawn from the studies are the importance of starting scores that comments receive based on user identity and the role of moderation labels as providing additional information in the ratings process. This leads into a discussion of implications for future ratings of online discussion forums. I finish by discussing the limitations of this set of studies, and suggesting future research designs in this area.

Contributions

The contributions of this dissertation can be divided into overall contributions of the combined work, and specific contributions in each chapter. This section will begin by enumerating the contributions of the overall, descriptive work about Slashdot and its readers. Then, it lists separate contributions from the different projects that comprise the dissertation; these contributions include empirical findings, conceptual schemes for thinking about the use and impacts of ratings, and methodological innovations.

Overall contributions

There are five contributions made by the overall work presented here. First, this is the first examination of how the Slashdot moderation system operates. Although some technical reports have explained the mechanics of the Slashdot system, and the source code is open for inspection, no work has examined the actual use of that system or how different features affect use. Describing Slashdot contributes possible design guidance to online community practitioners interested in repeating Slashdot's success. Describing Slashdot helps researchers in two ways as well. Although two dissertations have studied Slashdot previously (Halavais 2001; Poor 2004), neither provided full descriptions of features or use, as these works were concerned with Slashdot outcomes more than operations. Several recent studies, both in academia and the public press, have compared virtual communities of various sorts to Slashdot (Rheingold 2003; Dave, Wattenberg et al. 2004; Gillmor 2004; Viegas, Wattenberg et al. 2004; Bryant, Forte et al. 2005). Providing details about Slashdot helps those who make comparisons to that community in the future. In addition to helping future researchers of this particular community, the description of Slashdot helps add essential information to a growing body of work on online communities in general. Studies of Usenet (Smith 2002), MUDs (Bruckman 1997), Wikipedia (Bryant, Forte et al. 2005) and blogs (Facca and Lanzi 2005) all add to an understanding of the different social practices inherent in public, online discussions. Adding Slashdot to that mix is important because of the site's large size, public nature, longevity, and unusual use of tools like ratings systems.

Second, this is the first study to describe the Slashdot user population, which is shown in Appendix A as responses to the Slashdot user survey. Before that study, little was actually known about the characteristics of Slashdot users. Again, these findings are of use to those studying Slashdot in the future, and those comparing Slashdot to other sites. The description of user characteristics also adds to the literature on online community and Internet participation in general. While some surveys of general populations address online community use (Horrigan 2001), and a few surveys of specific online communities address use (Preece, Nonnecke et al. 2004), the main method for describing users of online communities in the past has focused on interview data and observation. The survey of Slashdot users is the first to describe the demographics of users who participate in online discussion. While these users may be idiosyncratic to the site, their description adds to future studies of online community users, and will allow for differentiation to be made between users of different types of communities.

The third contribution made by the overall work is the description of a relatively unique type of recommender system. Most recommender systems aggregate numerical ratings from users (Terveen and Hill 2002). When a Slashdot moderator rates a comment, they are moving the score of the comment higher or lower than it currently is. In addition, moderators do not actually assign scores of +1 or -1 to those comments, but rather choose from a list of labels like "Insightful" or "Offtopic" that then affect the score of the comment. This is a variant of recommendation that can't be found in the literature, and as such contributes to the literature through the description of its use and application. Besides extending the literature on online ratings systems, describing the Slashdot recommender variant is of use to designers of online communities interested in applying ratings to their endeavors.

Fourth, this works describes how Slashdot "pre-rates" comments by assigning them starting scores based on the previous history of the user posting them, as described more fully in Chapter 2. The literature on recommender systems highlights a problem with making recommendations when no content has been rated yet (Terveen and Hill 2002; Herlocker, Konstan et al. 2004), referred to as the "cold start" problem (Schein, Popescul et al. 2002). If the content hasn't been rated, then it is hard to make recommendations, especially in the case of collaborative filtering which depends on many ratings to make personalized recommendations. This could be an especially serious problem in an online conversation, where posting happens quickly and the lifetime of a conversation thread is short. Slashdot's use of pre-ratings offers a possible solution to this cold start problem, by offering a spread of scores at the moment of comment creation.

Fifth, this work extends research in large-scale online conversations from Usenet to Web discussions. Previous work on large-scale, online discussion forums have focused on Usenet interaction (Whittaker, Terveen et al. 1998; Smith 1999; Sack 2000; Butler 2001; Jones, Ravid et al. 2002), because it was a unique example of massive participation, and was decentralized for more convenient data collection. Web-based online discussions like Slashdot have slightly different design characteristics than Usenet, which may affect how participation occurs. This work shows that Slashdot discussions have many of the same characteristics of Usenet posting, in terms of thread depth for example. This provides evidence that findings from Usenet research can be extended into Web-based discussions.

Contributions in Chapter 3

There are three contributions to the literature made in Chapter 3. The first contribution of this chapter is to divide users into categories of those who change defaults to take advantage of ratings, those who leave default settings intact, and those who change their settings to suppress the use of scores. Both cognitive psychologists and economists have reflected that people presented with more information than they can process will often make sub-optimal choices for the sake of reducing their cognitive burden. This tendency has been attributed to human limits of cognition (Landauer 1991), and may be associated with the theory of “satisficing” (March and Simon 1958) wherein users make “good enough” decisions to save themselves the cognitive work of deciding between all possible choices. This work provides an example of how users do this in an online discussion system with many options for changing default settings, and associates user characteristics with that behavior.

The second contribution of this study is to identify the concept of personalization clusters based on the changes users have made to the value of moderation labels in the Slashdot system. Content recommender systems have traditionally made a single recommendation to the entire user population. An example is rating a movie as “4 star”, which does not take user motivations into account. Collaborative filtering systems have personalized recommendations by matching a user’s preferences with other users that have stated similar preferences, and some research has created “ontologies” of use in collaborative filtering systems by looking at similarities between ratings (Middleton, Shadbolt et al. 2004). This work extends that to the field, and shows how groups can be imputed from interface settings, as well as from stated preferences. The schema based recommendations proposed in Chapter 3 show that users modify the scores of comments based on the labels they received through the moderation process, and that they change groups of labels. For example, a user who increases the score of comments labeled as

"Insightful" might also choose to decrease the scores of comments labeled as "Troll". By seeing what clusters of modifications are made by users, one can propose an alternative approach by which ratings are used to recommend content to a small set of user types, rather than to each individual user personally or to a global population.

A third contribution of this work is to include user story views as a unit of analysis, providing a novel perspective of user interaction with online content. Other analyses of online community readership have been focused at all the behavior by a user, or at the conversation thread level (Whittaker, Terveen et al. 1998; Smith 2002). Combining users and conversations as a single unit of analysis permits a different granularity of description, which in this case is used to understand in how many of the conversations they read a user decided to make threshold changes. This formulation could be of use to other researchers of online discussions interested in how users behave differently across different units of interaction.

The fourth contribution of this chapter is to the literature on adaptable vs. adaptive interfaces. Adaptable interfaces are those that a user can change to tailor software to their use, where adaptive interfaces anticipate needs of the user. The literature on human-computer interaction has a long debate on the merits of adaptable vs. adaptive computer interfaces (Findlater and McGrenere 2004). Maclean (MacLean, Carter et al. 1990) theorized several types of users and how they would tailor interfaces differently given their use. Empirical work on preference setting has shown that most users rely on system defaults rather than change the user interface (Cypher 1991; Mackay 1991; Page, Johnsgard et al. 1996). Some researchers have used artificial intelligence agents to anticipate user needs in computer interfaces (Fischer and Girgensohn 1990; Fischer and

Reeves 1992), though this approach can sometimes not match user activities. This work adds to that literature by demonstrating the feasibility of having the activities of users more likely to change interfaces act as the adaptive function in dynamically changing interfaces. The behavior of these "lead users" in changing how they view comments on Slashdot is a good predictor of when other people will also make a change. This means that adapting interfaces based on the behavior of other users may be an effective technique in customizing interfaces.

Contribution in Chapter 4

There are three contributions to the literature to be found in Chapter 4. The first contribution of this study was to define the measures of participation for new users in a persistent online forum. By combining measures of comment scores, rate of posting comments, probability of posting a second comment and overall comments I was able to analyze different types of participation by new users on Slashdot. This may be of use to researchers who study other online communities as measures of new user outcomes.

The second contribution to be found in Chapter 4 is to show that observation and feedback both play roles in understanding new user behavior. Previous work on new users in online communities have largely focused on how long new users occupy an online discussion forum (Butler 2001), or conduct qualitative studies of interactions in environments like Multi-User Dungeons (Curtis 1992; Muramatsu and Ackerman 1998; Reid 1999). By examining the role of alternative explanations of learning for new users, this work is able to indicate some possibilities for designers of online communities who need to socialize new users. For example, online community practitioners might be able to create socio-technical tools that encourage new users to observe experienced participants. Designers may also decide to target new users for initial feedback that might shelter them from initial discouragement, or perhaps increase the drop out rate if that is seen as good for the community.

A third contribution of Chapter 4 is to provide online examples of mechanisms articulated by the literature on situated learning. Recent work on Wikipedia found that new users become engaged with the site at least partially by interacting with more experienced users (Bryant, Forte et al. 2005). This is consistent with the perspectives of communities of practice (Wenger 1998) and situated learning (Lave and Wenger 1993) in which novices interact with experts in various degrees until becoming expert themselves. In this chapter, we found that neither being replied to, nor receiving ratings from other users predicted whether the user would return to post a second comment. However, observation of others was an important predictor as measured by the number site visits the new user made. This extends the work on how new users are socialized in new communities by breaking down different types of interaction the new user may have with others.

Contributions in Chapter 5

There are three contributions to the research literature in Chapter 5. The first contribution in the chapter is to validate the theory of herding or information cascades in comment ratings on Slashdot. Herding in this context is an incorrect consensus reached because moderators are influenced by previous moderations either to remain silent or to contribute another moderation in the same direction (Banerjee 1992). In Slashdot, we find a decreased chance for reversing this herding process when comments are later in the life of a thread, deeper in the thread structure, or start with a lower score. Overall, only half of initial moderations that are rated as “unfair” by the meta-moderation process are reversed by subsequent moderation. This confirms that the theory of herding applies to everyday practice. It also adds an example of the theoretical problem with ratings that there is a benefit for producers of ratings to wait until others have done the work before evaluating the content (Avery, Resnick et al. 1999). Both researchers interested in herding behavior, and practitioners worried about the effects of initial ratings on scoring outcomes may be interested in how the disposition of comments in Slashdot affects herding.

A second, methodological contribution in Chapter 5 is the concept of the “half-conversation life”, or the elapsed time in which half the comments are posted. This is a potentially useful measure of the rate of activity within an online discussion, particularly in online systems that may be asynchronous, but have rapid posting activity early in the life of the conversation.

A third contribution is the confirmation of theory on rating systems that says that people using ratings may not trust the rating system if evaluations are not equally applied to all content (Dellarocas 2003). We find in Chapter 5 that comments late in the life of a conversation, or deep within threads do not receive the adequate attention. This may relate to the underprovision problem (Avery, Resnick et al. 1999), in which the cost of evaluating and assigning scores is exacerbated by costs imposed by the interface. For example, having to scroll to the bottom of a string of comments, or click on a thread to access comments deeper in the structure, add cost to the evaluation that unbalances the equal provision of ratings.

Slashdot

This section describes Slashdot in detail to provide context to the findings presented in subsequent chapters. Slashdot is a news and commentary site dedicated to technology issues, especially those focused on open source software. It attracts about a over 600,000 unique visitors a day, as measured by unique IP addresses requesting pages from the site. Paid editors select about two dozen news stories each day to appear on the site, providing a one paragraph summary for each and a link to an external site where the story originated. Each story becomes the topic for a threaded discussion among the site's users. Most of the commentary occurs in the first few hours after a story is posted, in part because the story loses its prominence on the front page of the site as other stories are posted. Around 75,000 comments are posted each day to the site. This section will provide details on the operation of Slashdot, and context for the chapters to follow.

Slashdot History

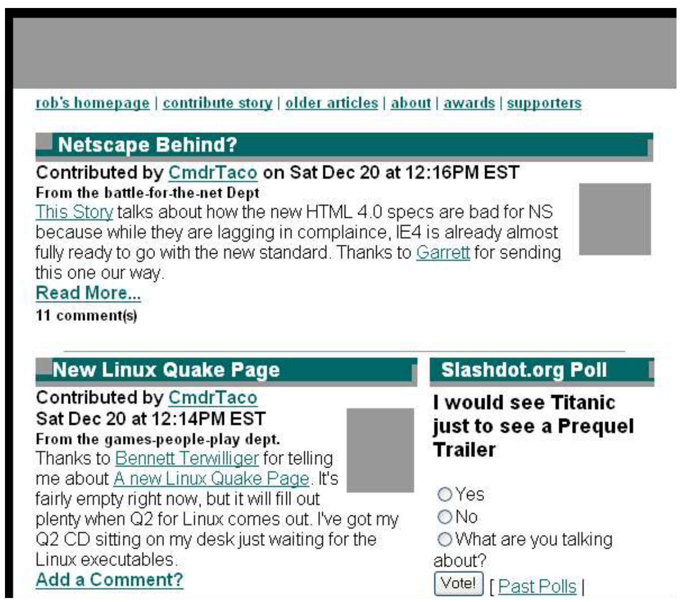
Rob "CmdrTaco" Malda was a Hope College student in 1997 when he posted static web pages to his personal computing space, the content of which was typically related to stories he had found around the Web and found to be interesting. Calling the site "Chips & Dips", he would get email from friends about things they had found as well. The site grew in popularity, and in October 1997 it was moved from college servers to a dedicated server, registered as "Slashdot.org".

> *"Being the cocky young lad that I was, I decided that I would register the most unpronounceable name I could think of. I didn't know exactly what I would be putting on my site, but I knew that whatever it was, it was going to be a creation purely for the age of the Internet."- Rob Malda reported in* (chromatic, Aker et al. 2002).

By April of 1998, Malda had enlisted the help of friends to create a publishing interface for the Web, using mod_perl and MySQL as key elements of the system. The basic design of the site was to post links to news or information somewhere else on the Web, with a small description of what could be found at that location. Visitors to the site could post comments about the item in a threaded discussion format using database driven forms provided by the site.

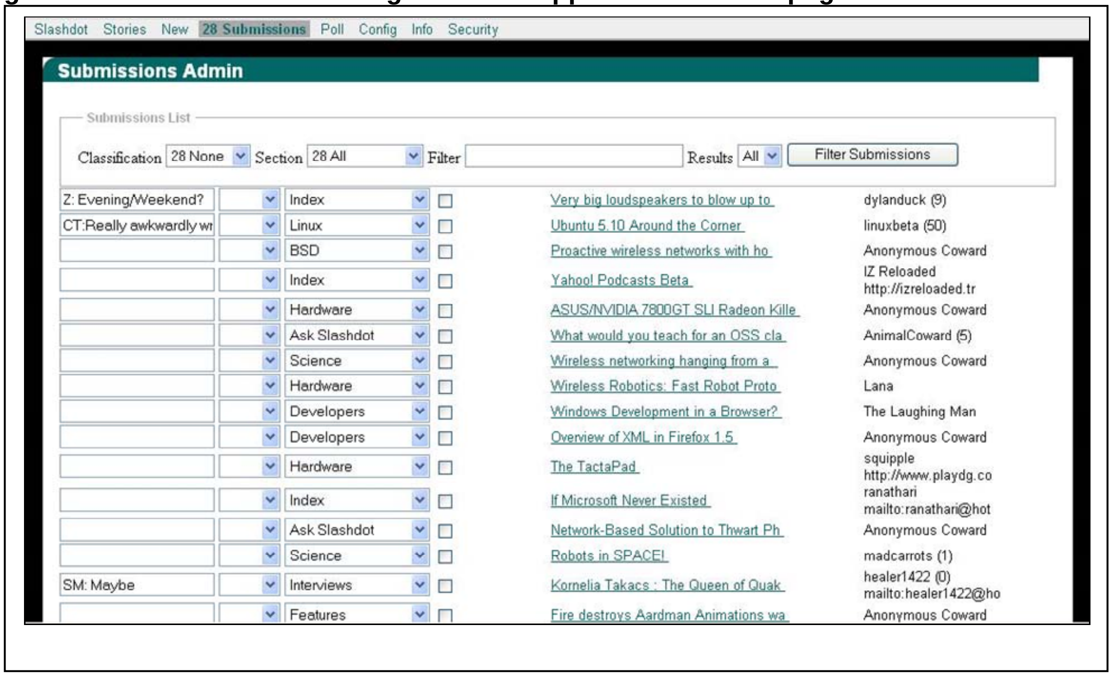
> *"Building on Usenet, IRC, and even MUDs, it (Slash) creates a reasonable system in which the knowledgable and the foolish can mingle, swap places, and somehow produce something that is worth reading. Frankly, it scares me that it works as well as it does."* (chromatic, Aker et al. 2002)

Figure 1.1: Slashdot circa 1997



As the user base grew, instances of misbehavior and information overload also increased. In order to control these effects, Malda and his colleagues programmed a form of moderation in which comments received scores assigned by themselves that would indicate the relative value of those comments. As the number of comments posted to the site increased, they extended the ability to moderate comments in this manner to 25 trusted users, and then to 400. Eventually, they made all registered users eligible for moderation privileges.

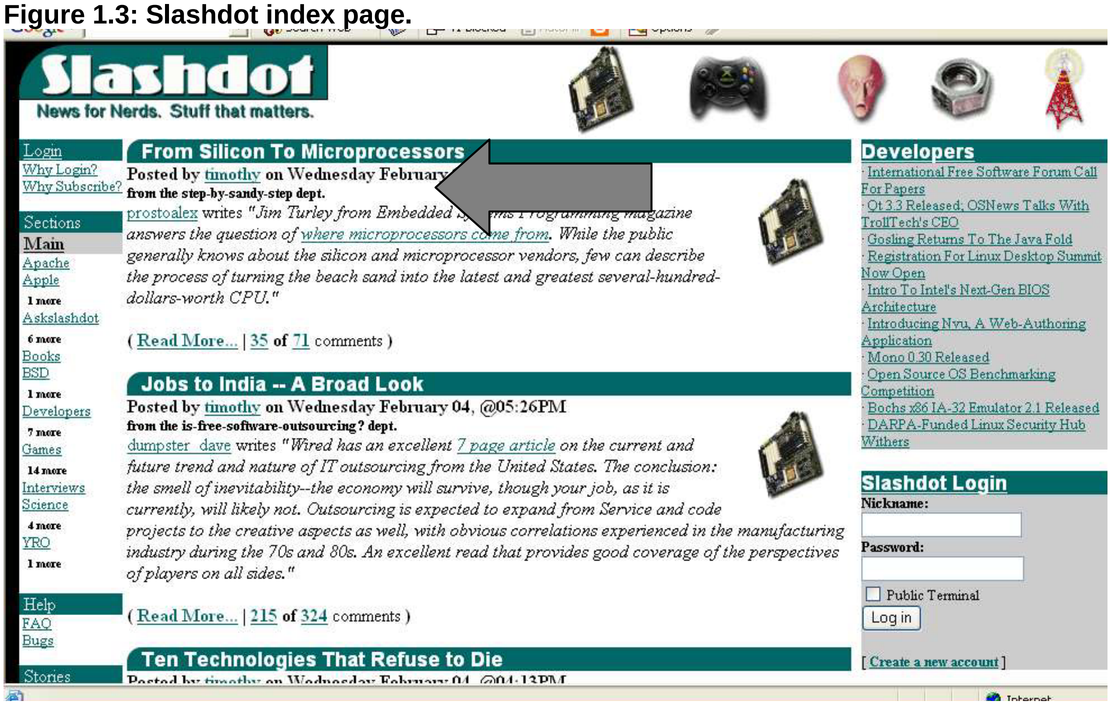
Figure 1.2: Interface for choosing stories to appear on the front page.



Slashdot operations

Stories are submitted to Slashdot by users, who find them from other news sources on the Web. Between 300 and 400 stories are submitted per day to the site, though many of them point to duplicate sources. Editors, paid employees of Slashdot, review the submissions, checking the originating link and reading the article referenced. Figure 1.2 shows the interface for selecting stories to appear on the site. From this list of stories, editors choose around two dozen to appear on the site over the course of the day, depending on the news cycle. Not all Slashdot content points to other sources on the Web. Original content includes user-provided book reviews, interviews in which questions are generated by Slashdot users, articles or reports submitted by editors and users, and the “Ask Slashdot” section in which a user presents a question to the readership, which can answer in the discussion forum.

Figure 1.3: Slashdot index page.



Stories appear in chronological sequence, with a summary written by the submitting user, a timestamp, a link to the original content, and the name of the editor who posted the story. Figure 1.3 shows the index page for Slashdot, displaying how stories are arranged on the site. Not all Slashdot stories appear on the main page, called the “index” page. Slashdot has separate sections for topics like “Games” and “Books” where additional stories not deemed of general interest are posted. All stories that appear on the index page also appear in the topic pages, but not vice versa.

The gray arrow in Figure 1.3 shows the main link to accessing the user comments area. By selecting “Read More”, the user is taken to a string of comments posted by other users. Entering the user comments area allows the user to both read and post comments to the discussion. Users may reply either to the story, which creates a top-level comment, or to one comment, in which case their posts appear deeper within the threaded structure. Figure 1.4 highlights the user interface for adding comments to a thread and the resulting structure. Both registered and anonymous users may post comments to the site. A story may be commented on for up to two weeks after it has been posted, but as shown in Chapter 5, most commenting happens during the first few hours that a story has been posted. After the story moves from the index page into the archive, commenting activity drops precipitously.

Figure 1.4: Interface for commenting on Slashdot.

After a comment has been posted, it can be rated by a "moderator". Slashdot users achieve moderator eligibility by having a positive reputation, which results from their participation on the site. In a recent week in October 2005, 67,000 registered users were eligible to become moderators and about 5,000 per day were selected. When selected, a user is given five moderation points, to be used within three days. Moderators are chosen by an automated script, which periodically assigns moderator privileges based on the number of comments in the system. Users are notified of their moderator status through a note that appears on the top of the index page, as well as by the addition of a UI element that appears in comment pages. Figure 1.5 shows the interface through which users are notified of their moderator status.

Figure 1.5: Notification of moderator status

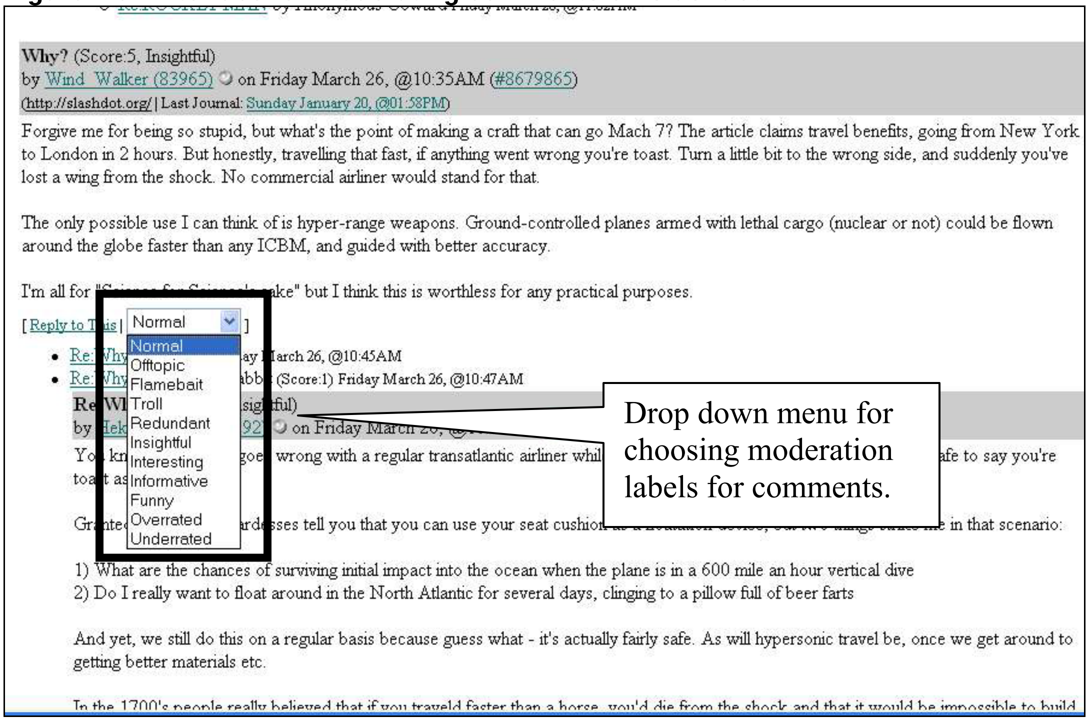
A moderator rates a comment by assigning a label to it, rather than a numerical value. Users choose from a list of descriptors for the comments, such as “Offtopic”, “Troll”, “Insightful”, “Funny”, or “Overrated”, each corresponding to a -1 or +1 adjustment in the score of the comment. Unlike most rating systems that employ an absolute scale (e.g. 1-5 stars), Slashdot moderation is relative to a comment’s score. Moderators are intended to assign rating labels to comments that they think deserve a higher or lower score. Table 1.1 lists the labels that can be used in the moderation process, and the value they assign to comments.

Table 1.1: Moderation labels and associated values

Label	Value
Offtopic	-1
Flamebait	-1
Troll	-1
Redundant	-1
Insightful	+1
Interesting	+1
Informative	+1
Funny	+1
Overrated	-1
Underrated	+1

Figure 1.6 shows the interface and label options as they appear to moderators in a comment stream. Comments start at a score ranging from -1 to +2, depending on the user posting the comment. Comments made by registered users with negative reputation in the system start at -1, those by anonymous users start at 0, those by registered users with normal reputation start at 1 and those by registered users with high reputation start at +2. Final scores for comments can range from –1 to +5, depending on starting scores and modifications made by the moderators.

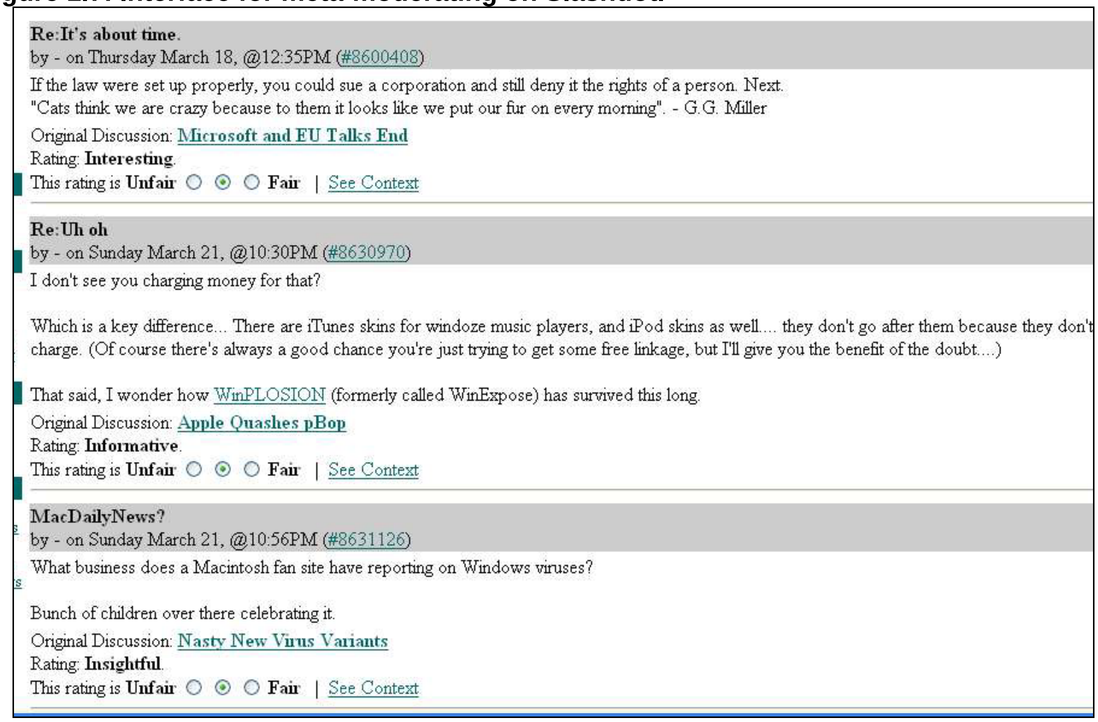
Figure 1.6: Interface for moderating comments on Slashdot.



Besides affecting the score of the comment being rated, moderation also affects the system-tracked reputation of the user who posted the comment. Each user who creates a profile on the site has a reputation score, called a "karma" score by the Slashdot moderation system. These scores range from -50 to +50, though users themselves only see labels like "Positive" or "High" when viewing their karma. Labels were introduced after the creation of the karma system to prevent the gaming of reputation that was made easier by access to the numerical score. Besides the gain or loss of karma points through the moderation process, users may gain karma through reading stories and having stories they've submitted posted to the site. A person's karma score affects the starting score of comments they write, and how often the user is eligible to moderate others' comments.

To "*remove bad moderators from the M1 (moderator) eligibility pool and reward good moderators with more delicious mod points*" (Malda 2003), Slashdot developed a meta-moderation system. Meta-moderators are presented with a set of moderations that they then rate as either “fair” or “unfair” or neutral. For each moderation, the meta-moderator sees the original comment and the reason assigned by the moderator (“Troll”, “Funny”, etc.), and the meta-moderator can click to see the context of comments surrounding the one that was moderated. The final score of the comment is not shown, nor are the other ratings of that comment, just the individual moderation. Consequently, moderations that use the labels ‘Underrated” or “Overrated” are not eligible for meta-moderation. Every rating, other than those labeled as “Underrated” or “Overrated” that is made on Slashdot receives five ratings through meta-moderation. A wider portion of the user base is eligible to participate in meta-moderation than is eligible for moderator status, with the only requirement being that the user not have a negative reputation score. Thus, more meta-moderation occurs than moderation. Figure 1.7 shows the interface for meta-moderating comments. Each user chosen to meta-moderate receives ten moderations to evaluate. Meta-moderation is used by the editorial staff to identify moderators who are consistently being rated as “unfair”, and may affect how often or whether the user seen as unfair gets to moderate in the future.

Figure 1.7: Interface for meta-moderating on Slashdot.



In abstract, Slashdot's comment and moderation system resembles a tree structure, with one story receiving many comments, one comment receiving (potentially) many moderations, and one rating assigned by moderation receiving many meta-moderations.

Figure 1.8: An abstracted view of contributions users make to Slashdot.

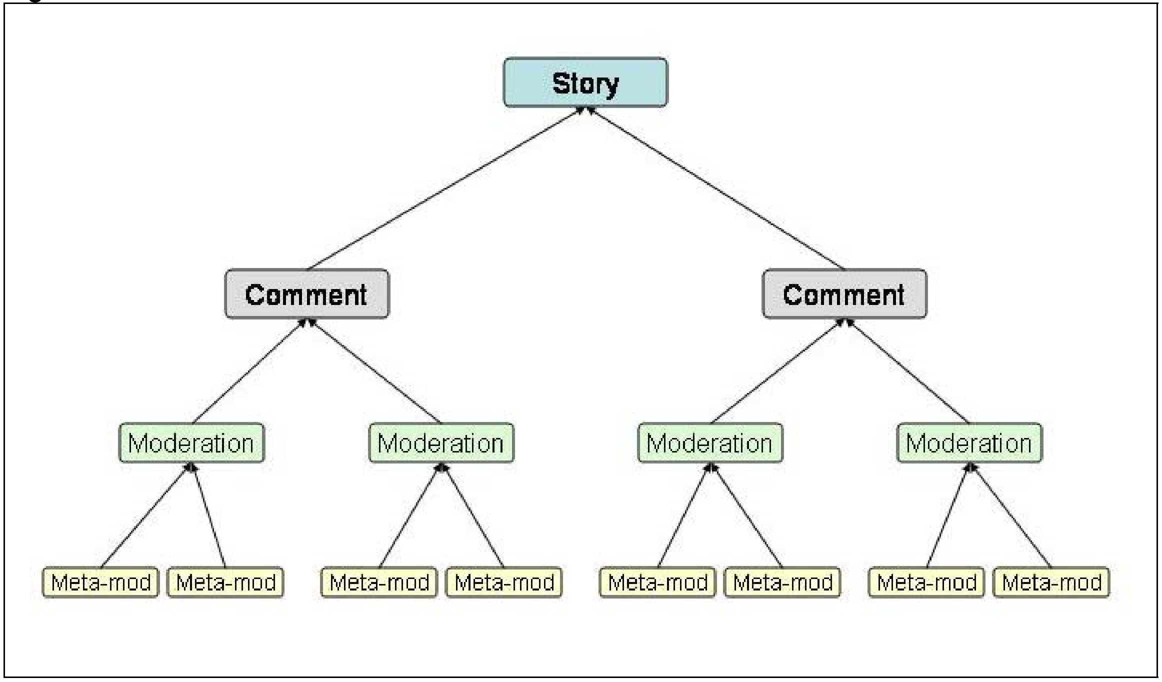


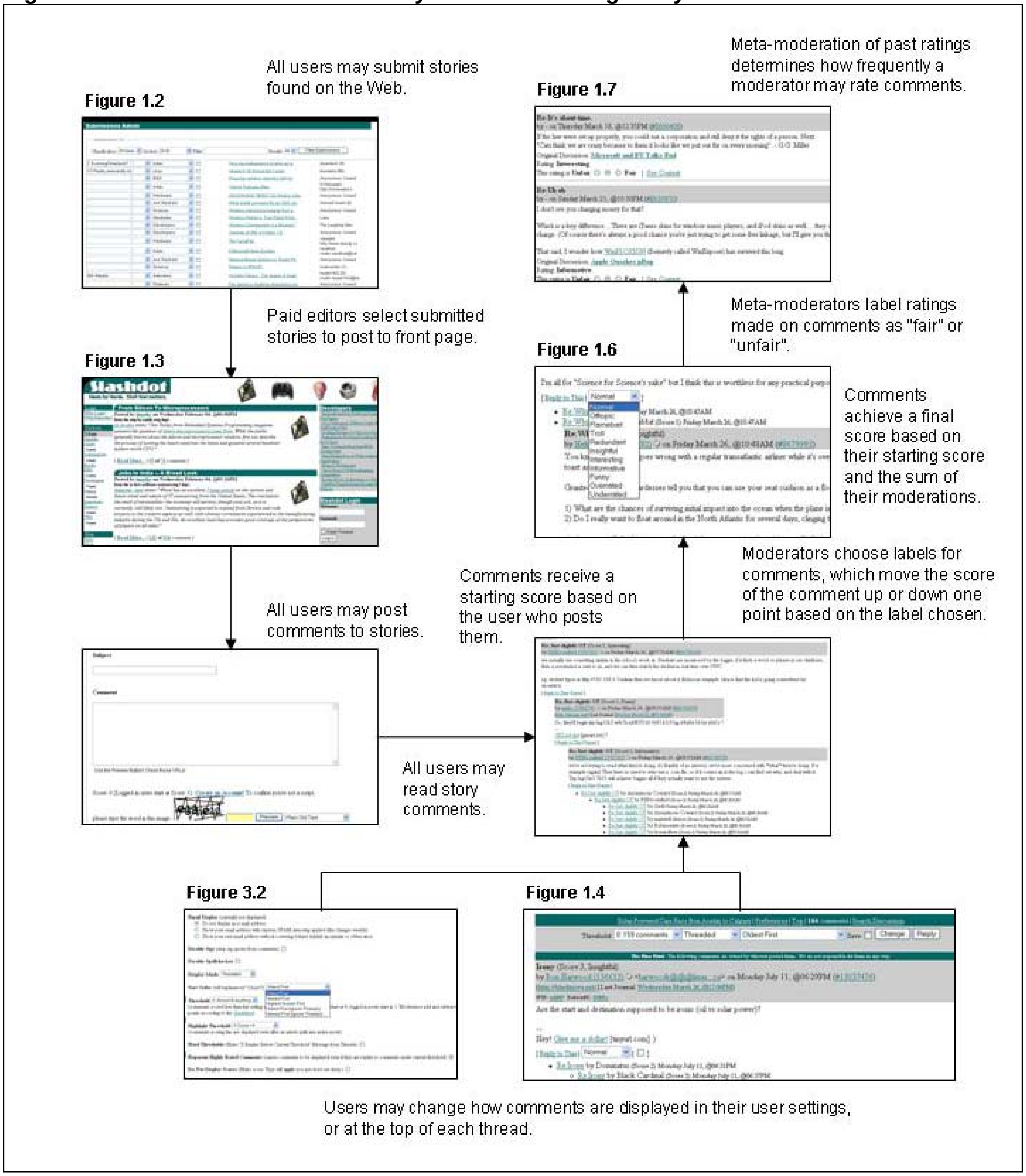
Table 1.2 summarizes the different types of users that may be mentioned, describing their role on the site and eligibility requirements. There are eight active editors currently. There are over 900,000 registered users, though only around 80,000 are active on any given day. There are around 575,000 anonymous users who visit the site on average per day. The numbers of both moderators and meta-moderators are dependent on the number of comments and ratings that occur on the site.

Table 1.2: Types of Slashdot user.

User type	Description
Editors	Paid Slashdot staff responsible for selecting stories to appear on the site. Also have unlimited moderator privileges. Currently 8 editors work on the site.
Registered users	Slashdot users who have created accounts on the site. These users may read, post comments, create journals, and are eligible for moderation and meta-moderation. There are currently over 900,000 registered users.
Anonymous users	Slashdot users who do not create accounts. These users may only read and post comments. Slashdot terms these users "Anonymous Cowards".
Moderators	Registered users who are selected to rate comments. Moderators are chosen from registered users who have positive reputation, have visited the site in the past week, and are in the 92.5% of oldest user accounts.
Meta-moderators	Registered users who are selected to rate moderations. Meta-moderators are chosen from registered users who do not have negative reputation, and have visited the site in the past week.

Figure 1.9 shows the overall lifecycle of activity on Slashdot. Users may affect their settings in their personal profiles at any time, but most activities are sequential, with story submissions leading to story posts, then to comments, moderation of comments and then meta-moderation of the moderations.

Figure 1.9: Overview of Slashdot story and commenting lifecycle.



Slashdot conversation characteristics

Part of Slashdot's editorial policy is that no posts are deleted from the database. *"We believe that discussions in Slashdot are like discussions in real life- you can't change what you say, you only can attempt to clarify by saying more. In other words, you* *can't delete a comment that you've posted, you only can post a reply to yourself and attempt to clarify what you've said."(Malda 2003)* This exacerbates the information overload problem, as Slashdot received high traffic and is committed to retaining all comments.

Another policy of the site creators is that anonymous posting be allowed within the system. *"We think the ability to post anonymously is important. Sometimes people have important information they want to post, but are afraid to do it if they can be linked to it. Anonymous Coward* (ed. Slashdot term for anonymous users) *posting will continue to exist for the foreseeable future." (Malda 2003)*

These features are salient because it creates an online conversation system with low barriers to enter the conversation, high message load, few constraints on user misbehavior, and the ability to change pseudonyms frequently. As will be shown in the background section, these features would usually be associated with breakdowns in the online conversation. Slashdot's rating system has allowed the site to preserve these features while still providing structure for participants.

Chapter 2
Theoretical Motivation and Background

Chapters 3, 4 and 5 have separate sections that motivate the individual inquiries. This section, rather than replicating statements made in the individual chapters, addresses general themes related to the use of rating systems in online conversations. This narrative is intended to show that there is a history of large-scale, many-to-many interactions with outcomes salient to the current state of online conversation, that previous methods of structuring these systems have not allowed a scale that should be encouraged, and that recommender and reputation systems may be a useful tool for these systems if such systems can overcome known barriers like the need for sufficient rating and the tendency of raters to "herd" after initial ratings.

Type of online discussion

This dissertation considers a particular type of online interaction, which is large-scale, public, text-based, many-to-many and persistent. This section provides background information about the features of these systems and how these types of online interactions have been described in that past.

This work is concerned with online communication that takes place simultaneously among many participants, in forums where messages are relatively persistent, where participation is publicly visible at least to other members of the forum, and where the main method of interaction is typically through textual messages. Often, these types of online conversation systems involve anonymous or pseudonymous participation, large numbers of participants, and are topic centered, though these latter characteristics are not necessary conditions. Although online discussion participants may choose to meet offline, the environment itself is designed as a wholly online space, not support for pre-existing offline activities. The relative permanence of messages and the many-to-many focus of online discussions differentiate them from instant messaging, Internet Relay Chat (IRC), one-to-one email, intranet messaging systems, and typical blogs that are used as a broadcast mechanism. Examples of online discussion systems to which this research may apply include email distribution lists, Usenet newsgroups, and Web-based bulletin boards. Often, online discussions are elements of more complex online interactions, like scientific collaborations, multi-player online games, or online content creation sites like Wikipedia[5] or the Debian[6] Project.

The most common shorthand term for online conversation systems like this is "virtual community", with alternatives being "online community" and "digital community". The term "virtual community" is traced to Rheingold's 1993 description of interactions on a popular Bulletin Board System (BBS) named the WELL (Rheingold 2000)[7]. The use of the "community" has proven contentious, with a great deal of popular attention debunking (Galston 1999) and defending (Powazek 2002) whether online community is a real form of community or not. Matei (Matei 2005) traces how the term "community" in the phrase was an intentional construct of the participants of the WELL, tracing discourse in that space on the subject. The WELL, a derivation of a liberal newsletter called the Whole Earth 'Lectronic Catalog, wanted to encourage the new medium to be seen as a communitarian enterprise.

[5] http://en.wikipedia.org/wiki/Main_Page
[6] http://www.debian.org/
[7] The WELL has become a web community owned by Salon (http://www.well.com/).

Rheingold (Rheingold 2000) updated his 1993 book in 2000, adding a chapter where he claimed he would have avoided the word "community" had he known the controversy involved. The term he preferred in the updated version of his book was "online social network". Online social networks are most fully covered by Barry Wellman (Wellman 1997; Wellman and Gulia 1999), who de-emphasizes the emotional-affective dimensions of online interactions evoked by the term "community" in favor of behavioral and relational measures common in social network analysis. Others (Wasko and Faraj 2000) have applied work on "communities of practice" (Wenger 1998) to the online medium, focusing on the information and expertise exchange that can happen in online interactions rather than affective bonds.

Other terms used by researchers avoiding the virtual community debate include "flash forums" (Dave, Wattenberg et al. 2004) and "virtual publics" (Jones and Rafaeli 1999), emphasizing the short lifespan of many online conversations. According to Jones et al (Jones, Ravid et al. 2004), virtual publics are "symbolically delineated, computer-mediated spaces… that enable a potentially wide range of individuals to attend and contribute to a shared set of computer-mediated interpersonal interactions." This term highlights the affordances of being public, namely wide participation, low barriers to entry and differing levels of anonymity.

Although these terms separately evoke important characteristics of persistent online interactions, none sum them up adequately to abandon the term "virtual community" or take precedence over precise, if awkward, descriptions of specific online social systems.

Many types of online technologies have been termed "virtual communities" in the past. Large email distribution lists, Usenet newsgroups, Multi-User Dungeons (MUDs) and Bulletin Board Systems (BBSs) have all been described as virtual communities at one point or another. Each of these systems allow for many-to-many, technologically-mediated interactions between participants, so share many characteristics of interest here. The following section briefly describes the history of online communication, then describes these virtual communities in more depth. The designers of Slashdot had participated in these different systems before creating the site, and those experiences shaped how they developed the rating system on Slashdot (chromatic, Aker et al. 2002). Also, the next section will highlight how older online discussion systems had illustrated the problems of scale in online conversations, the need to socialize newcomers to persistent online systems, the types of user misbehavior that could occur online and the resultant need for governance mechanisms, and the mechanisms that were developed in these cases.

History of online discussion

From its inception, the Internet has been a communication device. The value of ARPANET was originally seen to be in resource sharing, i.e. the ability of remote researchers to use software hosted on the large mainframe machines at the time. Abbate (Abbate 2000) argues that ARPANET would be a minor footnote had its use not switched from resource sharing to communications via email. By the middle of the 1970's, a significant population of users was communicating online, in various contexts. As more nodes were attached to ARPANET, email communication continued to grow.

Projects related to ARPANET were intended to have a research focus, though offtopic endeavors, like a science fiction discuss list, were tolerated. As the ARPANET developed and more sites were added to the network, the switch from research to communication increased (Ritchie and Thompson 1974; Hafner and Lyon 1996). During this period, email distribution lists and conferencing systems were developed, foreshadowing information overload problems that would recur in later online conversation systems.

By the late 1970's, users dissatisfied with the centralized control and expense of the government sponsored ARAPANET were creating alternative online discussion systems, three of which included features salient to current online conversations: Usenet, MUDs and BBS enabled virtual communities. These systems experienced the benefits and deficits of large-scale, public, many-to-many interactions, and provide many examples of mechanisms used to structure user behavior and decrease information overload. The following section describes each system and its salient features.

Usenet

While email discussion lists were early developments in the progress of ARPANET and the later Internet, it was the development of Usenet that extended online conversation to large-scale audiences. Developed in 1979 by Duke University students Tom Truscott and Jim Ellis and University of North Carolina student Steve Bellovin, Usenet was initially used to exchange information about the Unix operating system between Duke and the University of North Carolina over telephone lines (Abbate 2000). Handing out written descriptions of the project as well as software tapes at conferences, Truscott and Ellis freely distributed the program and encouraged fellow Unix enthusiasts to sign on to the endeavor (Pfaffenberger 2002). By 1984 nearly 1,000 sites were participating in the network (Quarterman and Hoskins 1986). Usenet allowed connections between sites with fewer resources than had been typical in the ARPANET system. By lowering the barriers to entry for sites, it allowed for more organizations to be part of the computer network and broadened the base of computer network users. This meant that non-programmers who had alternative agendas than sharing computer code could communicate directly with each other.

The content of Usenet is organized into different "newsgroups" in which users "post" content. A message may be cross-posted to many newsgroups and contain meta-information that provide context in terms of the poster and related messages (Smith and Fiore 1991; Whittaker, Terveen et al. 1998; Fiore, Tiernan et al. 2002).

Usenet newsgroup participation offers prototypical examples of the types of problems that can occur in large-scale, persistent, online interactions. Pfaffenberger (Pfaffenberger 2002) describes historical examples of socialization problems that occur in current settings. Trolling, spamming, flaming, flooding and other such colorful terms to describe user misbehavior originated in Usenet. Newsgroups reacted to these misbehaviors in many different ways. Occasionally, a newsgroup would increase centralized control, and introduce moderation to the interactions of that group. Sometimes newsgroup members would exit a group with too many misbehaving users and form new groups to avoid them. Some newsgroups became inactive since the misbehavior was too prevalent to realize the intended activity of the group.

Large-scale participation, which can lead to information overload, has also been a common occurrence on Usenet newsgroups. In 2000, over 151 million messages were logged by the Netscan project as having been posted to Usenet (Smith 2002). A study of 500 newsgroups found that the groups received an average of 24 messages per day (Whittaker, Terveen et al. 1998). Other studies report similar instances of high traffic in Usenet newsgroups, linking the high rate of posting with information overload (Butler 2001; Jones, Ravid et al. 2004).

Management of both user misbehavior and information overload in newsgroups has been difficult, as Usenet administration is distributed with no central authority to coordinate efforts (Smith 1999). When conflict occurs in Usenet newsgroups, or the perception of information overload is high, a common tactic is to splinter the community into more specialized groups (Fisher and Lueg 2002). Although cross-posting of comments across groups with similar interest is common (Whittaker, Terveen et al. 1998), few newsgroups individually have massive number of participants.

Slashdot adopted the ideas of topic based conversation, threaded conversation structures, unmediated user interactions and large-scale interactions from Usenet. The site did not adopt the decentralized structure of Usenet, which may be an effect of the structure of Web sites compared to Usenet protocols. They also did not splinter the site when populations grew large, or groups decided they had opposing goals. This may again result from how Websites operate as opposed to newsgroups. Finally, Slashdot maintained more centralized authority than newsgroups typically used. Although server administrators had broad powers like choosing which newsgroups to carry from Usenet, within the groups themselves power was equally distributed among members. Slashdot created differences in power levels by incorporating the right to moderate comments, and the administrators themselves kept broad abilities to regulate content and users at multiple levels.

Multi-User Dungeons (MUDs)

Multi-User Dungeons (MUDs) allowed for many-to-many interactions initially based on role-playing games. MUDs provide early examples of user misbehavior and the resulting mechanisms developed to discourage such activities. Many of the socio-technical design mechanisms used in Slashdot were derived from the experiences designers of that site had with MUDs.

Multi-User Dungeons (MUDs) were first developed in the late 1970's by two students named Roy Trubshaw and Richard Bartle. Bartle had been involved with the then newly developed paper-based role playing game "Dungeons and Dragons" and wanted to extend the social interaction of that game to online gaming (King and Borland 2003). Being based in a table-top role playing game, the first MUD inherited conventions like numerical ratings to track player progress, the use of pseudonyms as acceptable practice, the concept of collaboratively created narratives, and the use of centralized authority as arbiter of conflict ("dungeonmasters" in the offline game and "wizards" or "gods" in the MUDs).

The first MUD and most of those that followed it used textual descriptions of locations, objects and characters to create a sense of place in the online environment. Described as "text-based virtual realities" (Curtis and Nichols 1993), these games allow users to interact with the system and each other by describing their actions, using a set a preset commands to change the game environment. Many MUDs use an adventure game format, where a player assumes a character identity and then navigates the world, killing generated monsters, solving puzzles, collecting objects and sometimes fighting with other characters. Other MUDs have dropped the "swords and sorcery" themes, encouraging social interactions and cooperation instead (Bruckman 1997).

Like Usenet, MUDs experience a variety of social dilemmas, especially with the existence of anonymous and pseudonymous interactions. As Curtis (Curtis 1992) describes in the social interactions on LambdaMOO *"This protective anonymity also encourages some players to behave irresponsibly, rudely, or even obnoxiously. We have had instances of severe and repeated sexual harassment, crudity, and deliberate offensiveness."* Reid (Reid 1999) observed several fantasy and social MUDs, describing the social disturbances and subsequent reactions in each. She concluded that social control in such spaces is enforced through hierarchical power systems enabled by a combination of technical features and social rules. "Gods" and "Wizards" use access to deletion and editing mechanisms to enforce social rules in MUDs.

MUDs remain relevant examples of how users came to a shared online space and needed to negotiate issues of governance and socialization. Curtis (Curtis 1992) described many of the social dilemmas that rose in early text-based MUDs. Muramatsu and Ackerman (Muramatsu and Ackerman 1998) describe similar patterns in another fantasy role-playing MUD, noting that social structures evolved to support various types of cooperation and conflict. They also found that game administrators played valuable mediation roles in governing the game MUD being studied.

MUDs have most often responded to issues of social control by increasing centralized authority (Reid 1999). This limits the size of the group that can be brought together, as human moderation "gods" in MUDs cannot process large amounts of interaction. For example, current Massively Multiplayer Online Role-Playing Games (MMoRPGs) like World of Warcraft and Everquest offer multiple servers to not only handle the technical difficulties of massive participation, but the social burden of placing many players in the same space. Conflicts between players are still mediated by "gamemasters" who have special access to the system.

Slashdot borrowed several features developed by MUDs when designing their online conversation environment. The use of numerical systems to track user progress, the use of pseudonyms, and the use of multiple authority roles are all part of their current operations. However, Slashdot has not adopted the mechanisms of splintering as a response to large populations, which may be possible because the task of commenting on news stories requires lower effort than the more involved game tasks of MUDs.

Bulletin Board Systems

Bulletin Board Systems (BBSs) are networks in which users may call into a central computer and access a database that includes messages and software. Especially in early instantiations, Usenet and MUDs could be considered specialized forms of BBSs. Unlike the Internet, BBSs are not typically connections of many networks, but rather centrally controlled servers. The first BBSs were developed in the late-1970s, and continued in popularity through the 1990's. Wide adoption of internet protocol standards drew people away from local networking, but many companies, such as AOL and Compuserve, that started out as BBSs have changed into other computer network services (Abbate 2000).

Because BBSs depended on modem dialing into the central computer, the networks remained relatively small and geographically located. However, these limitations seemed to increase the feelings of social connectedness between participants, leading to a sense of community among participants (Rheingold 2000). BBSs demonstrated the ability of online interactions to form connections between participants. Positive effects like access to information and affective support were reported, as well as negative effects like identity deception and harassment (Horn 1998).

The salient features that Slashdot borrowed from BBSs were direct connections between multiple participants, and that some amount of misbehavior was likely to result from such contact. BBSs also provided examples of how users who were not necessarily those in control of the site could fulfill roles that helped structure the interactions. BBSs did not have the scale of Slashdot, were usually hosted by one person or a small group allowing for more centralized control, and were often geographically centered, which are key differences.

Methods of governing online discussions

Online interactions are hampered by the effects of information overload and user misbehavior, which has engendered numerous mechanisms for ameliorating those problems. This section briefly describes the nature and origin of problems associated with Computer-Mediated Communication (CMC), and then proceeds to describe common methods to manage online discussions. The goal of this narrative is to show that most of the mechanisms that have been developed do not adequately ease the effects of massive participation without losing the benefits of such participation.

Effects of computer mediation on communication

The purpose of this section is to briefly describe the characteristics of CMC that require the use of special mechanisms to help participants. CMC allows for large-scale participation, since the technological burden of adding new participants is not high. Adding email addresses to a distribution list, or allowing more members in a newsgroup, will not significantly affect the performance of the technology that supports those media. This has the possible positive effect of including large groups of users and maintaining peripheral awareness of online communities through the capacity for low level participation (Resnick 2001), but also creates the possible negative effect of information overload. Information overload in online communities happens when the quantity of available messages is higher than the ability of a consumer of the site to use those messages (Butler 2001).

CMC also constrains the information we can perceive about fellow discussion participants. This has possible positive effects in that it can reduce the biases often found in face-to-face communication, but has possible negative effects in that anonymous interactions seem to increase user misbehavior (Sproull and Kiesler 1991). The need for richer information channels may depend on how much common ground is necessary in the interaction being accomplished (Olson and Olson 2001).

Mechanisms for structuring online communication

As shown above, many problems with computer-mediated communication stem from either information overload caused by large-scale participation or degraded common ground caused by constrained information channels. Methods of organizing online communities have focused on directly redressing one or both of those problems through the addition of technological tools and/or social mechanisms. This section will describe the methods that have used these techniques for structuring online communities, and then discuss why there are opportunities to explore alternative methods.

Increased media channels

Some systems add media channels back to computer-mediated communication in order to increase the capacity for reaching mutual understanding between participants. There is a long history of systems that add media channels back into CMC environments, especially in the field of computer-supported cooperative work (Olson and Olson 2002). A review of the research in this area is beyond the scope of this narrative, but communication channels that have been replaced in CMC include visual, audio, gesture, co-referentiality, chronomics, proxemics and even the ability to touch objects (Clark 1992).

These systems have been developed to support small group activity, usually structured around work tasks. Adding media channels back to CMC environments necessarily constrains the number of participants that may use the system. Consequently, media-enhanced interactions as they have been developed to this point will not support the large-scale interactions of interest to this work.

Centralized editing and censoring

In some online discussion systems, a single editor (or small group of editors) maintains control over a broad range of “write” permissions. Write permissions may take the form of an ability to add or delete members, and edit or delete content. This person, or group, is sometimes called an editor, or moderator. In MUDs these centralized authority figures are called wizards, gods or game masters. To not confuse this role with the meaning of a Slashdot moderator, I will use the term "editor" to describe these users with strong centralized authority.

Centralized write permission may be used to mitigate information overload by restricting how many new users may enter an online discussion, vetting content before it is distributed to an entire group or in some cases editing messages for length. User misbehavior is often reduced through centralized control by censoring users, vetting content and or editing content for inflammatory material.

Email lists are an example of an online communication mechanism that use centralize authority to structure interactions, and common list software provide documentation on various methods for editing, vetting, or deleting content (L-Soft International 2002). A survey of list moderators found that 32% of list moderators described their main duties as filtering content, as opposed to 14% who responded that their primary duty was to prevent flaming, and 12% who listed conversation facilitation as the major goal of moderation (Collins and Berge 1997). The same survey found that list moderators felt that reducing noise, reducing off-topic conversation and preventing flames were the top three reasons for moderating an email list. Slowing down conversation and increased moderator effort were listed as the top two reasons for not moderating.

Some Usenet newsgroups are moderated as well, with newsgroup administrators editing content, often for grammar and spelling according to one early survey (Morris 1993). The same study found that time limitations caused group moderators to restrict the amount of time spent editing, doing so only when clarity of the message would suffer otherwise.

Multi-User Dungeons (MUDs) also typically centralize authority for editing and censoring content and users. Typically referred to as “gods” or “wizards”, these super-users typically can edit participants, delete characters from a system, or censor the actions of a user (Curtis and Nichols 1993). Reid (Reid 1999) described various methods by which control is exerted in MUDs. Studying both adventure and social MUDs, she found that users with central authority, wizards, used their control over the code of their spaces to resolve conflict. However, she warned that often these interventions were seen as involving favoritism, and often called into question the legitimacy of the authority of the MUD. When centralized authority is viewed as autocratic, it can be harmful to future participation in a space. Smith (Smith 1999) described the development of complex mediation rules in a MUD to resolve conflict between regular users and administration.

There are two characteristics of centralized control of online discussion spaces that point toward exploring other mechanisms for large scale interaction: scalability and protection of minority opinions.

As a group grows in size, the number of messages becomes too large for a single moderator or even group of moderators to constrain. Jones et al (Jones, Ravid et al. 2002) describe online interaction in terms of an S-shaped curve. Initial critical mass is hard to build, but once achieved leads to an explosion of contributions. If the overload caused by the sudden growth becomes too difficult to sort through, conversation breaks down to pre-critical mass levels. Centralized content moderation is unlikely to manage this explosive growth, and hence would lose the benefits that come from mass interaction.

Online discussions may include a variety of viewpoints, some of which may be unpopular with those with centralized authority. One goal of a robust conversation environment is the ability to express minority opinions (Cristiano 1999), which may be difficult to foster if a central editor controls which posts are distributed.

Distributed editing and censoring

Another option for guiding group behavior is to diffuse the authority to edit or censor to a wider group of users. While most online spaces allow for some type of user contribution, granting the ability to affect other users' contributions is rare. These systems address information overload problems by distributing the editorial role over a larger population of participants, and manage problems of user misbehavior by allowing users to police content for malicious changes (Bryant, Forte et al. 2005).

Often, spaces will allow for personal control of views, though the control is distributed among users as a type of customization. For example, killfiles have been used in Usenet to allow users to filter out the contributions of specified newsgroup participants, and "ignore" commands in email and online games have been used for the same purpose.

Distributing editing and censoring privileges to a wide user base is a method used by the site Wikipedia[8]. This site is a user-generated encyclopedia built on the Wiki platform. Wikipedia allows for any user to edit, delete or contribute to an entry on the site. To control for malicious edits, like erasing and entire entry, the site mirrors "last, best copies" and enables other users to replace the broken content. A recent analysis of author conflict on Wikipedia showed that most malicious edits are fixed quickly, though

[8] http://en.wikipedia.org/wiki/Main_Page

sometimes contributors can become locked in a re-editing loop where each undoes the others' actions (Viegas, Wattenberg et al. 2004).

Distributing the authority to edit and censor globally in an online community remains rare, and may only work for certain types of endeavors. Besides Wikipedia and other Wiki projects, there are few examples of truly distributed editing and censoring. Open source software development is sometimes seen as the prototypical example of this type of activity (Raymond 2000), though studies of open source projects show that centralized authority is relatively strong in terms of controlling membership and which content is distributed (Mockus, Field et al. 2002; Sandusky and Gasser 2005).

Wikipedia has characteristics that make it particularly disposed to distributed editing and moderating. Encyclopedia entries are relatively static, with no particular "shelf-life" for access by incoming users. Slashdot stories, on the other hand, have relatively short life spans. Discussion on news may be a quicker overall process than the creation of a knowledge artifact like an encyclopedia. This means that the principles that seem to work in Wiki systems and similar endeavors may not translate to other types of online interaction. This work adds to the literature on distributed governance of online communities by adding the activity of news discussion to what is known about distributed governance.

Recommender and reputation systems

The methods described above offer solutions that may not map to text-only, public, large-scale, relatively ephemeral interactions like those found in online discussion forums. One solution that may be more applicable to such environments is the use of recommender systems to leverage the past experiences of some users to inform the choices of current users. This section describes the literature on recommender systems, including possible hurdles to their implementation in online discussions.

Stated simply, recommender systems offer information about a choice that a consumer has to make between several alternative products or services when the consumer does not have personal experience with those alternatives. To reduce uncertainty about which alternative to choose, the consumer (or user) turns to others who have had previous experience with the range of choices. In short, any time the past is predictive of the future, recommender and reputation systems can help people make informed choices about what to pay attention to or consume and with whom to interact (Resnick and Varian 1997).

Seeking information about a choice from people who made that choice previously is common in the offline environment. Formalized systems like movie ratings, restaurant reviews and buying guides depend on experts to provide information about the relative worth of a product or service. An example is Consumer Reports, a nonprofit organization founded in 1936 that tests a wide range of products on multiple criteria, and then makes recommendations about which to purchase to its members. People seek recommendations from less formal sources like family and friends as well. For example, a person deciding which movie to see might select one based on the fact that a friend saw it and said they enjoyed it, a tendency typically referred to as “word of mouth”.

Online recommender systems use computation to aggregate the past experiences of some users to predict future preferences for other users. Items being recommended can include content, products, services and other users. Reputation systems are tools that recommend other users, rather than some item of content (Resnick, Zeckhauser et al. 2000). Although Slashdot does use a reputation system, none of the work in this dissertation deals specifically with that tool. Consequently, this section will focus recommender systems for content rather than people.

Recommender systems can differ widely on how preferences are computed, and how preferences are collected. The following sections describe these characteristics of technology-enabled recommender systems.

Computing preferences in recommender systems

Recommendations may be specific or general. For example, if I recommend a book on Amazon, I am making a general recommendation to the public at large. If I tell my friend to read a book, then I am making a specific recommendation. As Terveen and Hill (Terveen and Hill 2002) say:

"A recommendation may be directed to specific individuals or 'broadcast' to anyone who's interested. For the person who receives it, a recommendation is a resource that helps in making a choice from the universe of alternatives… A recommendation may be based not just on the recommender's preferences but also on those of the recommendation seeker."

Generalized online recommender systems include a range of systems. Amazon allows users to create lists of books on a certain topic, so that a user searching for a book on a certain topic may find a collection of similar books recommended by a person familiar with them. Apple's iTunes allows users to see the music play lists of others. Epinions[9], a site dedicated to sharing experiences with a wide range of consumer products, has users provide simple ratings of items and write about their experiences on the site. Even though these types of systems may use technological tools to aggregate

[9] http://www.epinions.com/

ratings or collect lists, they do not match recommendations to users through any computational mechanism. Rather, users must seek out like others, or trust that a recommendation is globally applicable.

Alternatively, some recommender systems use mechanisms to match recommendations to users based on their preferences. These systems can match preferences either through content similar to what a user has shown interest in previously, or through users who display similar interests. Tapestry, perhaps the earliest recommender system, was designed so that shared documents within Xerox PARC could be retrieved based on ratings and replies from those who had already read them (Goldberg, Nichols et al. 1992). This project coined the term "collaborative filtering" to describe technologically-enabled recommendation in which recommendations are tailored to specific user preferences. GroupLens (Resnick, Iacovou et al. 1994; Konstan, Miller et al. 1997) uses collaborative filtering to recommend articles in Usenet news. GroupLens operates by asking users to rate a certain number of Usenet articles as useful or not. Those ratings are compared with other users rating the same content using statistical similarity methods. If Person A has a rating on an item similar to Person B, then other items Person B rated might be of interest to Person A. Ringo was a tool that allowed users to rate music with which they had experience, and then used correlation to match users of similar tastes and make further listening recommendations (Shardanand and Maes 1995). Another well-known collaborative filtering system is MovieLens, based off the GroupLens project, a movie recommendation site (Herlocker, Konstan et al. 2004). MovieLens solicits ratings on movies that users have already seen, and uses those ratings to recommend movies that the user did not rate, but that people who had assigned similar scores for the rated movies did.

Eliciting preferences in recommender systems

As mentioned above, generalized recommender systems depend on users seeking out recommendations that they feel match their preferences or are general enough to apply to all information seekers. Collaborative filtering, on the other hand, depends on knowing about the preferences of the recommendation seeker as well as the recommendation provider.

One method of eliciting user preferences is to impute future preferences based on their previous choices, which is sometimes referred to as “content recommendation” (Terveen and Hill 2002). For example, if a user buys a Bruce Springsteen album on a music sale site, the system might log that information and guess that they are interested in other albums by the same artist. This can be expanded by examining other purchases made by users who also bought a Springsteen album, and recommending those items. This is not collaborative filtering, in that users are not matched on similarity, but rather the content is the unit being matched. There are several potential problems with this type of preference mining. Content recommendation is limited in that only near term matches are possible, and a user will not be recommended anything dissimilar to what they have seen in the past (Balabanovic and Shoham 1997). Also, matches assume single motivations for seeking recommendations. For example, one might buy a jazz album online as a present for a friend, but not be interested in future recommendations for jazz albums. Another potential issue is multiple use of single online accounts. An Amazon buyer might share and accounts with family members, and consequently have confounded recommendations based on the purchases of that member.

Another way of soliciting user preferences is to analyze implicit choices they have made and impute decisions from them, sometimes referred to as social data mining. PHOAKS[10] is a system that uses search tools to find URLs in newsgroup postings, and recommend web pages based on a series of calculations about how and where the URL appears in a newsgroup (Hill and Terveen 1996). Google uses a type of recommendation service as well to match search requests. Using its PageRank software and software agents that comb the web, Google captures the way Web pages describe a link for future matching. For example, "University of Michigan", "UM" and "Umich" may all be labels that a person using Google might use while searching for the website for the University of Michigan. PageRank takes those labels, and matches them to how creators of web pages have used matching labels to describe URLs. If I search for "Umich" Google will match me to a URL that others have described with the "Umich" label. In effect, Google is capturing implicit recommendations for URLs based on how different people label the link to those URLs (Bianchini, Gori et al. 2005). Social data mining is a useful method in that it does not require additional work for users to state their preferences. The problem with this method is that it only allows generalized recommendations, though key term searching, for example, allows users to more easily search for recommended items.

Many recommender systems depend on explicitly solicited user preferences to match users with similar interest. MovieLens requires users to rate movies in order to correlate stated preferences with those of other users. If the user does not provide their preferences, there is no basis to make matches with other users. Using explicit ratings

[10] People Helping One Another Know Stuff

resolves problems of multiple goals found in implicit data mining, but depends on relatively high user effort to input preferences.

Some online recommender systems use a hybrid of preference solicitations to recommend content. Netflix[11], an online movie rental site, combines explicitly stated user preferences with rental history to compute collaborative filtering recommendations. Tivo uses explicit ratings in terms of their "thumbs up" and "thumbs down" tools and implicit ratings in terms of shows chosen to watch to make recommendations to users about shows they have not rated, but may enjoy (Ali and vanStam 2004).

Inherent barriers in recommendation and reputation

Slashdot uses a recommender system to provide information about how previous users valued comments after they had read them.

For recommender systems to be effective, they must overcome a variety of potential inherent barriers. Scarcity, for example, is the problem of eliciting sufficient feedback on items to make future recommendations (Balabanovic and Shoham 1997; Avery, Resnick et al. 1999). Users may benefit from the recommendations of others without subsequently enriching the system with their own recommendations. Scarce ratings can occur because of small user populations, or because social loafing (Karau and Williams 1993), in which responsibility for action is distributed across a group leading to decreased participation. Especially in recommender systems that depend on explicit statements of user preferences, eliciting sufficient feedback may be difficult.

The first rating in a system can have a disproportionate effect on the direction of future ratings. This tendency is referred to as "herding" (Banerjee 1992), or information

[11] http://www.netflix.com/

cascades (Bikhchandani, Hirshleifer et al. 1989; Bikhchandani, Hirshleifer et al. 1998). An example of this tendency in the psychology literature is to the tendency of pedestrians to look up when they see someone else doing so (McPhail 1991). This can be problematic for online rating systems in that content may be incorrectly scored because of earlier ratings. There was an indication of this found in MovieLens ratings, where users ratings were affected differently by being shown different scores from previous users (Cosley, Lam et al. 2003).

Users also need to trust and understand the recommendation being provided. Users may not trust recommendations that do not reflect their experience. For example if they are recommended a movie they did not like, they may lose overall trust in the system. Users have shown a preference to have how their preferences were computed explained in the display of results (Herlocker, Konstan et al. 2000). One technique for adding feedback about the correctness of ratings has been the use of meta-rating techniques, in which the accuracy of an inferred preference is itself rated (Schafer, Konstan et al. 2002). There may be a difference in the provision of ratings when the ratings themselves are made publicly or privately (Dellarocas 2003). Ratings made privately may be trusted less than those publicly tied to another user, as consumers of the rating would have more context in the latter condition for determining the motivations of the rater.

Use of recommender systems in online conversation

Recommender systems may be useful for structuring online conversations. These systems disperse the work load of identifying valuable content among many users, allowing for scalability. Recommender systems may support minority opinions by identifying clusters of users who have shared perspectives. By allowing for filtering based on user preferences, online conversation rating systems offer tools for resolving information overload without constraining membership. Although Tapestry and GroupLens are both examples of the application of recommender systems to online conversations, neither was widely used. The Slashdot case operates on a larger scale than either previous example. This scale, as well as the long time use of the Slashdot recommender, makes this an interesting case to extend the literature on recommender systems.

Conclusion

The goal of this narrative was to show that online conversations have traditionally faced a similar set of problems, the solutions for which have been sub-optimal for encouraging large-scale, public interactions. Recommender systems offer a toolset that might help resolve some of the tensions between managing large-scale online interactions without losing the benefits of that scale. Slashdot provides an interesting "in the wild" example of a recommender system used in online interactions.

Chapter 3

Follow the Reader: Filtering Comments on Slashdot[12]

Abstract

Large-scale online communities need to manage the tension between critical mass and information overload. Slashdot is a news and discussion site that has used comment rating to allow massive participation while providing a mechanism for users to filter content. We find that many users are actively using ratings to restructure how comments are shown to them. However, many users are not using ratings to guide their reading behavior, which may be caused by friction: the perceived cost of changing a setting versus the perceived benefit for having done so. We recommend leveraging the efforts of the users that do use the rating system to reduce the cost of changing for all other users. One strategy is the creation of statistically motivated schemas that represent types of reading behavior. Another strategy is to dynamically set filtering thresholds for comments, based in part on the choices of previous readers, which are far better predictors of readers' choices than content features such as the number of comments or the ratings of those comments.

[12] Work done with Erik Johnston and Paul Resnick, pending submission.

Introduction

Massive, many-to-many online discussions may have a range of benefits, including information sharing(Ackerman, Swenson et al. 2003), the creation of new forms of social capital (Resnick 2001), scientific discovery (Finholt and Olson 1997) and political discourse (Kelly, Fisher et al. 2005). However, users may never realize these positive outcomes if worthwhile content is buried in the rush of messages typical in online discussions involving hundreds or thousands of participants. Information overload occurs when an individual is unable to process or use all of the available inputs. Early research in psychology showed that humans have a limited cognitive capacity to accept new information (Miller 1956). This limitation is exacerbated in CMC settings where technology allows large numbers of participants in online discussions.

Jones et al. (Jones, Ravid et al. 2002) examined Usenet newsgroups and found that users are more likely to respond to simpler messages in situations of overload; that users will end participation as overload increases and that users generate simpler responses as overload increases. Jones and Rafaeli (Jones, Ravid et al. 2004) also found that in larger Usenet newsgroups, as measured by number of postings to the group, messages became shorter. Previously, Jones and Rafaeli (Jones and Rafaeli 1999) proposed that communication online takes an S-shaped pattern of frequency of occurrence. Early in the existence of a conversation space, or “virtual public” to use their term, there is a struggle to achieve critical mass of people contributing to the conversation. A sharp increase after that critical mass is achieved results in information overload, and communication levels off as participants are dis-incentivized by the rate of messages. Jones and Rafaeli (Jones, Ravid et al. 2004) have described this as a tension between the critical mass needed to benefit from "shared public online interpersonal interactions" and the breakdowns that occur in information overload conditions. Butler (Butler 2001) found that large listservs lost a higher proportion of their users to attrition than smaller groups. If there are too few interactions, users may decide they do not have much to gain from participating themselves in that group. If there are too many interactions, users may become dissatisfied with the amount of "noise" in the group and seek less active groups.

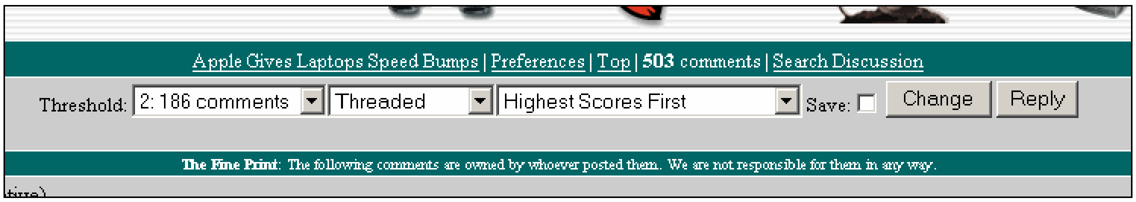
Several strategies have been employed to help readers make sense of complex online environments. Netscan provides statistics and visualizations about Usenet participation, helping to illustrate newsgroup interaction dimensions and styles (Smith 2002). Visualizations have also been used to provide context to email discussions (Fisher 2004), Wikipedia interactions (Viegas, Wattenberg et al. 2004) and cross-linking in blogs and personal Web pages (Adamic and Glance 2005). Ackerman (Ackerman, Swenson et al. 2003) has used a combination of software agents, social rules and automatic text summarization to “distill” comments from an online discussion forum into discrete discussion summaries.

Rating systems have been used widely to provide information about people, products and content online (Terveen and Hill 2002). An early attempt to apply ratings to online messages was the Tapestry (Goldberg, Nichols et al. 1992) system, which had users rate content in an intranet messaging site, and then made recommendations for new articles based on collaborative filtering strategies. GroupLens (Resnick, Iacovou et al. 1994) applied this idea to Usenet and introduced the technique of automated, personalized recommendations based on the ratings of people with similar tastes, a technique that is now commonly referred to as collaborative filtering or recommender systems. GroupLens displayed its predictions of user interest in a message as a guide, but did not automatically use the predictions to sort or filter messages.

As described more fully in Chapter 2, Slashdot collects ratings from selected readers who act in a “moderator” role. Moderators assign labels such as “Informative”, “Funny”, or “Troll” to comments. Based on those labels, each comment accumulates a score from -1 to +5. Slashdot not only displays the score of each message, but also allows users to sort or filter the available messages based on those scores.

Here, we examine the impact of ratings on readers. First, we look for evidence about whether users prefer that ratings be used to sort or filter their messages. Then, we suggest new features that could be introduced, to make the ratings more useful to more readers. Evidence about users' preferences comes both from responses to subjective survey questions, and from objective measures of users' choices of setting when reading comments.

Figure 3.1: Options for changing comments settings temporarily within discussion forums.



Literature on user access of customization features indicates that few people take advantage of advanced interface options. Mackay (Mackay 1991) conducted interviews with and collected automatic records of customization activities of 51 members of a project at MIT over a four month period. She found that users engaged in a continual analysis of the cost of learning to customize versus the benefit of having done so. A significant barrier includes the lack of time to learn the specific tools available. Triggers for changing preferences included when changing allowed the user to avoid learning new behaviors. Customizing to make an environment more aesthetically pleasing was generally avoided. Mackay concluded that users "satisfice" rather than optimize, that is they do the minimum necessary to use the software.

Page et al. (Page, Johnsgard et al. 1996) studied 101 users of a text-editing program and found that 92% of participants did some form of customization. They also found a strong relationship between how much the software was used and the amount of customization that took place. The more the system was used, the more customization users engaged in. This research suggests that users only change options when absolutely necessary, for instance when their work is being seriously impeded, or they are heavy users and find certain repetitive tasks easier to customize. Mackay mentions satisficing as a possible explanation for this behavior. Satisficing is a term from Simon (Simon 1996), and relates to the bounded capacity for humans to make rational decisions. Cognitively, people are unable to hold all the variables necessary for making a choice in mind, so they instead choose a "good enough" solution. March and Simon (March and Simon 1958) argue that only in rare occasions are people concerned with the optimal solution, and that people compare the marginal-benefit of making a decision with the cost of that decision in terms of risk.

Consequently, we must be circumspect in making inferences about the preferences of readers who have not made changes to the default settings. Our approach is to focus on *lead* users, those who have made explicit choices about their viewing settings. We extrapolate from the revealed preferences of those lead users to infer what other, less proactive users, might like the system to do, if there were no friction preventing them changing how they view comments.

Viewing Slashdot comments

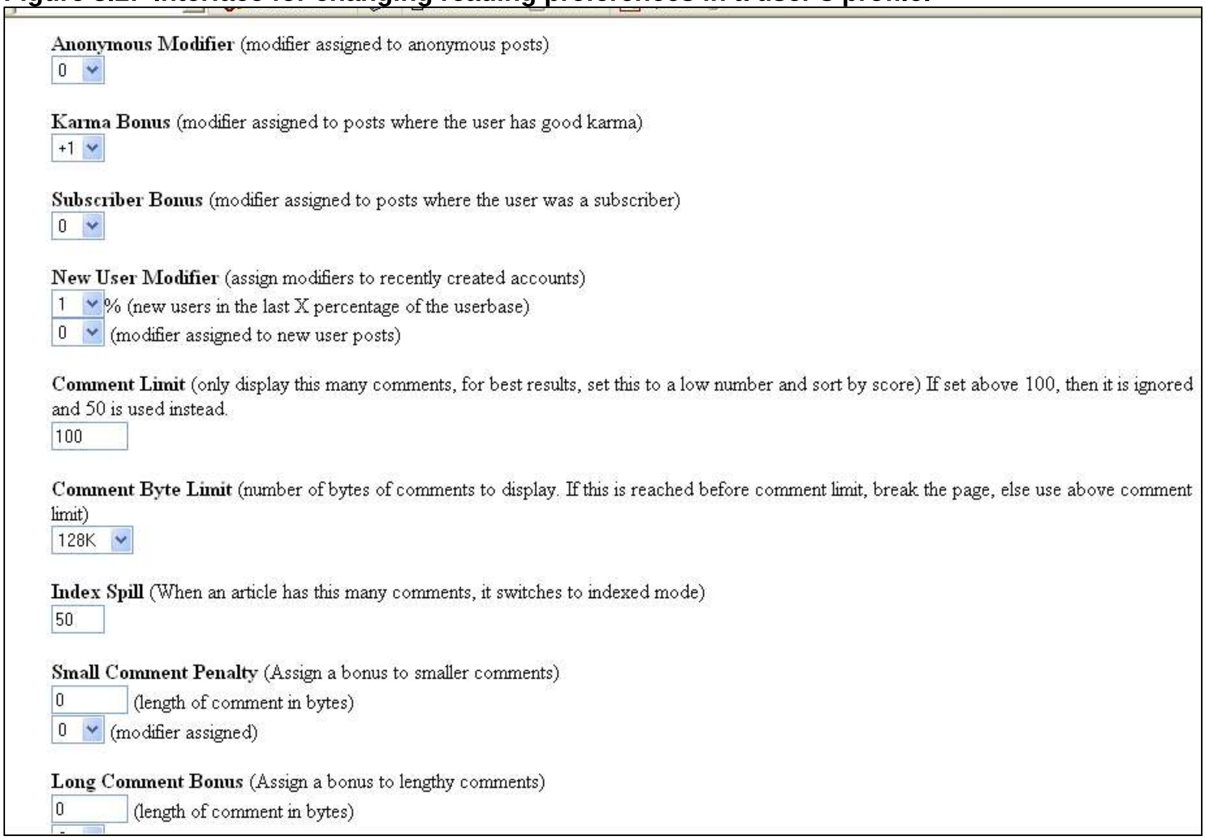
The default view for comments in a Slashdot forum is a threaded structure with a filtering "threshold" of +1. Responses are indented and responses to responses are further indented. The responses to any comment or listed in chronological order, with the oldest ones first. At the top level, the direct responses to an initial news story, full text is shown for comments that have a score of 1 or higher. For comments deeper in the thread (i.e., replies to other comments), full text is shown for comments rated 4 or higher, a single line is shown for comments rated 1-3, and the comment is omitted if its score is below the threshold of 1.

Any reader, anonymous or registered, can make a one-time change to the display of comments for the current story, using the options shown at the top of Figure 3.1. Clicking on any of the links titled "n replies below your current threshold" will set the threshold to -1. Through pull-down menus shown at the top of Figure3.1, a reader can set the threshold to any value between -1 and +5. A reader can also sort messages based on their ratings; at each level of the response hierarchy messages with the highest scores are displayed first, rather than those that were posted first.

In addition to one-time changes on a per-story basis, registered users may also make permanent changes in their personal profile. Any of the settings that can be elected on a one-time basis can be set as the user's standard setting, the starting setting when reading comments about any new story.

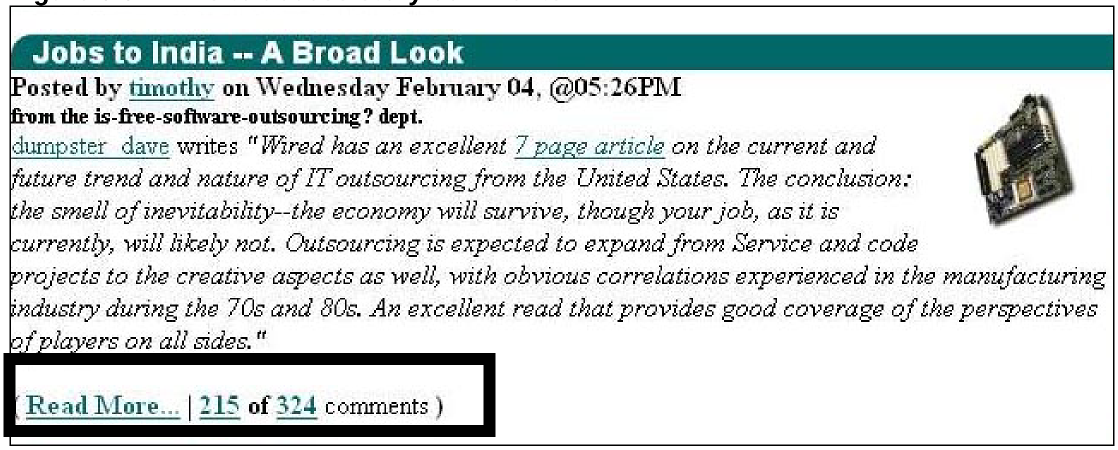
Registered users can also make other changes in the personal profiles that are not available as one-time changes. Options here include special penalties or bonuses based on such things as length of the comment, the reason given for moderation and whether the user is anonymous, or registered with high karma, which is described in Chapter 2 as the reputation score registered users have on the site. These settings generate personalized scores for comments. A comment moderated as "Funny" may end up with a score of +3 for one user, who has given additional bonus points to funny messages, but as -1 for another user who has assigned a penalty to comments rated as "funny". In total, there are 36 options that registered users can check to change their viewing patterns. Figure 3.2 shows the interface used to change these settings.

Figure 3.2: Interface for changing reading preferences in a user's profile.



The way that a reader arrives at a page can affect the initial display settings, in particular the threshold level. As shown in Figure 3.3, each story on the index page is associated with several links. Following the "Read More" link will take a reader to the comments page with the user's standard threshold from their profile (or the default of +1 if the reader is not logged in). Clicking "X" in "X of Y comments" does the same thing; X is the number of comments that are at or above the threshold that will be set. If a reader clicks on the "Y" in "X of Y comments", the threshold is set to -1 when the page of comments is displayed.

Figure 3.3: Links to access story comments.



Methods

Slashdot server logs were analyzed to determine patterns of customization and reading behavior of site members. The first log is a pre-existing Slashdot database used to track user profile settings. This log includes moderation modifiers, viewing preferences, and other variables that affect how comments are displayed. The second server log is a general user information table that records user history like account creation date, reputation level, comments made and similar items. The third log tracks user requests for pages from the site. This log contains time stamps of user page requests, user identification numbers, the URL of the page requested and whether any interface changes were the result of a specific user selection. Slashdot has previously only logged aggregate number of page requests, and this log added the capacity to tell for each user which page was being requested. The page request log represents an 80 hour period in mid-July, 2005.

The two user preference logs contain information for the 875,573 users who had created accounts on Slashdot. Of those users, 90,273 logged in and requested a total of 2,416,331 pages. Slashdot pages are divided into several sections, only some of which display user comments. Of the number of page requests listed above, 47% were for pages that would display user comments. Additionally, 2,613,181 anonymous page requests from 409,932 IP addresses were also logged. Given the nature of IP information, it is not possible to claim that each IP corresponds to one user, but it can give a rough estimate of the pattern of use, as argued in a recent survey article on mining Web logs for usage patterns (Facca and Lanzi 2005). Of these page requests, 2,341,628 were for pages that would display user comments. For anonymous users, the index page that is a gateway into the site is most often displayed as a static page. An artifact of the logging procedure was that these static page requests were not logged, so that the pages with user comments are a smaller portion of overall anonymous user hits than represented here. All requests for pages with user comments are dynamic, and thus recorded here.

We also conducted a survey with registered Slashdot users to learn more about their characteristics and attitudes. The sampling frame for this survey was the list of registered Slashdot users, based on the Slashdot assigned unique identifying number.

User identification numbers were compared against IP addresses to assure that multiple numbers were not held by single individuals. Each day a script chose 10% of the registered Slashdot users to receive an invitation to participate in the survey. Potential respondents received an invitation to participate at the top of the index page of the site, an area commonly reserved for site messages, including notification to moderate and meta-moderate.

Between June 15 and June 20, 2005, 8121 respondents participated in the study. The overall response rate for the study period was 19.1%, with some variation per day. The survey responses may be suspect for several reasons. First, there may be non-response bias: the people choosing to respond may be those most aware of and most favorably disposed to the Slashdot moderation systems. Second, the instrument may not be eliciting responses to the phenomenon to of interest here. For example, readers may think the ratings are helpful simply because they create feedback to writers and incentives for them to post better comments, even if they do not use them for sorting or filtering when reading. These potential problems with the survey are not extraordinary for surveys in general, and there is no secondary evidence to indicate that the responses are indeed unrepresentative of Slashdot as a whole. Appendix 1 contains the instrument used for this survey.

Slashdot survey respondents showed many shared characteristics. Respondents were 98% male, and 62% had completed a college or graduate degree. 71% of users were between 18 and 34 years old. Slashdot users also reported high levels of technology use. 84.5% of respondents reported visiting 2 or more news and discussion sites besides Slashdot each day. 13.5% reported visiting 6 or more news and discussion sites per day. On a scale of 1-7 where 7 indicates “Extremely” when asked about how expert they consider themselves with computers, Slashdot respondents rated themselves an average of 6, with 24% of respondents placing themselves in the highest category of computer expertise.

Ratings are useful to readers

The survey asked users to indicate the extent to which they felt Slashdot’s moderation system was important in identifying good comments, and the response was generally favorable. 84.7% of respondents agreed somewhat or strongly with the statement “The moderation system is important in identifying good comments.” Only 8.5% disagreed somewhat or strongly.

The choices readers make about how comments are displayed provide a behavioral measure of their reaction to the views where comments are sorted or filtered based on the comments’ scores. If a reader never strays from the default filtering threshold, we can not infer whether he prefers that setting over all the others, or whether he simply has not explored the other options[13]. On the other hand, there are a number of explicit choices that users can make that reveal a preference for an alternative view that also uses scores to filter or sort the messages, or for an alternative view that does not use scores in that way. In this section, we consider what fraction of the users reveal a preference for a view of comments that does or not make use of the scores.

Table 3.1 shows how frequently users make changes to either employ ratings in customizing how comments are displayed, or take steps to suppress the role of comment scores as shown by logs of user settings. Users may occupy more than one state, so percentages in the table are not cumulative. The grayed area represents settings changes a user makes to employ ratings to customize their view. The cells in white are settings that suppress comment scores.

Table 3.1: Percentages of users who used ratings to change their views, or made changes to suppress ratings. Categories are not mutually exclusive.

User action	N	Percent
Permanently change threshold to 2-5	26,349	30.2%
Permanently change sort order to “Highest first”	14,132	16.2%
Temporary change threshold to 0-5	5,944	6.8%
Temporary change sort order to “Highest first”	927	1.1%
Change moderation label values	13,803	15.8%
Permanent change threshold to -1, no sort	5,209	6.0%
Temporary change thresholds to -1	1,572	1.7%

[13] For grammatical convenience, we refer to Slashdot readers generically as male. Of survey respondents, 98% were male, suggesting that the alternative we often adopt, of making generic users female, would be misleading in this case.

with forum dropdowns		
Temporary change thresholds to -1 through navigation	12,473	14.3%

To acquire an understanding of how different categories of users are employing ratings differently to customize how they see comments, we divide users into three groups: those that never strayed from their default setting, those that explicitly chose score-based views and those who customize comment views to suppress the role of comment scores.

Readers Who Leave Defaults Intact

Users can change their viewing settings either permanently via their user profiles, or on a one-time basis. Originally, the default threshold for Slashdot comments was 0. At an unrecorded time in 2000, the default threshold was changed to +1. 45.4% of users who logged in during the study period had their permanent viewing threshold set to either 0 or +1 and had not changed any of the other settings in their user profile that related to moderation score. In particular, they did not specify sorting based on scores, and did not change the score bonuses associated with labels such as funny, troll, or interesting.
Of these users who had not changed their permanent settings in their profiles, 95% also did not make any one-time changes to the threshold or sort order. This means that 43.1% of the total number of registered users who viewed comments during the study period made neither temporary nor permanent changes to how they view comments, and we are not able to infer whether they liked having comments filtered based on their scores.

Readers Who Choose Sorting or Filtering

Of registered users who logged in during our study period, 39.2% had set their permanent profiles to include either a threshold of 2 or higher, or sorting based on highest scores first, or both. An additional 14.6% made changes to the bonus scores assigned to particular comment labels, and maintained a threshold higher than -1 in their permanent profiles. An additional 1.4% of users did not make such permanent changes, but at least once made a one-time change to select a view threshold other than -1 or sorting based on scores. Overall, 48.3% of users made explicit choices in at least one of these three ways to use comment scores to affect their view.

Readers Who Suppress Use of Scores

The only viewing settings that suppress the use of comment scores are a filtering threshold of -1, and a sort order based on chronology rather than scores. We have already seen that some users made no changes to their view settings, and some made permanent or one-time changes to settings that made use of comment scores. This group of score suppressors includes 5.7% of the user population who set their permanent profile to have a threshold of -1 and no sorting based on scores, and never made a one-time change to a setting that used scores. The score suppressors also includes 0.003% of the user population who had not changed their permanent profiles and whose only one-time changes were to set the filtering threshold to -1. As mentioned above, users can temporarily change their thresholds to -1 by navigating the user interface, specifically by selecting the option “N comments below your threshold”. We found that 14.3% of users made at least one change to a -1 threshold through navigation rather than through the drop-down menus. Of all registered users who visited the site during the study period, 3.8% navigated to -1, but made no other changes to either permanent or temporary settings. By including those users who navigate to -1 rather than choosing it, we find that 15% of the users had acted to suppress scores at least once.

Summary

We have divided users into three groups: 45.4% who never strayed from the default setting; 48.3% who explicitly chose score-based views, and 15.0% who were score suppressors, with some overlap between groups. The ratio of score-choosers to score suppressors, 3.2:1, is an estimate of the portion of Slashdot users who like using scores as a reading aid at least some of the time.

The estimate is conservative for three reasons. First, users who never strayed from the default, and thus never expressed a preference, were actually employing a filtered view, with a threshold of 1 or 0. Many of them probably did prefer their settings over other available settings, but because they had not had to explicitly choose those settings, we count them as having indeterminate preferences.

Second, we included as score suppressors people who had not suppressed the use of scores in their permanent profiles but had suppressed scores on a one-time basis. Many of them may actually prefer as a general rule to have a filter threshold, and only sometimes suppress its use. In our conservative interpretation, however, we count these users as score suppressors, since they never made an active decision to rely on scores, and they sometimes made an active decision to suppress them.

Third, even for users who use ratings for filtering sorting, ratings might be playing a role in their viewing behavior. Scores are still displayed at the top of each message, and they may be scrolling and "berry picking" to select highly rated comments from the forum.

The 3.2:1 ratio means that, of registered users whose behavior showed a preference one way or the other, over three quarters preferred, at least some of the time, a reading view based on comment scores. This estimate is consistent with the high percentage of users who gave favorable responses on the survey. We conclude that moderation scores are, indeed, useful to most Slashdot readers, though a minority prefer not to use them.

These patterns lend some support to the idea that ratings are being used to change how comments are displayed, and that users seem to be employing ratings to explore the offerings at either end of the spectrum. Another possible implication of these patterns is that Slashdot readers are using ratings to view comments with different purposes. Readers who start out at +5 and maintain a high threshold might be readers with little time and want to get the core arguments in the thread. Users who opt to go to -1 may have more time to explore. Looking at the average threshold that a user chooses, 15.3% of users always change their temporary threshold to -1 and 11.4% always change their threshold to +5. It could be that the viewing population can be divided into groups of "explorers" and "exploiters". Later sections will address how to identify the characteristics of different ratings use types.

Friction occurs on the site

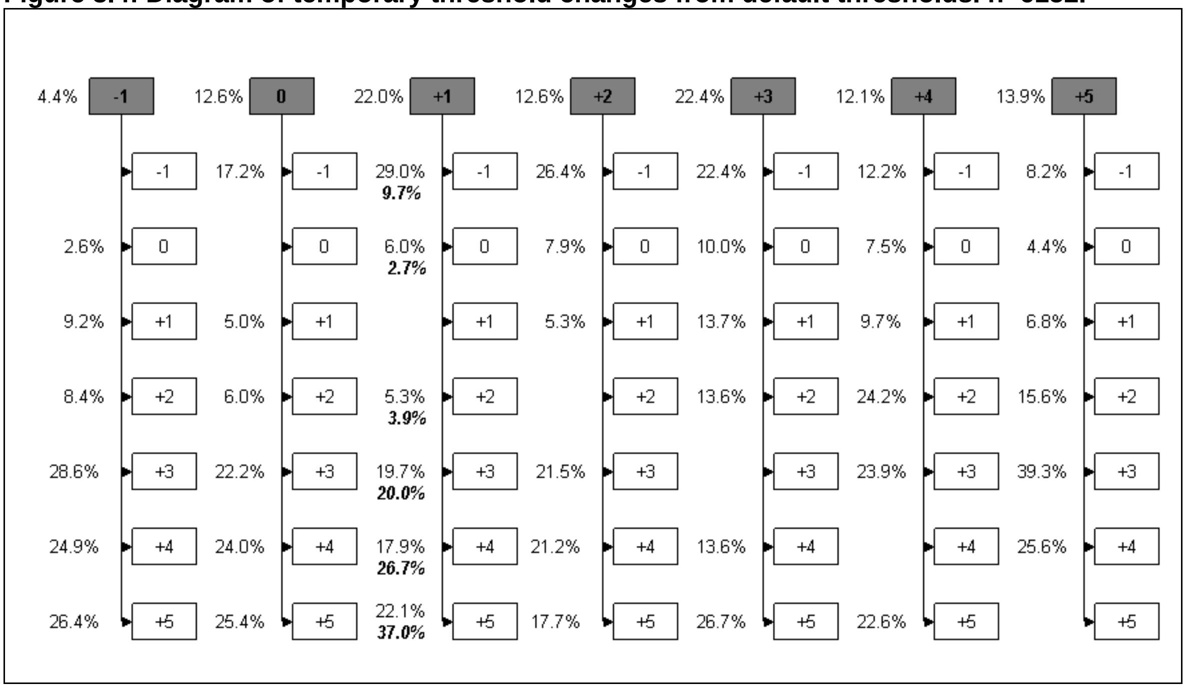
"Friction" is a shorthand term for the factors that prevent a user from using the site interface to change how comments are displayed. When viewing a forum, how comments are displayed may be sub-optimal, as in too many comments, or not ordered in a sensible way. The best strategy for the user might be to change the viewing threshold or sort order, yet the cost of doing so might be higher than the expected benefit from the change. Each choice presents a cognitive burden; each mouse click takes time, and each page reload involves a wait and then a redisplay that may destroy the user's reading context.

Are these forms of friction preventing Slashdot from changing to more preferred settings, especially making one-time changes? Here again, when people take no action we can't tell what their preferences are. However, we can again make some inferences based on the patterns of changes made by those who do make changes.

First, of registered users who visited the site during the study period, we found that 10% made at least one temporary change to how comments are displayed. Of these users who make temporary changes, we found that they tend to make changes relatively often. We computed the ratio of stories where users change the threshold over total stories they viewed. This ratio excludes the first change that the user makes in order to account for the use of a threshold change as a selection factor in identifying users. For users who made at least one change, and viewed more than one story, on average they changed settings on 32% of the stories they read. Half of the users changed settings on 22% or more of stories they read, and a quarter of users changed settings on 60% or more of stories they read.

Second, we find that readers who make one-time changes to filtering thresholds tend to make big jumps, not move to adjacent thresholds. Figure 3.4 below shows the pattern of changes from different thresholds for registered users. The shaded boxes are the starting level, with the numbers next to them representing their proportion of all threshold levels set by registered users. All anonymous users start with a default threshold of +1; their changes in threshold are recorded separately from registered users, in bold italics under the registered user score for that threshold level. These numbers reflect only the first change on the first story that a user reads. This is done to account for multiple changes within story that would cloud the pattern of threshold changes.

Figure 3.4: Diagram of temporary threshold changes from default thresholds. n=6152.



As shown in Figure 3.4, the most common moves are to high thresholds of 3 or greater. The other extreme, -1, is also somewhat popular, but the middle values of 0-2 are less popular. Most importantly, for all starting thresholds, moves to thresholds adjacent to the starting threshold are less frequent than moves of two or more places.

The patterns of big jumps suggests that there is significant friction from clicking, waiting for a page reload, and then getting reoriented after that reload. It seems natural to think that, for any user the distribution over stories of optimal filtering threshold would be a smooth, single-peaked distribution. That is, if n is the peak, the threshold that is most frequently the optimal one of the reader, then n-1 and n+1 should be optimal more often than n-2 and n+2.

Without any friction, then, we would expect that the most frequent changes would be to adjacent thresholds. With friction, however, when the adjacent thresholds are the optimal ones, they may not be sufficiently better than the default threshold, and the user would not bother to change to them. Only when a threshold farther away is optimal would it be sufficiently better than the default to overcome the friction. The pattern of actual threshold changes is consistent with this explanation.

Reducing friction

New system features may be able to reduce the friction that keeps users from finding their optimal settings. One idea is to identify useful schemas for score bonuses that users could select either for the permanent profiles or on a one-time basis. Another idea is to automatically change filtering thresholds. We consider each in turn.

Schema Based Score Bonuses

When moderators on Slashdot rate a comment, they are actually assigning a label that moves the comment's score up or down based on the value of the label. Registered users have the option to assign additional weights in either direction to the effect any given label has on the score of a comment. Table 3.2 shows the labels associated with comment ratings, the typical value attached to the label, and the percentages of users who have added weight to the comments in either direction. This table acts as a distribution of the direction of weights applied to moderation labels.

Table 3.2: Percentage of moderation label changes of those users who have changed at least one label setting.

Label		Frequency	Percent
Flamebait (-1)	Up	1515	5.4
	Down	11497	40.7
Funny (+1)	Up	12713	45.0
	Down	5413	19.2
Informative (+1)	Up	12451	44.1
	Down	262	0.9
Insightful (+1)	Up	13074	46.3
	Down	365	1.3
Interesting (+1)	Up	11144	39.5
	Down	253	0.9
Offtopic (-1)	Up	1051	3.7
	Down	9455	33.5
Redundant (-1)	Up	677	2.6
	Down	10229	36.2
Troll (-1)	Up	1270	4.5
	Down	12718	45.0

Just 3.2% of registered users have personalized the score bonuses assigned to moderation labels. When users do change score bonuses, however, they tend to change more than one: 32.6% changed only one; the median was three; 13.6% changed all eight. A noticeable minority, about 5% of registered users, apply positive bonuses to moderation labels like “Troll” and “Flamebait”, essentially guaranteeing they will view material the system is largely designed to demote. Compared to other users, those who had applied positive weight to negative labels had higher average moderations spent (54

vs. 49; t=-2.67, p<.01), higher karma (12.8 vs. 10.6; t=-8.82, p<.01) and had made more comments (179 vs. 67; t=-26.64, p<.01). This is counter-intuitive, as one would expect users less entrenched in Slashdot to be the ones who might want to see such content. One explanation might be that experienced users enjoy the entertainment value of comments that are rated negatively, "living in the muck" as it were.

If changes to score bonuses based on labels occurs in clusters of changes, it might be possible to find patterns of weights that indicate types of use. For example, people who added positive bonuses to the "Interesting" label may be more likely also give a positive bonus to "Insightful" than to "Funny." This suggests that users might find it helpful to be able to choose, with a single click, a schema that identifies a cluster of settings for score bonuses. We identify several possible schemas: gem seekers, serious thinkers, muck rakers, and humor seekers.

Grouping of moderation labels

To detect sensible schemas based on moderation labels, we grouped the most common changes in label values that users made. Some users are reinforcing the rating directions of the moderation labels by assigning additional weights to positively connoted labels. One possible schema related to this tendency might be a grouping for "Gem Seeker" users who further reinforce rating directions tied to labels to constrain the comments to only those that are seen as most cogent. To see if this pattern currently exists, we looked at those users who added weight to the "Interesting" label to see which other label modifications occur concurrently. Table 3.3 shows the number of people who had added weight to the "Interesting" label who also weighted ratings for other labels either negatively and positively. The percentage column indicates what percentage of the number who weighted "Interesting" positively that represented. Overall, 11,144 users added positive weight to the "Interesting" label.

Table 3.3: Percent of users who change label values up or down who have also modified "Interesting" comment with higher ratings.

	n	% in "Interesting"
Informative (up)	9,889	88.7
Insightful (up)	10,080	90.5
Funny (Up)	7,363	66.1
Funny (down)	1,133	10.2
Off-topic (down)	5,852	52.5
Troll (down)	7,120	63.9
Flamebait (down)	6,803	61.0

For users who added weight to the "Insightful" label, we found that they were very likely to also add ratings value to "Informative" and "Insightful". Conversely, they also tended to further decrease the rating values associated with negative labels. These users are elevating weights in the rating system to create a schema whereby highly rated and lower rated comments are winnowed dramatically. This shows that some users have already customized a "Gem Seeker" view that could possibly be of use to future readers.

As shown above, some users assigned positive ratings to typically negative labels like "Troll" and "Flamebait". It could be that these users are setting a cluster of modified labels that bring poorly rated comments to light. A schema for "Muck Rakers" might be of use to those who find entertainment value in the personal insults and deceptions that constitute "Troll" and "Flamebait" comments. A total of 1270 users modified the "Troll" moderation tag to increase the score of comments that received that label. Table 3.4 shows what those users who modified "Troll" as positive did with other labels.

Table 3.4: Percent of users who change label values up or down who have also modified "Troll" comments with higher ratings.

	n	% in "Troll"
Flamebait (up)	802	63.1
Off-topic (up)	469	36.9
Funny (Up)	484	38.1
Funny (down)	190	15.0
Interesting (down)	83	6.5
Insightful (down)	74	5.8
Informative (down)	74	4.5

This pattern is slightly different than that found in the "Gem Seeker" schema. While "Muck Rakers" do seem to promote comments with negative labels, they also are not widely demoting comments with positive labels. In fact, for users who modified the "Troll" label positively, they more frequently added, rather than subtracted, additional weight to the positive labels. This describes a schema where users are not wanting to see only negative comments, but do want to pull them up below the default thresholds to increase the chance of viewing them along with the positively labeled comments.

Grouping “Serious Thinkers” and “Humor Seekers” seems largely to depend on how users react to the Funny label. Around 10% of users who rate up Interesting, Insightful and Informative took the additional step of rating Funny down. There seems to be at least one grouping of users who prefer to view only those comments that other users have not labeled as funny. On the other hand another group of users who rate Funny highly also increase the ratings for other labels including Informative, Troll and Insightful. These users seem to want to read comments labeled as Funny, despite their other actions with different labels.

This data supports a view of Slashdot users with heterogeneous goals in reading comments who use the label weighting system to enact those goals. Users without the ability to similarly customize, either because they are anonymous users without the option or high-friction registered users, may benefit from viewing schemas based on those users who do create custom ratings.

Automatic Threshold Changes

Given the apparent friction, that users who would prefer other settings will not necessarily select them, there are opportunities for the system to assist in those selections. For example, based on the pattern of one-time changes a user makes, the system could offer to make changes in the user’s permanent profile.

Here we focus on automatic threshold changes, where the system infers, for a particular story, that a user would be likely to prefer a different setting. Given the costs involved in manual changes to settings, we have limited information about whether, for particular stories, users would have preferred different threshold levels than the ones they actually used.

However, we can get some indicator about what kinds of stories would benefit from one-time increases in filter thresholds by examining the behavior of lead users, those who seem to have less friction in making one-time changes. We examine the behavior of users who read at least five stories during our study period, and made at least one one-time increase in their filtering threshold. There were 884 such users; they read a total of 428 stories during the study period, and increased their threshold while reading 23% of them.

We constructed a logistic regression model to predict when our lead users decided to make one-time increases in their filter thresholds. Several factors enter the equation as independent variables. First, of course, is the default threshold set in each user's profile: as Figure 4 showed, threshold increases were far more likely from a +1 threshold than at +4 threshold (an increase is not possible from a +5 threshold). The second factor is the total number of comments that had been written about a story at the time the user read that story and its associated comments. The third factor is the quality of those comments, as measured by the percentage that achieved scores of 3 or higher.

Analogous to collaborative filtering, where the opinions or behavior of other people leads to recommendations about items to attend to, the threshold choices of other users can be mined to predict when users increase thresholds: if many past readers increased their thresholds, the current reader may be more likely to as well. To operationalize this variable, for each reading of a story we computed the percentage of previous reads of that story by other lead users where the users increased their filtering thresholds.

Table 3.5 shows results of estimating the binary logistic (logit) regression model. The base result is for users with a permanent threshold of -1. When interpreting a logit model, positive coefficients indicate higher probabilities with the amount of the coefficient indicating how strongly that variable affects the probability of the event occurring. P-values with an asterisk (*) are statistically significant at the .05 alpha level or better.

Table 3.5: Logistic regression predicting that a lead user will increase their comment threshold.

Pseudo R-squared	0.16		
n	26,335		
	Coef.	**Z**	**P>|z|**
Default thresh = 0	0.164	2.78	.005*
Default thresh= 1	0.251	4.52	.001*
Default thresh= 2	-0.063	-1.04	.296
Default thresh= 3	-0.778	-13.42	.001*
Default thresh= 4	-1.251	-15.59	.001*
Number of comments at time of view	-0.001	-5.11	.001*
Percent of comments in forum at score 3 or higher	-0.700	-4.19	.001*
Percent of lead users who already viewed that change	5.714	48.68	.001*
Constant	-1.962	-28.18	.001*

Users with thresholds set at 0 or +1 in their permanent profiles were more likely to make a one-time increase in the filtering threshold than those with a permanent threshold of -1. This makes sense since presumably those who have permanent thresholds of -1, which turns off filtering, are less likely to want to turn it on. On the other hand, those with permanent thresholds of 3 or 4 were less likely to make one-time increases. They have relatively little to gain from increasing thresholds from their already high levels.

If users were merely trying to read a manageable number of comments for each story, we expect them to raise their filtering thresholds when there were more comments above their threshold. Somewhat surprisingly, this is not the case: both more comments and a higher percentage of comments getting scores above 2 led to *decreases* in the probability that users would increase their threshold. In the case of number of comments, a coefficient of -0.001 indicated a very slight effect on the final probability.

While statistically significant, the level of impact of the total number of comments is small. For a lead user with initial threshold of +1, reading a story with the median number of comments (181), with the median percentage of highly rated comments (30), and the median percentage of previous lead users increasing their thresholds (15%), the predicted probability of this user increasing his threshold while reading the story is 46%. If, instead, there were 350 comments (the 75^{th} percentile), the probability of increasing his threshold would decrease to only 44%. The impact of the percentage of highly rated messages is also comparably small. If that percentage increases to 22% (the 75^{th} percentile), the predicted probably only makes a slight further decline, to 43%.

On the other hand, the behavior of prior users is both statistically and practically significant. If the percentage of prior users who increased their threshold moves from 30% to 42% (the 75^{th} percentile), the predicted probability jumps all the way to 60%. This large impact suggests that a collaborative filtering style interface, where the threshold decisions of early readers affects the automatically selected thresholds of later readers could be quite effective. It is not that users choosing a threshold observe the actions of lead users, but rather that the lead users predict what others will do.

It may be that the actions of prior users already take into account the number of comments and the percentage of them that were of high quality. Thus, in the regression model, there would not be any additional effect attributed to those factors. This can not explain the negative coefficients, however. One explanation may be that more comments and more highly rated comments are indicators of better conversations, and readers increase thresholds only for lower quality conversations.

Conclusion

We found that registered Slashdot users could be divided intro three categories: those who never change the default comment display, those who use ratings to modify the comment display, and those who change the comment display to suppress ratings. Although a significant portion of users did not change from system set defaults, an even larger group of users employed ratings to modify how comments were displayed in forum. Interestingly, a non-trivial segment of users made modifications to diminish the effects of comment rating. For those users who don't make changes, or only make changes rarely, we find some indication that friction exists, preventing ready changes to how comments are displayed. This friction is related to the perceived cost of changing opposed to the perceived benefit for having done so.

A possible remediation for this friction is to support dynamic comment view changes based on the behavior of other users. One aspect of this is the use of schema labels as meta-data to create groups of user types, allowing high-friction users to take advantage of the customization done by others. Another idea for of dynamic filtering is to use the behavior or Slashdot readers who seem to have low friction and active reading patterns, i.e. lead users, to automatically change comment settings.

This research indicates that Slashdot readers have different goals in reading comments, and widely support those goals by leveraging comment ratings. However, some users do not take advantage of comment scores in displaying how comments are viewed, at least in part because of the friction involved in making a change. Using the activities and preferences of those who do make such changes, we can consider design recommendations to enhance the use of ratings beyond the ways they have been used previously. Creating schemas based on ratings use, and predicting preferences using the behaviors of lead users both offer new ways to leverage the feedback provided by ratings systems online.

Chapter 4

Follow the (Slash) dot: Effects of feedback on new members in an online community[14]

Abstract

Many virtual communities involve ongoing discussions, with large numbers of users and established, if implicit rules for participation. As new users enter communities like this, both they and existing members benefit when new users learn the standards for participation. Slashdot is a news and discussion site that has developed a system of distributed moderation to provide feedback about the value of posts on their site. This study examines three explanations for how new users learn to participate in a digital community: learning transfer from previous experiences, observation of other members, and feedback from other members. We find that new user behavior is affected by a combination of their viewing behavior, the moderation feedback they receive, and replies to their comments.

Introduction

According to a recent report by the Pew Internet and American Life Project (Madden and Rainie 2003), 25% of Internet users in the United States participate in online chat rooms or discussions, a number which has grown over recent years. A person entering an online discussion is often joining a mature social system with existing

---

[14] Published in collaboration with Erik Johnston. Lampe, C. and E. Johnston (2005). Follow the (Slash) dot: Effects of Feedback on New Members in an Online Community. International Conference on Supporting Group Work, GROUP '05, Sanibel Island, FL, ACM Press.

members and a developed sense of how members behave. The entrance of new members into established discussions may be both potentially beneficial and potentially harmful to the operation of the online forum. New members may provide additional energy and ideas to persistent digital communities. However, the textual nature of most online discussions, and heterogeneity of online forums across the web, may make it difficult for new members to detect rules for how to behave. When new users have trouble conforming to discussion standards they may increase information overload for the entire community and become more vulnerable to a wide variety of deception from misbehaving users.

Many methods have been used to socialize new users in an online space. Frequently Asked Question (FAQ) documents often explicitly state a community's values and procedures. In some spaces new users must wait before being able to contribute content to ensure that they have time to observe the normal methods and types of participation. Some digital communities have sections allocated specifically for new users, often referred to as "newbie gardens". Online role-playing games often provide spaces like this for new users to learn system commands and interaction standards free of harassment from more experienced players. Another common method of socializing new users is to provide direct mentorship from more experienced community members. A further method of shaping new user behavior is the use of feedback provided by the larger community, often in the form of rating systems that provide evaluations of new contributions.

This research examines the role different mechanisms for learning might have in shaping new user behavior. Are users coming to an online community with all the skills they need to post highly rated comments? Do new users observe others to determine how to write comments? What role does feedback from other members play in shaping new user posting behavior?

Effects as new users join online groups

Online groups that are successful attract new members on an ongoing basis. Internet users interested in online discussion seek out groups that provide the maximum benefit for their investment of time and effort. Joining persistent, large groups makes sense to the new member as they are able to see a wider array of viewpoints, and have their own messages viewed by more people (Hiltz and Turoff 1978; Palme 1995). A new member might observe the site as a passive participant before deciding to submit comments.

Established discussion spaces also benefit from having new members. New participants refresh interest and activity on a site (Powazek 2002). New members can also replace users who have left the site for various reasons, keeping critical mass (Kim 2000). The existence of low barriers to entry and exit in most online discussion sites means that membership remains in a constant state of flux, which has the benefits of eliciting new viewpoints, renewing commitment, and maintaining a healthy size population. However, there are problems that can occur when new members enter established communities.

New members may increase information overload

People have limited ability to perceive and process information (Miller 1962), as well as limited attention spans. The propensity of the online environment to create information overload was discussed by Hiltz and Turoff (Hiltz and Turoff 1985), who recommended designing computer mediated communication (CMC) systems specifically to reduce overload, including such elements as voting, moderation and sanctioning of anonymous members. Jones et al. (Jones, Ravid et al. 2002) examined Usenet newsgroups and found that users are more likely to respond to simpler messages in situations of overload; that users will end participation as overload increases and that users generate simpler responses as overload increases. Previously, Jones and Rafaeli (Jones and Rafaeli 1999) proposed that communication online takes an S-shaped pattern of frequency of occurrence. Early in the existence of a conversation space, or "virtual public" to use their term, there is a struggle to achieve critical mass of people contributing to the conversation. A sharp increase after that critical mass is achieved results in information overload, and communication levels off as participants are discouraged by the rate of messages. Butler (Butler 2001) similarly found that more active listserv's not only had more users entering the discussion, but that they lost users at a greater rate than smaller structures.

New members may violate norms

In offline communities, new members often learn how to behave by following the nonverbal cues of fellow participants, an ability which is often lost in the CMC context (Sproull and Kiesler 1991). Online discussion spaces often have well-developed standards of behavior for how to proceed with conversation, including what constitutes a good post, how often one should post, and how to interact with other members. Violations of these norms can lead to misunderstanding, flame wars and other types of social breakdowns that occur in online communication. Many online spaces have a vocabulary that is specific to that venue (Crystal 2001). Ignorance of special terms and requests for clarification can derail conversation and irritate experienced members. It is often difficult to tell when a breach of etiquette is the result of innocent ignorance from a new user, or willful misbehavior.

The digital community benefits when new members learn rules quickly. Less time needs to be spent by experienced members in explaining terms and expectations (Whittaker, Terveen et al. 1998), which leads to more attention placed on discussion, the central activity of the community.

New members may be vulnerable to deception

Many types of misbehavior that take place in virtual public spheres specifically target new users. “Trolling” is posting a comment designed to trick people into aggravated responses. “The well-constructed troll is a post that induces lots of newbies and flamers to make themselves look even more clueless than they already do, while subtly conveying to the more savvy and experienced that it is in fact a deliberate troll. If you don't fall for the joke, you get to be in on it. (Raymond 2004)”. Trolling often takes advantage of new user naïveté to elicit angry responses.

In Multi-User Dungeons (MUDs) and Massively Multi-Player Online Roleplaying Games (MMORPGs) a category of player specifically targets new players, killing their carefully crafted characters (Reid 1999). In response, these spaces have protected new members by leaving all players immune to harassment or by creating “newbie gardens”, areas where new members can operate safely. This form of protection is not commonly available in other types of digital communities, where new user contributions are often immediately compared with experienced members of the site.

New members may be ignored

If the new member receives no attention from the community, they are likely to abandon the space for not appreciating them appropriately. New members who submit contributions to an online discussion have a hope that other members of the community will value their contribution. New members might expect to feel that their contribution is worthwhile and the discussion is worth their effort if their comment receives attention in the form of replies. Appropriate attention from other community members is likely to lead to future participation.

Variables affecting participation outcomes

To understand how new users learn how to participate appropriately for their roles in online communities we examine several methods for learning how to properly contribute in a new community.

- Previous Experience
  - The new member has skills developed prior to joining the site either through formal education, or participation in similar forums with complementary standards.
- Observation
  - The new member observes successful, experienced members and emulates them.
- Feedback
  - The new member participates by posting a comment and their future contributions are shaped through direct feedback from other members through moderation and discussion.

The importance of previous experience in determining new user participation is based on the theory of learning transfer (Perkins and Salomon 1992), which argues that learning in one context enhances related performance in another context. Observation is one component of the theory of situated learning [14], where new members observe experienced members before starting their own participation. Learning through participation and feedback is grounded in behaviorist theories (Ilgen, Fisher et al. 1979), which claims that people can be shaped by feedback to learn new tasks. We believe that each of these play a unique role in the learning process and by identifying the contribution of each, we will be able to make design recommendations that take advantage of the unique qualities of each of them.

Transfer of learning from other digital communities

It is possible that there is a universal standard for posting in online discussions, and that learning to post valued comments in, say, Usenet groups transfers to Web discussion boards. Several guides of "netiquette" are available (Crystal 2001), and users participate in multiple discussion spaces at the same time (Madden and Rainie 2003). It is unlikely that participants enter each virtual community tabula rasa, but rather transfer skills learned in other fora. This type of near transfer suggests that learning from a separate context enhances the ability to perform in a new context (Perkins and Salomon 1992).

Situated learning through observation of successful participants

Many theories of learning point to the importance of observing others engaged in similar behaviors. In a classic study, Bandura (Bandura 1973) found that children who observed adults attacking a doll were more likely to engage in that behavior than children who did not observe the behavior. In small groups, members often decide how to behave based on the actions of authority members of that group (Forsyth 1999). Lave and Wenger (Lave and Wenger 1993) use the idea of apprenticeship to explain the different structure of learning in communities of practice. Being an apprentice is not the same as being a pupil. Apprenticeship for Lave here is learning as a peripheral participant that is learning to become a member of a community of practice through increasingly involved participation. At the outset, new members participate simply by observing, learning the values and practices of the community *before* attempting to use them. One starts out at the periphery by observing, but as one becomes a more central participant, feedback from other participants becomes increasingly important. In a study of lurkers, Preece et al (Preece, Nonnecke et al. 2004) found that one of the main reasons given for remaining inactive was that the lurkers were observing the group to learn more about it. Many digital communities also include pages of "Frequently Asked Questions" (FAQs) that are intended to provide explicit guidelines on community expectations. Often, standards for behavior are more implicit, and must be discovered by the new participant.

Feedback from experienced users

Finally, there are many ways in which new members of a discussion space may receive direct feedback from experienced participants. Feedback has been shown to affect behavior depending on a variety of factors, including the perceived legitimacy of the feedback presenter, the ability of the recipient to understand the feedback and immediacy of the feedback given (Ilgen, Fisher et al. 1979). Occasionally, an online discussion space will indicate how many times a message has been read. New users may also receive feedback from existing members that indicates their contributions were valued. Two ways of providing feedback are examined in this study. First, in some online communities users provide direct feedback in the form of ratings for contributions, assigning a numerical value to comments or rating a comment up or down from its current ranting. Secondly, replies to a comment indicate to the author of the parent comment that they've not only been read, but that someone was affected enough by their comment to post a reply (Whittaker, Terveen et al. 1998; Smith 2002).

Research questions

To tie together participation outcomes for new users with alternate explanations for how they learn to participate, we generated several research questions, which we then attempt to answer using data from an active online community.

Q1: How do new users behave when they first enter an established online community?

Q1a: How do users react differently to different types of attention from other users?

Q2: Is there a gap between how well new users think they understand valued participation, and how they are actually rated by other community members?

Q2a: Is this potential gap associated with their previous experience?

Q3: How are measures of learning transfer, observation and feedback related to participation outcomes for new users?

Q3a: How is previous experience related to the first contribution a new user makes?

Q3b: How are learning measures related to how comments are valued by other users?

Q3c: How are learning measures related to whether new users post comments, and the rate at which they post?

Methods

We studied users of a popular discussion site that uses a comment rating system to determine the relationship between these three methods of new user learning and initial participation outcomes. This section describes the digital community we studied, data collected from that site, and a description of how the data is being used to describe the associations articulated above.

Slashdot: News for nerds. Stuff that Matters

Slashdot is a news and commentary site dedicated to technology issues, especially open source software. It attracts about a third of a million unique users each day. As stories are posted, users of the site may comment on those stories. Each story typically engenders several hundred comments, with some stories resulting in over a thousand comments. After a comment has been posted, it can then be rated by another user with moderator eligibility.

Slashdot is a large, active digital community with a strongly developed culture. This culture is expressed in special terms used by Slashdot, like calling anonymous posters “anonymous cowards”; in-jokes that Slashdot participants share, and elements of the Slashdot interface being used in the comments themselves, for example a signature file including admonitions that anyone disagrees with the poster is “-5: Wrong”, which references the Slashdot moderation system. The structure of the site has accreted over time to respond to changing user needs. New members entering the site receive feedback from moderators, but it is unclear how that feedback encourages or discourages participation.

Data collection

Data collection included an analysis of server logs of 11,079 new users who made accounts on Slashdot between November 1, 2004 and December 6, 2004. Whenever a Slashdot user loads a page, posts a comment, or rates another's comment, a record of the interaction is kept on the server log. These records are associated with specific users, stories, and times, allowing us to compare interactions between users. Also, logs of user characteristics, including account creation date and reputation score as of December 6, 2004 were gathered.

Slashdot users may participate anonymously on the site without registering. Consequently, some portion of new users identified with this study had experience with the site prior to creating an account. Our assumption is that this is an independent error term, not correlated with any of our independent measures, and consequently omitting it has no effect on our analysis.

Other server information related to new users included logs of all moderations that took place during the study period, including ratings of new user comments. Additionally, we collected data on all comments made during the same time frame to compare new user activity with other users during the same time period.

Besides the log analysis, we conducted surveys with 233 users who had created their accounts since November 1, 2004. Respondents were recruited by an invitation to participate that appeared on Slashdot's homepage in late November. Survey respondents were tracked with a unique identifier that allows responses to be matched with log data. Survey invitations appeared for only a two day period, during which time only 3,341 users identified in the dataset visited the site. The overall response rate for the study was

8%. 109 of the 233 survey respondents had not made any comments since creating their user accounts. We checked responses between commenters and non-commenters and found no significant differences between their responses.

We used the account creation date to identify new users. While it is possible to have multiple accounts, or post without creating an account, the culture of Slashdot discourages both of these behaviors. Since individuals can create multiple accounts on Slashdot, we also matched user IP addresses to see if new users were experienced users with new accounts. We asked in the survey if the user has more than one account on Slashdot. In the server logs we found several instances of multiple accounts from the same IP address which we excluded from the study. No survey respondents reported other accounts or had IP addresses that matched those of other user accounts.

Before examining individual user outcomes, we examined overall participation rates on Slashdot. Out of 11,079 new users selected for study, 1763 users (16%) made 6467 comments. Of new users who commented, 55.1% made only one comment. The maximum was 248 comments. Of those who made any comments, the mean was 3.7, and the median was 1 comment. New users who commented had a median of 28 minutes elapse between the creation of their account and the posting of their first comment.

As has been found in other online settings (Jones 1997), drop out rates of new users are high. To get a sense of how many people abandon their accounts on Slashdot, we studied the 1000 users of the site who had created their accounts between November 1, 2004 and November 10, 2004. Of these users, 5% had visited only one Slashdot page by December 6, 2004. 25% of this sample only looked at 10 or fewer pages. The median number of site pages loaded by a new user was 39, and the mean number was 101. The maximum number of pages loaded by one of these users was 4035, with 27.5% of users viewing more than 100 pages over the study period.

Measuring alternate explanations of learning

As mentioned above, we have focused on three explanations for how new users learn to participate in an environment like Slashdot: learning transfer from previous experience, observation of other members, and feedback from other members.

To approximate previous experiences with commenting, we asked news users questions in the online survey about their experience in online forums, self-rated computer expertise and education level. We asked survey respondents to rate their own experience in discussion sites other than Slashdot on a scale from 1 to 7. On average, respondents scored a 4.3 out of 7 for experience in other sites, with a high percentage in the highest category. We also asked respondents to rate their own expertise with computers. Respondents rated themselves as very expert with computers, an average of 5.97 on a scale of 1 to 7. Survey respondents were also asked about their level of education. Over 50% of respondents reported having a college degree or graduate degree, with an additional 34% claiming some level of college experience.

Observation behavior is captured in two ways. First, the amount of time between when a user creates their account and when they post their first comment is collected. While some exceptions will exist, it could be that the longer the time spent between these two events indicates more time viewing the site. Since the exceptions to this might be significant, we also collected page request information for users. An error in logging prevented the collection of timestamp information on individual page requests, so we only know overall page requests during the study period. Using the overall page requests made by the user over he study period, we coded users into low frequency and high frequency site viewers.

User feedback is captured in two ways. Each comment from our new users begins with the default score of +1, which can change based on moderations from other users. We collected the final scores. Additionally, we mark the number of replies a comment receives as a form of feedback, as it indicates to the new user that their comment was not only read, but regarded enough by other users to engender a reply.

Table 4.1 summarizes the measures of learning described above. Measures in dark gray are associated with learning transfer from previous experiences, those in light gray with observing other users, and those in white with feedback from other users.

Table 4.1: Summary of measures used to detect different types of learning.

Variable	Explanation
Online experience	Survey measure of how much experience users felt they had in online discussion forums.
Computer expertise	Survey measure of how expert the respondent felt they were with computers.
Education level	Survey measure of the last educational degree the respondent received.
Observation time	Time between when a new user creates an account and time they post their first comment.
Hit frequency	How frequently the new user posts page views from the site.
Score	The score a user's comments receive through moderation.
Replies	The number of replies a user's comments receive from other users.

New user participation outcomes

There are several participation outcomes for new users that we use as estimates of their integration into the Slashdot forum. We developed three indicators of desirable new user participation on the site: scores of comments written by the new user, the rate at which comments are posted, and the overall number of comments made.

The scores of comments accumulated through the distributed moderation system act as one measure that the contributor is valued by the community. The higher the average score of comments posted by a user, the more likely that user is valued. For this work, we look at the score of the first comments users make, as well as the average scores of their comment. We look at first comment scores because they show whether the new user starts as a highly valued participant, or if they start being valued little by the community initially, as measured by comment scores, but subsequently improve over time.

The amount of time that a user waits between posts may also indicate desirable participation. In particular, delays between the first comment a user makes and the second may indicate they were turned off by the response to the first comment, while less delay may indicate that the user was drawn into the system.

Another measure of successful participation is the number of comments they post during the study period. It is beneficial to have members post comments, although too many posts may indicate problematic participation.

Table 4.2 summarizes measures of participation outcomes for new Slashdot users included in this study.

Table 4.2: Summary of measures of participation outcomes.

Measure	Explanation
First score	Score that a new user's first comment receives through moderation.
Probability of second comment	The probability that a user will post a second comment after having posted a first.
Time to post second comment	Time it takes a user to post a second comment after posting a first
Score change	Difference between the scores a new user's first and second comments receive.

Number of comments	Total number of comments a user made over the study period.

Results

The results of our analysis of new user behavior are structured to attempt to answer the research questions raised earlier. The first section describes initial user contributions, and how participation differs among users who receive different scores on their first comments. The second section addresses user beliefs about their ability to create valued messages and how their actual scores do or do not match that perception. The third section analyzes participation outcomes in terms of the measures of alternative learning methods described above.

How do new users behave when they first come to Slashdot?

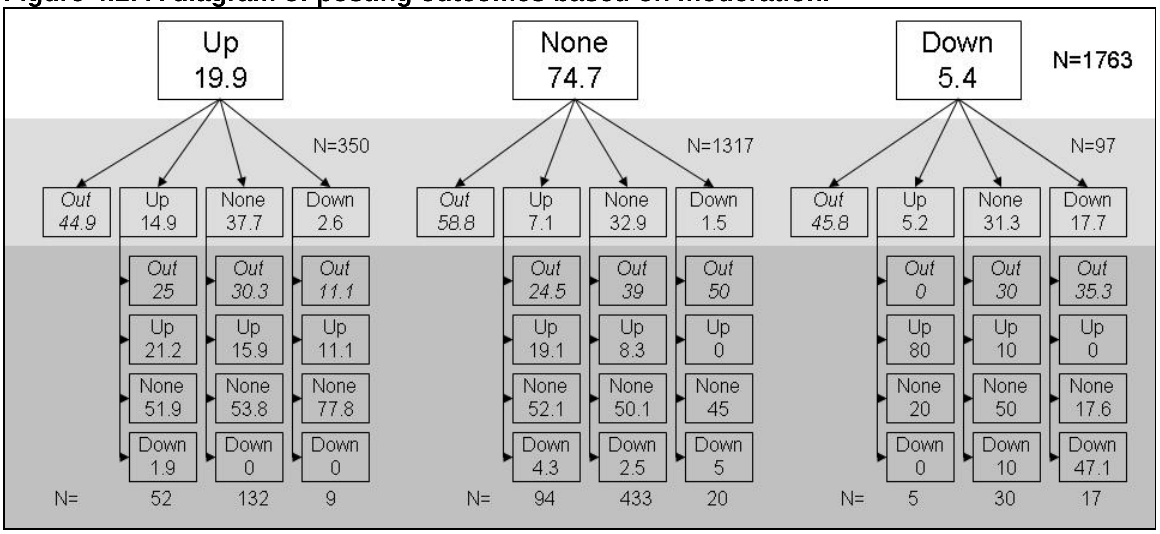
To describe more fully how initial moderation affects participation, we examined posting patterns for the first three comments made by users. How many users start with negative ratings, yet go on to future success? Which types of moderation are associated with users ceasing to post comments? Figure 4.1 shows the moderation outcome of the first three comments new users made in the dataset being studied. In the chart, the possible outcomes are "Up" for when a user receives positive feedback through the rating system, "None" for when the moderation receives no ratings, "Down" for when the comment receives negative feedback through moderation and "Out" for when the user does not make additional comments.

New users who received no moderation were less likely to make a second comment than users who received either positive or negative initial feedback through moderation. Even when a user receives feedback on their first comment, lack of feedback on the second is associated with approximately 30% of users to ceasing commenting.

There is some indication that receiving two negative moderations in a row make it unlikely that a user will receive a positive moderation. Though the numbers in this analysis are too low to be certain of the pattern, each path followed to the 5th comment show no occasions where two down ratings were followed by a future up rating. Another interesting pattern is the recovery of users whose first comment was rated negatively. In those cases where the second comment received positive rating, 4 out of the 5 cases where there were third comments were also rated positively, and none of them received a negative rating on their third comment. However, there is also a finding that some people that receive initial negative feedback continue to make comments that are commented negatively. This propensity increases at each level suggesting that negative feedback is the goal of some users, or alternatively that the user pool at this level only contains those contributors who are unable to write valued comments.

These descriptions of moderation outcomes indicate that some change is happening to users between the times they post comments. The direction of moderation does seem to have some relationship with how future comments will be rated, but with this data it is unwise to make a strong causal claim. It does seem possible given these patterns that Slashdot participants are using the moderation system as a sounding board to craft their future comments. This is consistent with Goffman's argument (Goffman 1963) that individuals acting in a public place are "performing" for some perceived audience. In the Slashdot case, users might be using the feedback provided through moderation to adjust their performance.

Figure 4.1: A diagram of posting outcomes based on moderation.



New users believe they can detect a good comment.

One sign that new members of Slashdot are detecting the values of a discussion space is that they agree about what makes a comment highly valued. In Lampe and Resnick (Lampe and Resnick 2004) we found that moderators on the site widely agree on what constitutes a good comment, as measured by their agreement on the score of a comment. To analyze whether new users similarly agree that they can detect a good comment, we asked several survey questions about their confidence in their ability to detect a highly rated comment.

Most new Slashdot members felt that they could readily detect expectations for posting a good comment. Means reported below are on a scale of 1-5, where 1 indicates low agreement and 5 indicates high agreement. In four questions related to confidence in their ability to detect valued comments on the site, users strongly agreed that they knew what a good comment was (mean=3.95, C.I. $3.83 \leq x \leq 4.07$), could tell why a comment had received the score it did (mean=3.72, C.I. $3.59 \leq x \leq 3.85$), felt the expectations for writing a good comment were clear (mean=3.43, C.I. $3.29 \leq x \leq 3.58$) and could write a comment that would be highly scored (mean=3.66, C.I. $3.51 \leq x \leq 3.81$). In addition, the values for the purpose of the moderation system were strongly shared amongst new users, who agreed strongly that moderation should be used to promote well written contributions (mean=4.24, C.I. $4.12 \leq x \leq 4.36$) as opposed to supporting particular viewpoints (mean=1.96, C.I. $1.80 \leq x \leq 2.12$).

New members also strongly agreed when asked if discussion on Slashdot was worthwhile compared to other sites (mean=4.10, C.I. $3.97 \leq x \leq 4.23$), and that the moderation system is important in fostering discussion on Slashdot (mean=4.19, C.I. $4.05 \leq x \leq 4.32$). The strong agreement on the questions reported above seems to indicate that new members of Slashdot believe they know what constitutes a good comment.

To determine whether a user's impressions of how well they understand how to write a valued comment is related to their eventual ability to write a valued comment, we used the Spearman's rho statistic to correlate survey responses with the average score of comments a user received. Spearman's rho is a measure of relationship between ordinal variables, and consequently a better choice here than the more common Pearson's correlation statistic. The average score users' comments received was poorly correlated with whether the user reported they were confident they could write a comment that would be rated highly ($r=0.14$, $p<0.15$, $n=113$), whether they felt the expectations for highly rated comments was clear ($r=0.09$, $p<0.32$, $n=114$), whether they felt they understood why a comment received the score it did ($r=0.10$, $p<0.29$, $n=120$) or how concerned the user was with the scores their comments receive ($r=0.03$, $p<0.72$, $n=120$).

Even though Slashdot users widely agreed that they knew what constituted a highly rated comment on the site, that belief does not seem to be associated with actual production of highly rated comments. This could be because Slashdot users actually have very different opinions about what would constitute a highly rated comment, and only believe that other agree with them. It could also be that some users are creating comments they know are not going to be highly rated.

Mechanisms that affect user contributions

In this section we examine first comment scores, whether and how quickly a second comment is posted, total number of comments posted and average ratings of comments posted by a user. For each of these user outcomes, we look at our operationalizations of user learning to determine effects.

First comment scores

When a new user creates their first comment, they have not had the opportunity to benefit from direct community feedback, but they may be affected by learning they bring from other experiences, or by observing the site prior to posting a first comment

Table 4.3 reports an ordinary least squares regression predicting the initial score a comment will receive based on measures of previous experience and observation. This model shows that the measures of previous experience and observation of other users poorly predict the how the first comment made by a new user on the site will be rated by others. It could be that the measures of experience and observation do not adequately represent their real values. Another explanation is that users entering Slashdot for the first time share many characteristics, including ability to write comments with little difference between users.

Table 4.3: Ordinary least squares regression predicting score of first user comment.

R-square	0.06		
df	5,61		
	Coef.	**t**	**P>|t|**
Constant	.826	0.753	.454
Forum experience	-0.055	-0.560	.578
Computer expertise	0.066	0.364	.717
Education level	0.173	1.207	.232
Page views	0.000	0.996	.323
Observation time	0.000	-0.866	.390

To assess possible explanations for these findings, we also analyzed the different independent variables separately with the score of the first comment. We did not find any significant relationships between any of the independent variables individually and the score of a user's first comment.

Likelihood of posting a second comment

Many things may happen after a user has posted their first comment. The comment may be rated positively negatively by other members of the discussion board. Some users may decide to reply to the comment. For some users, their first comment may be completely ignored.

As mentioned above, 55.1% of new users on Slashdot made only one comment to the site during the study period. As shown in Figure 4.1, what happens to the first comment a user makes seems to have some relationship with whether they will post a second comment or not. What predicts whether the new user will post a second comment? Of the different outcomes for a comment, do any of them predict that a second comment will follow?

Table 4.4 reports a logistic regression predicting the binary outcome of whether a new user will post a second comment: positive coefficients indicate higher probabilities. Whether the first comment a new user makes is replied to, and whether that comment was rated, or not, by the moderation system appear to be poor predictors of whether a user will post a second comment. Time between creating an account and posting the first comment, as well as how heavily the user requested page views were better predictors of the likelihood of posting a second comment.

Table 4.4: Logistic regression predicting if a new user will post a second question.

Pseudo R-squared	0.14		
n	1704		
	Coef.	**Z**	**P>|z|**
1st com replied to	-0.563	1.32	.251
1st comment was rated	-0.198	0.97	.325
Time between acct creation and first post	0.001	9.21	.002
Frequency of page views	0.005	37.29	.001
Constant	.720	2.08	.001

However, the R-squared value of this model indicates that many important factors that predict posting a second comment are not being accounted for.

Time to post second comment

The gap in time between when a user posts their first comment and when they post their second may be an important indicator of socialization. If the user has a negative first experience, it may take them longer to post again. Consequently, even if feedback does not affect whether a user will post a second comment, it might affect how they do so.

The average time between first and second post for those users who made a second comment was 2.6 days, and the median time was 5.7 hours. This disparity is caused by some outlying users who had large amounts of time between their first and second posts.

Time to post a second comment was not strongly correlated with online forum experience (r=-0.20, n=88), computer experience (r=-0.18, n=87) or education level (r=-0.13, n=87). For measures of observation, how frequently the user requests page views was not related to time lag between first and second comments (r=0.04, n=392). The amount of time users spent observing the site before posting their first comment also had only a weak relationship between the time difference in posting first and second comments (r=0.09, n=780).

Table 4.5 shows measures of user feedback in terms of whether the first comment a new user posted was rated up, down or ignored, as well as whether that comment was replied to. Using these different first comment outcomes as separate groups, we measure the median time difference between posting the first and second comments in terms of different first comment outcomes.

Table 4.5: Time to post second comment based on first comment outcomes.

First comment outcome	Median time to post 2nd comment
First comment rated down	50 minutes
First comment not rated	9.8 hours
First comment rated up	4.2 hours
First comment replied to	1.4 hours
First comment not replied to	7.1 hours

Users who received no rating on their first comments also took much longer to post a second comment. Users who received a negative rating on their first comment were the quickest to post a second comment. This could result from several factors. One might be that negative attention in the form of negative ratings causes users to want to prove themselves as positive contributors. It could also be that some of these users are writing inflammatory content, which they post more often and is rated negatively.

Table 4.6 shows the multivariate explanations that predict how many minutes pass between posting a first and second comment. Survey measures are excluded from the model as they were found to be unassociated with time in the univariate analysis. The amount of time between account creation and posting a first comment minorly reduces the time to post a second comment. The major factor seems to be whether the first comment a new user writes is rated up, which reduces the overall time to post a second comment.

Table 4.6: Ordinary least squares regression for time in minutes to post a second comment.

R-square		0.06	
df		5, 386	
	Coef.	**t**	**P>|t|**
Constant	7787.90	12.064	.001
Observation time	-0.002	-2.580	.010
Page views	-1.977	-1.327	.185
First comment up	-3939.80	-3.639	.001
First com. down	-1983.14	-0.974	.331
First com. replied	-1394.07	-0.719	.472

Users who received replies to their first comment took less than a third of the time than those who did not to post a second comment, though this difference did not appear to be a significant factor in the regression model. The strong difference in users who are initially rated down was also revised by the multivariate analysis; it does not look like being rated down on a first comment affects time to post a second comment when considered with other factors. It does seem that positive feedback from other users through ratings does reduce the overall time to post a second comment. However, this model also has a low r-squared value, indicating that many variables important to this measure are not included. It is likely that one of these missing variables would help explain the difference in findings about the importance of replies in the model versus the univariate analysis.

Score of second comment

Another measure that a difference nonrandomly occurs between a user's first and second comments to the site is the difference in scores between the first comment posted by the user and the second.

Table 4.7 shows an ordinary least squares regression predicting the change in score from the first comment to the second. This model shows that if the first comment receives a positive result, it has a negative effect on the score of the second comment, and vice versa if the first comment received a negative score. Whether the first comment was replied to does not seem to be a factor, though number of page views has a small, but significant effect on the score of the second comment. The findings for the reverse roles of initial up and down may be a result of regression towards the mean, i.e. that initial scores have an element of randomness, and second scores correct the arbitrary high and low scores of the first comment. This is supported by the correlation between the scores of the first and second comments a user makes ($r=0.174$, $p<.001$, $n=792$).

Table 4.7: Ordinary least squares regression predicting score of the second comment.

R-square	0.34		
df	5, 386		
	Coef.	**t**	**P>|t|**
Constant	0.262	3.009	.003
Observation time	0.000	0.412	.681
Page views	0.000	2.347	.019
First comment up	-1.790	-12.155	.001
First com. down	1.286	4.670	.001
First com. replied	-0.062	-0/238	.812

Number of comments posted

Although it can be dangerous to have too many comments posted, having new users create comments is a measure that they are involved in the digital community. Consequently, we looked at the overall number of comments new users made during the study period, and which factors were associated with frequency of posting.

New users made an average of 26 comments over the study period, with a median of 9 comments. 15% of new users only made one comment, and one user was responsible for 248 comments.

The overall number of comments posted was not strongly correlated with online forum experience (r=0.17, n=124), computer experience (r=0.10, n=122) or education level (r=-0.06, n=121). However, the overall number of comments was relatively strongly correlated to how frequently the user requested page views (r=0.52, n=69) and the time lapse between their account creation and when they posted their first comment (r=0.38, n=124). This could mean that users who read the site more often are more willing to post comments. A design implication of this might be that site administrators could detect frequent viewers early on, and mark them as potentially valuable participants early in their tenure on the site.

Table 4.8 presents an ordinary least squares regression model predicting the overall number of comments a user will make. Since measures of experience are unrelated to total comments in univariate analysis, they are excluded from the model.

In the model, observation time and frequency of page views are both related to the overall number of comments a user makes, whereas the direction of initial moderation and the first reply a comment receives are not. One could imagine an active user who both comments and reads frequently, but is no better at writing comments than any other user. Other work has shown that often a few authors are responsible for the majority of messages in an online conversation forum (Viegas and Smith 2004). This seems to present a consistent view of the role between reading activity and posting.

Table 4.8: Ordinary least squares regression predicting number of comments.

R-square	0.23		
df	5, 386		
	Coef.	**t**	**P>|t|**
Constant	4.551	4.176	.001
Observation time	-0.003	-2.741	.006
Page views	0.026	10.252	.001
First comment up	1.760	0.955	.340
First com. down	-1.585	-0.460	.646
First com. replied	-5.019	-1.537	.125

Discussion

New members entering an existing online community face a complicated, often overwhelming environment where it is hard to know how to act. Slashdot is an unusual digital community, with a distinct, techno-centric culture and design that has accreted over a long time. We feel this makes Slashdot an especially interesting case study in that new members of this space are likely to have an especially hard time learning the standards of practice for posting comments.

Previous experience did not seem to have a relationship with how highly rated a user's first comment becomes. Although Slashdot users report they have high levels of experience in other online discussion forums, that experience does not seem to translate into automatic success on the site.

New users felt they could write a comment that would be highly valued, but when their comments were rated there was no relationship between that belief and the actual score of the comment. This could mean that there is high variability in what users see as "valued" comments.

The patterns of moderation outcomes for the first three comments new users contribute suggest intriguing implications. It is clear that many users choose not to continue contributing comments after their first, though the reason for this is less clear. Being rated up on a first comment does not seem to affect whether a user will post again or not, but it does seem to affect how quickly a second post will happen. Active users as measured by time spent observing the site before commenting and frequency of page views are more likely to post a second comments and more comments overall.

The high rates of drop out among new members points to an alternate use of feedback through the rating system, namely encouraging users to select themselves out of the population. In formulating this research, we focused on learning *how* to participate in a way that will be valued. It could be that ratings provide a way to determine *whether* to continue participation on the site.

There is some indication from the change is scores from first comments to second that initial scores might be random, and future comments regress towards a mean score. If this is true, then there are serious implications for how rating systems might affect new users. New users who receive undeserved, somewhat random negative attention might prematurely drop out of a discussion forum.

Limitations and future work

This study is an initial examination of new user behavior in a persistent digital community, and is largely intended only to describe different participation outcomes. We look at the relationship between different user variables and how they participate on the site. By describing this case we are hoping to motivate continued examination of how new users become socialized when entering persistent digital communities.

This work depends on findings from one case: Slashdot. This site is an exceptional case in many ways, being one of the few online discussion forums to use distributed moderation, and having developed their structure over many years. We feel, for example, that the general finding that users who receive attention from experienced members will participate differently can be generalized to a wider variety of digital communities. In the Slashdot case, attention came in the form of ratings and replies from other members, while in the section above we mention other possible forms of attention site designers could use. We hope that rather than being an exception to other online interactions, the Slashdot case provides a leading example of how interactions might be shaped with different elements of observation and user feedback.

In the next stage of this work it is essential to more directly address causality. One experiment in consideration is to randomly assign new users on Slashdot into groups where they receive controlled amounts of feedback and measuring how their future participation differs.

Another important data collection effort for future work will be more open-ended interactions with Slashdot users, either through more sophisticated survey work, or through interviews. For example, many Slashdot users never create an account, and for those who do it is unclear why they choose to do so. If a Slashdot user participates anonymously for a year and then creates an account, it is wrong to say they are a “new” user. More qualitative work might be able to address this issue where server log analysis has not.

Although the survey provided interesting insight into the beliefs of the new Slashdot members, future work with this population should include additional measures of digital community experiences. Additional questions related to frequency and depth of participation in other forums would be valuable new information. Future work should also include more specific information about the specific pages being accessed by new users.

Conclusion

The findings from this study indicate that participation outcomes for new users are affected by a mix of previous experiences, observation of other members, and feedback received through ratings and replies. Each play distinct roles in whether and how a new user will participate on the site, and how that participation will be viewed by the larger community.

These findings are potentially important for designers of digital communities who need to plan for incorporating new members into their ongoing social structures. Since page loads were associated with valued contributions, designers could require a certain number of visits to the site before allowing posting. Initial ratings, either up or down, were associated with future posting activity, so sites might develop rules different for rating new users than for regular contributors.

Chapter 5

Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space[15]

Abstract

Can a system of distributed moderation quickly and consistently separate high and low quality comments in an online conversation? Analysis of the site Slashdot.org suggests that the answer is a qualified yes, but that important challenges remain for designers of such systems. Thousands of users act as moderators. Final scores for comments are reasonably dispersed and the community generally agrees that moderations are fair. On the other hand, posting happens quickly and much of a conversation can pass before the best and worst comments are identified. Of those moderations that were judged unfair, only about half were subsequently counterbalanced by a moderation in the other direction. And comments with low scores, not at top-level, or posted late in a conversation were more likely to be overlooked by moderators.

Introduction

Participants in online conversations have diverse goals. Some readers want to be informed, some to be amused. Some posters want to inform or amuse, some want to compete, and others want merely to be noticed.

[15] This work published with Paul Resnick (Lampe and Resnick 2004)Lampe, C. and P. Resnick (2004). Slash(dot) and burn: distributed moderation in a large online conversation space. Conference on Human Factors in Computing Systems (CHI), Vienna, Austria, ACM Press.

In conversation spaces with limited access and few participants, individuals can allocate their attention and informal social mechanisms can reduce disruptive behavior. In conversational spaces with low entry barriers and hundreds or thousands of participants, governance is more problematic (Kollock and Smith 1996). Such colorful expressions as trolling, flaming, spamming, and flooding have emerged to describe behaviors that benefit some people while disrupting others' ability to get what they want from a conversational space (Whittaker, Terveen et al. 1998; Pfaffenberger 2002). Even absent deliberately disruptive behavior, too many postings can lead to information overload. More participants in conversation spaces is empirically correlated with more turnover of participation (Butler 1999; Jones, Ravid et al. 2002), one indicator of user dissatisfaction.

Various methods have been used to limit the disruption that anti-social behavior can cause, and to help readers cope with information overload. Properties of messages (e.g., length) or their contents (e.g., shared word usage with other messages (Sack 2000)) can be identified automatically. Individual or group kill files can be created to censor particular authors or properties of message authors (e.g., frequency of posting or frequency of being responded to) can be calculated automatically and used to classify messages (Smith and Fiore 1991).

The judgments of other people, however, are often the best indicator of which messages are worth attending to. In small to medium size conversations, an individual can act as moderator, screening all candidate messages. This gives the moderator a lot of power, more than other participants are comfortable with in some situations. Moreover, a single moderator, or even a small team of moderators, simply can't keep up if there are too many messages to evaluate.

Beginning with the Tapestry system (Goldberg, Nichols et al. 1992), researchers and developers have explored ways to collect and use the judgments of the general readership rather than just a few designated leaders. These distributed moderation systems have only recently been deployed in large scale conversation spaces. There has been little opportunity to evaluate how well they function at classifying posts, how those classifications affect reader behavior, and how they affect posting behavior.

This paper focuses on only the moderation process itself. Even leaving aside questions of how moderation impacts readers and writers, fundamental questions remain. The most fundamental is whether shared norms can emerge about what constitutes a good or bad post, with most moderators following those norms most of the time, or whether tastes differ in fundamental ways, so that more personalized recommendations need to be made, using collaborative filtering techniques (Resnick, Iacovou et al. 1994; Shardanand and Maes 1995; Terveen and Hill 2002).

A theoretical investigation of incentives for provision of evaluations (Avery, Resnick et al. 1999) described several potential problems. One is underprovision. Some or all posts may get insufficient attention from moderators, or there could be long delays from the time a comment is posted until it is moderated. Another potential problem is premature negative consensus. Messages that receive early negative moderation might get insufficient attention from other moderators, and thus moderation mistakes would not be corrected.

Slashdot presents a unique opportunity to investigate empirically how distributed moderation plays out in practice. The site has honed its moderation system over several years and norms of usage have had plenty of time to develop. Thus, remaining problems should reflect subtle issues that are not immediately apparent or fundamental problems for which there is no easy fix.

Methods

We analyzed usage logs for the period extending from May 31, 2003 through July 30, 2003. The logs included information for each comment, moderation and meta-moderation that took place. User data included the karma scores of users and whether they were regular users or paid editors. The dataset includes 293,608 moderations, 489,948 comments, and 1,576,937 meta-moderations.

Our primary method of inquiry was to look for patterns in the usage logs. Because there are so many observations in our datasets, the differences we report are all strongly statistically significant, and we omit reporting measures of significance in most cases. We also conducted interviews with three Slashdot editors, reviewing early findings and asking for clarification and explication of certain phenomena.

We begin with summary statistics about levels of participation in the moderation and meta-moderation systems and the distribution of scores for comments. Next, we examine whether there was a community consensus about what constitutes a good or bad comment. Third, we examine how long it took to identify good and bad comments. Fourth, we examine whether moderations judged to be unfair by meta-moderators were corrected with subsequent moderations. Finally, we investigate whether there are some types of messages that receive unfair treatment or insufficient attention from moderators.

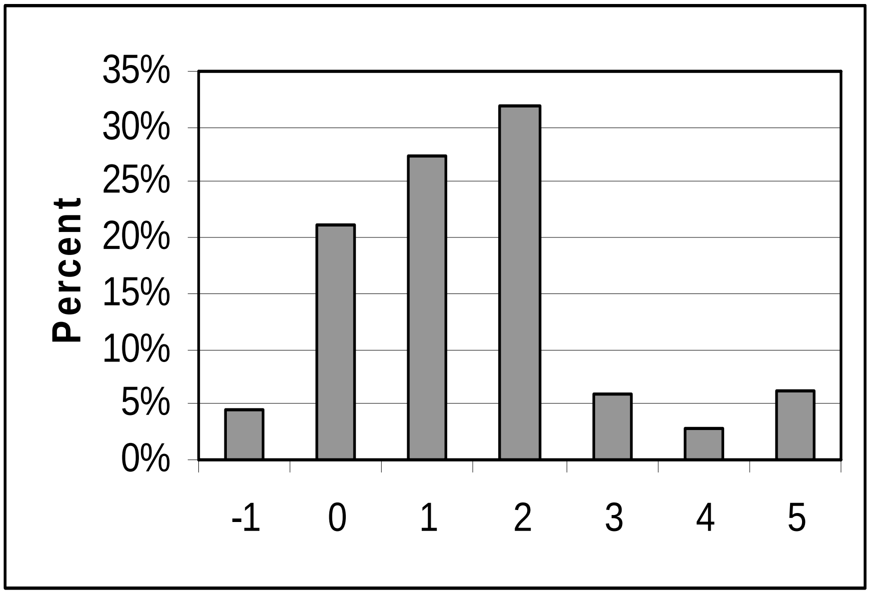
Participation Levels and Outcomes

There is widespread participation in the moderation and meta-moderation systems. 24,069 distinct users moderated during the two month period and the median number of moderations per moderator was 7 (mean 13). Because the system deliberately limited the amount of moderation any individual can perform, the maximum number of moderations completed by anyone other than paid staff was 164, less than three per day. Paid staff, who have unlimited moderator points, accounted for only 2.4% of the total moderations. 18,799 distinct users meta-moderated and the median per person was 25 (mean 84).

There is a partial but not complete overlap between moderators and posters. Of users who commented, 41% also moderated. Of moderators, 68% also commented while 32% (nearly 8000 users) were lurkers who never posted during the two month period. Participation overlap between commenting and metamoderation was similar, but somewhat lower. Of users who commented, 31% also meta-moderated. Of those who meta-moderated, 66% also commented.

During the study period, 28% of comments received at least one moderation during the study period. Of those that did, 48% received only one moderation. The highest number of moderations on a comment during this study period was 51, though historically there have been rare comments that have received over a hundred. In keeping with the stated guidelines, the overwhelming majority of moderations, 79%, were positive.
There was a reasonable dispersion of final scores, as shown in Figure 5.1. About one in four comments finished with a score of –1 or 0, about one in ten with a score of 4 or 5.

Figure 5.1: Distribution of final comment scores.



Reaching consensus

Is there a community consensus about which comments should receive up and down moderations? One indicator of disagreement would be the frequency of comments receiving both positive and negative moderations. Among comments that received moderation, 65% received only positive moderation, 20% only negative, and 15% received both.

Metamoderations provide a more direct indicator of the extent of community consensus about norms for moderation. 92% of all metamoderations indicated agreement with the moderations they evaluated. The rate was even higher for positive moderations, 94%. There was less consensus, however, about negative moderations, with only 77% agreement from meta-moderators.

While most users seem to diverge occasionally from total community consensus, true "rebel" moderators were rare. Only 14% of moderators were never metamoderated as unfair, but 72% of moderators received more than 5/6 “fair” metamoderations. For 453 moderators, about 2% of the pool, more than half the metamoderations disagreed with the direction of their moderations.

Moderation Delays

A comment is eligible for moderation for up to two weeks after it is posted. A major purpose of the distributed moderation system, however, is to help readers allocate their attention. For that reason, it is desirable for moderation to occur as quickly as possible.

We do not have data on the distribution of elapsed time from comment posting to reading. However, to get a sense of the time scale of conversations, we computed each story’s “half-conversation life”, the elapsed time until half of the total comments on the story were posted. The median half-conversation life among stories was 174 minutes, or just under three hours. The median time for a story to accumulate 90% of its comments was 1060 minutes, or about eighteen hours.

Among comments that received some moderation, the median time until receiving the first moderation was 83 minutes. Perhaps a more useful metric is how much time elapsed before a moderation first pushed a comment to a score of +4 or down to 0 or –1, as shown in Table 1. More than 40% of comments that reached a +4 score took longer to do so than 174 minutes, the time at which a typical conversation was already half over. More than 20% of the comments that were downgraded to 0 or –1 took at least that long. (Merely starting with a score of 0 or –1, without receiving a negative moderation, did not count as being downgraded in this timing analysis.)

Table 5.1: Time to reach benchmark scores.

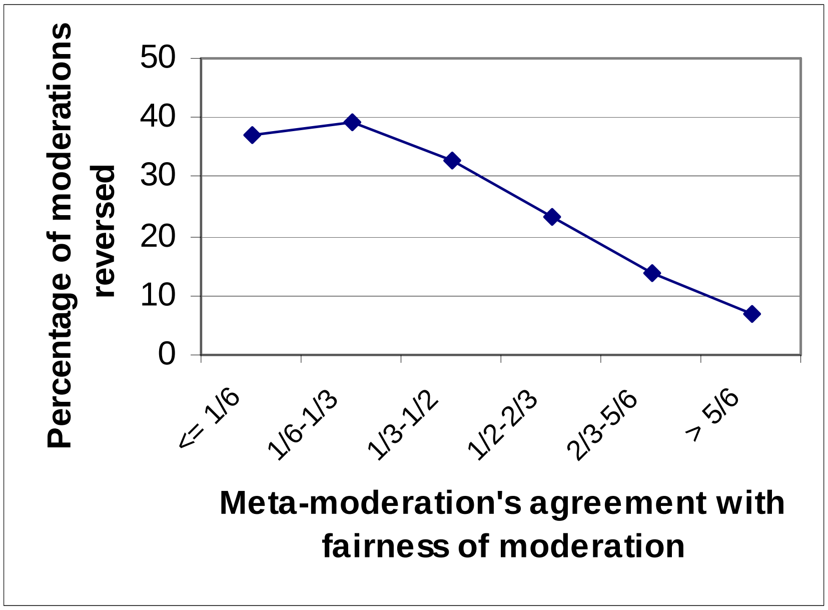
Percentile	Time in minutes to reach a score >= 4 (n=47,474)	Time in minutes to reach a score <=0 (n=28,277)
10	19	2
20	37	5
30	61	9
40	96	16
50	148	28
60	227	49
70	350	90
80	554	190
90	932	517

Reversing Unfair Moderations

We have already seen that most moderations conform to community standards, as expressed through the meta-moderation system. Ideally, after an incorrect negative moderation, someone else would moderate the comment positively, and vice versa. We call this a moderation reversal.

In practice, less fair moderations were more likely to be reversed, as shown in Figure 5.2. However, even moderations that all or almost all meta-moderations disagreed with were reversed less than half the time. Unfair positive moderations (as judged by at least 2/3 of the meta-moderators) were reversed 34% of the time, and unfair negative moderations were reversed 40% of the time.

Figure 5.2: Moderation reversals



Buried treasures

Theories from information economics suggest two reasons why comments of equal quality may not end up with equal scores through the moderation process. First, some comments may get less attention from moderators, so there is less chance that they will be moved from their current scores (Avery, Resnick et al. 1999). Second, there may be a herding or information cascade effect, where moderators are influenced by previous moderations either to remain silent or to contribute another moderation in the same direction (Bikhchandani, Hirshleifer et al. 1989; Banerjee 1992).

Either insufficient attention or information cascades could result in buried treasures, comments that should have high scores but do not. The previous section's results on low reversal rates suggest that incorrect moderations did cause some treasures to be buried (and some trash to be surfaced). Systematic biases that make some types of comments more likely to be buried would be even more troubling.

Moderators may give insufficient attention to comments with low scores, response comments (as opposed to top-level comments that start new threads), or comments added later in the conversation. Though moderators are encouraged to scan all comments, they can use viewing thresholds in the same way as other readers, so that lower-scoring comments would be hidden and responses would need higher scores than top-level comments would need to be visible. And if moderators look through all the comments posted so far and some moderators read early in the conversation, the early posts will be looked at by more moderators than will later posts.

In fact, comments with lower starting scores were less likely to be moderated. For example, 30% of comments starting at 2 received a moderation, compared to only 29% of those starting at 1, 25% of those starting at 0, and 9% of those starting at -1. Table 5.2, which compares initial to final scores, shows that comments that started with higher scores tended to finish with higher scores.

Of top-level comments, 48% received some moderation, compared to 22% for response comments. The mean final score for top-level comments was 1.73, as compared to 1.40 for responses.

Finally, comments posted later fared less well in the moderation process. We categorized comments into quintiles: the first fifth of comments on each story are classified as early, the last fifth as late. Of early comments, 59% were moderated, compared to 25% for comments in the middle of the conversation and 7% for late comments. The mean final score for early comments was 1.77, compared to 1.46 for comments in the middle of the conversation and 1.24 for late comments.

Of course, the lower probability of moderation and lower final scores do not necessarily imply problems of insufficient attention from moderators or information cascades. Instead, they may correctly indicate lower quality or less valued messages. For example, late comments may be less likely to contribute new ideas to a conversation. Below we describe three potential confounds, characteristics of comments or the people that posted them that may be the true cause of moderation differences and that may be correlated with the starting score, with whether a comment is at top-level, and with whether a comment comes late in a conversation. Table 5.2 shows correlations among the variables of interest. We then controlled for the potential confounds in regression analyses.

Table 5.2: Initial and final comment scores

		Ending score							Total
		-1	0	1	2	3	4	5	
Starting score	-1	93.4%	3.8%	1.2%	.6%	.4%	.2%	.4%	
	0	13.3%	76.3%	5.9%	1.9%	.8%	.6%	1.3%	
	1	2.0%	2.9%	72.6%	11.0%	4.1%	2.4%	4.9%	
	2	0.00%	0.00%	2.1%	71.0%	11.2%	4.9%	10.8%	
Number of comments		21,753	107,169	265,800	42,379	17,417	19,518	15,912	489,948

Anonymous Posts

The first potential confound is whether the poster chose to remain anonymous. Research on anonymous posting indicates that the higher the anonymity of the user, the more likely their contribution is to have lower value. This lower value can be expressed as off-topic, flaming behavior, or in lower quality submissions (Friedman and Resnick 1997). Anonymous posting is correlated with lower starting scores at Slashdot, since all anonymous posts start with a score of 0. As shown in Table 5.3, anonymous posts were more likely to be responses rather than at top-level, but they were less likely to come late in a conversation.

Table 5.3: Correlations of characteristics and outcomes

	Mod'd	s-score	F-score	Karm	Short com't	Long com't	Anon user	Top level	Early	Late
Modded	1.00									
Starting Score	0.05	1.00								
Final score	0.43	0.69	1.00							
Karma	0.06	0.91	0.64	1.00						
Short comment	-0.01	-0.18	-0.18	-0.17	1.00					
Long comment	0.09	0.09	0.12	0.09	-0.11	1.00				
Anonymous user	-0.05	-0.80	-0.58	-0.84	0.17	-0.07	1.00			
Top level	0.25	-0.03	0.10	-0.04	0.02	0.03	-0.03	1.00		
Early in conversation	0.34	-0.07	0.10	-0.05	0.07	-0.07	0.05	0.32	1.00	
Late in conversation	-0.24	0.04	-0.08	0.03	-0.02	0.06	-0.04	-0.13	-0.25	1.00

Karma Score

The second potential confound is the poster's karma level. Posters with higher karma may be more skilled writers, or better understand and follow the community's norms. Comments from users with higher karma start with higher scores. However, as the correlations in Table 5.3 show, users with higher karma were somewhat less likely to post at top-level or to post early in a conversation.

Comment length

Grice's maxims for optimal messages (Grice 1969) indicate that messages should be long enough to be informative, but not so long as to violate conversational expectations. Thus, exceptionally short or long messages may generally be judged to be of lower quality. In our dataset, the shortest 10% of messages (which we refer to as "very short messages) had fewer than 65 characters and the longest 10% ("very long messages) had more than

1089 characters. As the correlations in Table 3 show, very long comments were more frequent later in threads and very short comments had lower starting scores. Other correlations, however, were not consistent with message length being a confound: both short and long messages were more frequent at top-level than were medium length messages.

Tables 5.4 and 5.5 show that starting score, top-level posting, and late posting had an impact on moderation, even controlling for the potential confounds identified. Table 5.4 reports an ordinary least squares regression predicting the final score: positive coefficients indicate higher predicted scores. Table 5.5 reports a logistic regression predicting the binary outcome of whether a comment will be moderated: positive coefficients indicate higher probabilities. All the coefficients show that top-level comments, early comments, and comments with higher starting scores were more likely to receive moderation and to get higher final scores, even when controlling for the potential confounds.

Table 5.4: Ordinary least squares regression predicting final comment scores.

R-squared	0.52		
	Coef.	**t**	**P>|t|**
Starting score	1.080	259.68	.001
Karma	0.002	20.44	.001
Long comment	0.267	56.90	.001
Short comment	-0.290	-61.08	.001
Top level	0.234	67.71	.001
Early comment	0.416	109.91	.001
Late comment	-0.266	-73.81	.001
Constant	0.157	31.70	.001

The R-squared measure of fit for the predictions of final score was .52, suggesting that there are differences among comments that are important to moderation outcomes but are not captured by the variables in the regression model. Perhaps comments with low starting scores, not at top-level, or posted late in a conversation really are of lower quality, but that quality was not captured by the confounds identified above. Two further tests, however, suggest that that this is not the complete explanation, and that there is a problem of insufficient moderator attention to these comments.

Table 5.5: Logistic regression predicting if a comment will be moderated.

Pseudo R-squared	0.16		
	Coef.	**Z**	**P>|z|**
Starting score	0.043	4.14	.001
Karma	0.007	23.98	.001
Long comment	0.856	76.78	.001
Short comment	-0.119	-9.75	.001
Anonymous user	0.167	10.58	.001
Top level	0.789	99.28	.001
Early comment	1.324	158.86	.001
Late comment	-1.596	-115.59	.001
Constant	-1.604	-127.95	.001

First, we consider the delay until receiving the first moderation for a comment. Since this measure considers only comments that do receive moderation, it should be independent of the quality of the comments and reflect only the amount of attention from moderators. Table 5.6 shows that comments with higher starting scores received moderations sooner. Comments at top-level also received moderation sooner (median time to first moderation 46 minutes vs. 120). Comments early in a conversation also were moderated sooner (median time to first moderation 22 minutes for early comments, 79 for comments in the middle of the conversation, and 288 minutes for late comments.)

Table 5.6: Lower scoring comments took longer to receive first moderation.

Start score	Median time in minutes
-1	37
0	45
1	86
2	108

The second test was to look at the probability of reversing an incorrect moderation, as discussed in the previous section. Here, we restrict attention only to incorrect negative moderations, as those are the ones that can cause treasures (good comments) to be buried. Table 5.7 shows that the lower the current score for a comment, the lower the probability of reversing an incorrect moderation, suggesting that moderators attend less to comments with lower scores. Comments at top-level were more likely to have incorrect moderations reversed (44% vs. 35%). Comments early in a thread were also more likely to have incorrect moderations reversed (33% for very early comments, 19% for comments in the middle of a thread, and 12% for late comments).

Table 5.7: Errors were corrected less frequently for comments with lower scores.

Score of comment receiving "unfair" moderation.	% Reversed
-1	25%
0	32%
1	37%
2	46%
3	49%
4	57%

Limitations and Future Research

Additional data and analysis could provide even clearer evidence on the issues investigated here. By analyzing the contents of comments to identify typographic elements like bullets and bolded text, the presence of links to other comments, or other features, we could control for more potential confounds in the analysis of whether late comments, comments with lower initial scores, or not at top-level, had less of a chance to achieve high scores. With readership logs, we could measure the attention of moderators to particular messages rather than using time to moderation and other proxies. If a random sample of

Slashdot users rated a sample of comments as to what their final score should be, we could measure how frequently the distributed moderation system converged to correct final scores.

With both reader logs and assessments of correct final scores, it might be possible to distinguish problems of insufficient moderator attention from information cascades (Bikhchandani, Hirshleifer et al. 1998). That is, we could control for the amount of moderator attention and for the community's assessment of the correct final score when analyzing whether the previous moderation had any influence on the next moderation. If previous moderation still had an effect, it would imply an information cascade that could only be remedied by withholding from moderators the results of previous moderations. If previous moderation had no effect, then the problem of buried treasures could be remedied merely by redirecting moderation attention.

In addition to refining the analyses of moderation provision presented in this paper, in future research we plan to turn our attention to the impacts of moderation on readers and writers of comments. To what extent are readers making use of comment scores in allocating their attention and how could the scores be used even better? To what extent does the moderation system help newcomers to learn the norms of the community, encourage valued writers to keep participating, and drive away trolls?

Design implications

Slashdot's design, and the usage patterns that have emerged, highlight tensions among four design goals for distributed moderation systems. First, comments should be moderated quickly. Second, they should be moderated accurately according to the community norms. Third, each individual moderator should have limited impact on any particular comment. Fourth, the burden on moderators should be minimized, to encourage their continued participation.

Consider the tension among timeliness, accuracy, and minimizing the influence of individual moderators. In the Slashdot system, two to five people (depending on a comment's initial score) must provide positive moderations before a comment reaches a score of +4. This limits the impact of any individual moderator. But more than 40% of comments that reached +4 took longer than three hours to reach it; in three hours, the typical conversation was already half over. An alternative design would give more weight to early moderators, which would lead to earlier identification of treasures (and trash) but would give more power to those early moderators and lead to more errors caused by items having inappropriately high or low scores that would have to be corrected by future moderators.

There is also a tension between minimizing moderator effort on the one hand, and timeliness and quality of moderation outcomes on the other hand. At Slashdot, moderators choose which comments to attend to, and only provide feedback on comments that they think should be moved from their current score. This minimizes disruption to moderators' usual reading patterns. Our analysis showed, however, that it leads to biases. Comments with lower current scores, comments not at top-level, and comments later in a thread received slower moderation and lower scores on average than they deserved.

Alternative designs might cause treasures to be discovered more quickly and consistently, at the expense of a little more moderator effort. For example, there could be a special moderator's view of a conversation. It would hide comments below certain thresholds, as with the view presented to other readers. But comments the system had flagged as needing additional moderator attention would not be hidden. Recently posted comments and those with recent moderation would be flagged. Once a flagged comment had been presented to enough moderators, the system would infer from the lack of any explicit moderator action that the item was correctly classified and stop highlighting it for future moderators. All comments would reach their final score much faster, and the problems of uncorrected moderation errors and buried treasures would be reduced significantly.

Conclusion

Slashdot is an unusual site. Many more people participate in each conversation thread than is typical of conversation spaces on the Internet. Slashdot's mostly tech savvy, younger users, may be especially good at using the moderation tools. The design has accreted slowly, giving users plenty of time to adapt to it. Rather than limiting the value of this analysis, however, we believe these characteristics of Slashdot make it an especially valuable site to study. The scale of the site makes moderation a necessity rather than a luxury and patterns of moderator behavior that have emerged shed light on the fundamental tensions involved in distributed moderation systems.

Slashdot provides an existence proof that the basic idea of distributed moderation is sound. There is widespread participation. There seems to be a broad, though not perfect consensus about which comments deserve to be moderated up or down. Comment scores are dispersed so that they offer some information of potential value to readers.

Closer analysis, however, revealed that it often takes a long time for especially good comments to be identified. We also found that incorrect moderations were often not reversed, and that later comments, comments not at top-level, and comments with low starting scores, did not get the same treatment from moderators as other comments did. These findings highlight tensions among timeliness, accuracy, limiting the influence of individual moderators, and minimizing the effort required of individual moderators. We believe any system of distributed moderation will eventually have to make tradeoffs among these goals. There is still room, however, for design advances that require only modestly more moderator effort to produce far more timely and accurate moderation overall.

Chapter 6
Conclusion

The central objective of this research was to examine an online conversation rating system "in the wild"; to understand how participants in Slashdot use ratings to structure their community. The work was intended to show how readers use ratings to customize their view, how ratings affect new members entering Slashdot, and how moderators apply ratings to comments.

Online discussions have historically been plagued with the problems of information overload (Jones, Ravid et al. 2004) caused by large numbers of participants and trouble establishing mutual understanding between participants caused by the constrained information channels inherent in text-based interactions (Sproull and Kiesler 1991; Clark 1992). Although researchers and designers have applied many solutions to these problems in the past, solutions have not typically been applicable to large-scale interactions. Technologically enabled recommendation systems use the past experiences of some users to make suggestions to others (Resnick and Varian 1997), and address the issues of massive participation without curtailing it.

The research described here is designed to show how Slashdot has used a recommendation system to shape user participation. The main body of this dissertation is comprised of three inquiries related to rating use on Slashdot. The first inquiry, as described in Chapter 3, examined how readers use ratings to customize how comments are displayed. The second project, described in Chapter 4, described the role comment ratings may have in shaping the participation of users new to Slashdot. The third inquiry, described in Chapter 5, reported on the provision of ratings on Slashdot comments, looking specifically at whether comments are sufficiently and fairly rated.

In the next section, I briefly review the findings of the three studies. Next, I consider the overall significance of these findings in the context of structuring online discussions. I start with instances of the Slashdot rating system that might be generalized to other systems, and follow that with a discussion of possible changes that might benefit Slashdot moderation. Third, I describe the limitations of this research and possible responses to those limitations. Finally, I discuss future research in this area that would extend this work.

Findings

Before discussing the overall implications of this research, I'll briefly enumerate the findings from the individual studies. In Chapter 3, I found that comment ratings are useful to users, as measured by user survey responses and percentage of users who employ ratings to customize how comments are displayed. I also conclude that there is some "friction" that prevents users from readily changing their viewing threshold, given that users tend to make large changes to thresholds rather than incremental changes. Given that there is something that prevents users from changing thresholds, I recommend methods of using the choices of other users to reduce that friction. The first method for reducing friction is to create viewing schemas that support different goals for reading, based on the clusters of moderation label modifications made by some users. For example, a view for "Gem Seekers" might be derived from the choices of users who have modified those moderation labels that increase the value of "Informative" and "Insightful" comments as labeled by the moderation system. Another method of reducing the resistance to changing viewing thresholds is to model the behavior of "lead users", those users who seem more ready to change viewing thresholds either because they are more expert with the system or have lower tolerance for information overload. Chapter 3 shows that whether lead users who had previously entered a discussion thread had changed their threshold was a significant predictor of whether the new lead user would do so. This indicates it is possible to predict useful, dynamic changes in viewing thresholds based on what previous readers of the forum did.

Chapter 4 addressed the role of comment rating in shaping new user participation behavior. The chapter describes that whether a new user receives positive, negative, or no moderation on their first comment did not predict whether the user would make a second comment, or how many comments the user made overall, but did seem to relate to how much time passed between posting a first comment and a second. This chapter also examined the role of observation, as measured by user page views, in understanding new user behavior. How many page views a new user had made before posting was not associated with the score of their first comment, or how long it took to post a second comment, but was related to an increased probability of posting a second comment, the score of the second comment, and the total number of comments made. The conclusions drawn from this work is that ratings do provide some measure of feedback, and possible incentive, to new users but that other factors like observation also play important roles in shaping new user participation.

Chapter 5 examined whether comment ratings were sufficiently and fairly provided. This inquiry found that although only 28% of comments received explicit ratings, the distribution of scores was reasonably dispersed because of the automatic bonuses given to messages from some users. Second, we found that moderators generally agreed on the disposition of a comment, as measured by the variability of ratings on a comment and the meta-moderation scores of ratings. Third, this work described how moderations that were perceived to be unfair by the meta-moderation system were more likely to be reversed by subsequent ratings. Fourth, the chapter described how comments posted late in the life of a thread, deeper within a thread or with lower starting scores were less likely to get moderated than other comments, meaning that ratings are not entirely fairly applied to all comments.

Discussion

In general, the Slashdot moderation system seems to work. Readers employ the ratings to change how they view comments, new users receive some amount of feedback from the scores of their comments, survey respondents report high levels of satisfaction with the moderation system, and the moderators mostly agree on the disposition of a comment when making ratings, and reverse “incorrect” moderations.

However, there are still gaps in the overall use of ratings on Slashdot. First, 15% of users chose to customize their readings views to ignore ratings at least once while reading Slashdot. Second, new users may be adversely affected by initial ratings in terms of how quickly they post additional comments after their first. Third, not all comments were as likely to receive ratings, meaning that some worthwhile content may be ignored because of when and where it was posted rather than the content of the message. As discussed in Chapter 1 in the section on contributions of this work, these findings confim potential issues for online rating systems that have been discussed in theoretical descriptions of evaluation systems (Banerjee 1992; Bikhchandani, Hirshleifer et al. 1998; Avery, Resnick et al. 1999; Dellarocas 2003)

The usefulness of ratings in online discussions was not obvious before studying the Slashdot case. Although collaborative filtering systems had been applied to online discussion previously by GroupLens (Resnick, Iacovou et al. 1994) and demonstrated to have some utility for users, they were not widely deployed. This works extends the recommender system literature in two ways. First, it provides an example of ratings that occur on a very different object than is typical in other recommender system applications. Much of the work on recommenders has examined ratings of consumer goods, as in Epinions and Movielens (Terveen and Hill 2002). Online discussions are different from other "goods" being evaluated in that they can happen over a very brief amount of time, meaning that evaluation needs to happen quickly to be useful. Rating discussion may also be different because raters may evaluate a comment both on its quality and on whether they agree with its content. This creates more opportunities for abuse of the rating system. Interestingly, such abuse appears to be rate which does not seem to occur given the high level of agreement between Slashdot moderators on the disposition of comments.

Analyzing and describing specific features of the Slashdot moderation system also extends the literature on recommender systems by adding new case examples for future comparisons and study. The use of moderation labels rather than just numeric rating, the embedded reputation system that is tied to content recommendation and the mechanism by which raters only move the score of a comment up or down if they disagree with the current score of the comment are all unique features to the Slashdot rating system, and add to our understanding of how these systems may be used.

As shown in Chapter 3, some readers showed sophisticated use of moderation labels to craft comment views that addressed different motivations for reading Slashdot forums. For example, users who enjoyed reading comments labeled as “Troll” could add positive weight to those comments and see them. Other users who chose to decrease the score of comments labeled as “Funny” may have been reading comments for a different experience than entertainment. Even though the Slashdot system makes general recommendations rather than customized ones through collaborative filtering mechanisms, rating labels provide additional meta-information that allows for more sophisticated use than typically found in general recommenders. Chapter 3 recommends using moderation labels to create schemas for viewing comments, like “Gem Seeker” or “Muck Raker”. The overall conclusion I draw from the use of these labels is that rating metadata can be useful for creating heterogeneous recommendations. Collaborative filtering personalizes recommendations by matching users to like others through algorithms that find connections in user preferences (Terveen and Hill 2002). Using rating metadata allows for the Slashdot system to target types of users, rather than the whole population of users, as is the practice in general recommender systems, or each individual user, as is the practice in collaborative filtering mechanisms.

Second, pre-rating comments based on the history of the user posting the comment allows for a dispersion of scores without depending on every comment in a thread receiving attention from moderators. As shown in Chapter 5, only 28% of comments received any moderation during that study period. Since comments could start anywhere from -1 to +2 based on the identity of the posting user, the moderation system acted as a winnower, separating particularly good and bad content from the population of comments. Consequently, relatively low amounts of moderation could still provide useful feedback to readers about which comments to read. The usefulness of pre-rating might be extended to use the context in which the comment appears as well as the reputation of the user posting it. For example, a comment posted deeper in a thread might receive an additional pre-rating weight based on the score of its parent comment. If a sub-thread has a high rate of posting, then comments within that thread might receive additional weight, or lose weight if there have been no recent posts within that sub-thread.

Limitations

Slashdot has developed its system incrementally over several years, and targets a particularly technology savvy audience. As shown in Appendix A, the Slashdot audience is largely male, computer expert, familiar with other types of online communication, and well-educated. This audience, though targeted, is also very large, which is not consistently true for many other online discussion forums. These characteristics may reduce the ability to generalize findings from Slashdot to other online discussions. Particularly, it could be that it is impossible to map the findings from the Slashdot case to the design of a new online community, and that technological tools like the Slashdot moderation system need to co-evolve with the particular needs of an online social system.

Although Slashdot is an exceptional case, it may also act as a “leading edge” example of how people will interact as online interactions become more commonplace. The Slashdot system may not map wholesale to other online communities, but features like moderation labels and pre-ratings of comments may be usefully employed in future designs. Just as Slashdot borrowed features from Usenet and MUDs, future online conversation systems may benefit by using tools that Slashdot has developed.

This research uses server log data to infer user behavior and survey data to collect user opinions and characteristics. While these methods were useful for the specific questions asked in the individual chapters, they don't answer more nuanced questions of motivation and interpretation. Methods such as ethnographic interviews and content analysis could have been used to answer these more contextual questions. For example, in looking at new user behavior, I only examined the outcome of a user's first comment in terms of how it was rated, not in how good the comment actually was. In looking at reading behavior, I described several uses of moderation labels, but did not learn about perceived efficacy of customizing label ratings.

Each of the studies that comprise the dissertation depended on cross-sections of server logs and user surveys that represent "snapshot" views of behavior on Slashdot. It could be that moderation practices change over time, or that a different set of readers constitute the audience of Slashdot now. Cross-sections of user behavior do not show longitudinal changes to participation on the site, which may play an important, undetected role in these findings.

Selecting the three analyses of ratings use of Slashdot described in this chapter required ignoring other important elements of the Slashdot system. Slashdot, given its size and persistence, is a complex socio-technical environment. Some of the interactions not studied here could play an important, undetected role in explaining these findings.

These studies focused on the moderation system of Slashdot, but other characteristics of the site may be even more important in shaping user behavior.

Future Research

Opportunities for future research can be divided broadly into two categories: future research on Slashdot and future research applying these findings to other systems. In future studies of Slashdot, there are good opportunities for understanding the context of use by employing different methods combined with those used here. The following list explores some additional research questions for ratings use on Slashdot:

- The role of "trolls", socially deviant users, in affecting how moderation is enacted on Slashdot. What types of user misbehavior does moderation affect and what types does it ignore? What role does user misbehavior have in the design of conversation rating systems? Using interviews with trolls, editors and Slashdot users this study could examine the tactics trolls use to work-around the moderation system and how editors respond to those tactics. Secondary analysis of troll comments could be used to determine the success of such tactics.
- How does a moderation system where users move scores in a "better or worse than current score" mechanism differ from absolute ratings? An implication of this mechanism is that some raters may not score a comment because they feel it is at the correct level. This approach may also resolve some of the sufficiency issues found in recommender systems by not requiring that every item be rated. More work comparing absolute ratings versus this form of rating would be valuable to research on recommender systems.

- What is the relationship between the topic being commented on and the ratings of those comments? There is some indication that political stories engender more contentiously rated comments than technology stories. Are there other topics that will affect the provision of ratings? Two methods might be combined to address these issues. The first would be a lab study in which lab participants are randomly assigned to rate comments from different types of stories and see if the overall ratings differ. The second method would be to look at how moderators rate comments in existing threads of different story types and compare the ratings using variability within the scores and measures of unfairness in meta-moderation.
- Does the final score of a comment constitute its "correct" score? Since comments reach a final score through a series of moderations, and only the moderations themselves are meta-rated, it's unclear whether Slashdot users agree with the end dispositions of the comments. One method of detecting the difference between the "true" score and the score achieved by moderation would be to have users assign absolute scores to comments and compare those ratings to what scores the comments achieved through moderation. Using surveys and interviews, this inquiry would examine how much users trust the ratings of comments.
- What information do moderators and meta-moderators seek before rating a comment? Raters have the option of seeking information about the user who posted the comment, the context in which a previous rating was made and the history of moderation labels attached to a comment and more when deciding to moderate a comment or meta-moderate a moderation. This inquiry would use

server-log analysis and interviews with active Slashdot moderators to track what information, if any, a rater collects before assigning a label to a comment.

- What variables would be useful in additionally pre-rating comments? As suggested in the discussion section, pre-rating comments plays an important role in Slashdot moderation. Other variables that might be of use in pre-rating a comment include the score of parent comment, the average scores of comments in the thread, the rate of posting in a sub-thread, and the number of comments in a forum. The research on this subject would compare different variables and different weights to see which measures increase comment score dispersion and moderator agreement with final scores.

Besides future work on Slashdot, the research described here suggests research questions that may be best raised by studying other sites. These questions include:

- Can moderation be applied to other online discussion forums? It is possible that Slashdot is a unique case, and rating comments would not work in other contexts. Two ways to test this would be to apply a comment rating system to an ongoing discussion forum without a rating system, or to create a new online discussion system with a moderation system embedded from inception. Comparing the development of rating practices in these two contexts might show the possibility of imposing ratings in ongoing systems, or the difficulty in starting rating mechanisms in new systems.
- Are moderation labels useful in other contexts? One method to study this would be to take an online community that currently uses a numerical rating system for

content and replace it with a labeled rating system for groups of users. For example, Movielens could replace the familiar 5 star framework for rating movies with labels like "Fun" or "Matinee Only". Another method would be to create a new online community that only uses the moderation labels rather than numerical ratings and examine how use develops. For example, an online community based on mental health support might be created and choose to rate content with labels like "Inaccurate" or "Supportive". Finally, a moderation label system might be applied to an existing online community that doesn't currently use ratings at all. For example, Wikipedia might include a rating system with labels like "Incomplete" or "Needs revising" to provide cues about which content needs more work. This may be useful for the emerging practice of social tagging or "folksonomies". These are systems in which users assign labels to content like links and pictures, and those labels are used to match participant interests. The site "del.icio.us"[16] allows users to assign labels to Web page links and create categories based on those tags. Further research on labeling would help understand this emerging practice.

- What other content metadata can be applied in other settings? Besides moderation labels, user characteristics that determine pre-rated scores were important in Slashdot's moderation system. Other systems might benefit by using comment and user metadata to differentially weigh the rating of content without using collaborative filtering. For example, the site e-thepeople[17] allows users to rate posts in a political discussion. A test of this research question might be to

[16] http://del.icio.us/
[17] http://www.e-thepeople.org/

pre-rate comments based on political leanings imputed from previous rating behavior. This would allow a liberally-oriented reader to see comments rated as "High quality conservative". A study of the Amazon book review system might weigh or sort user books reviews by looking at the history of the user posting the book review, like agreement with their previous reviews.

Conclusion

Slashdot provides an example of how an online discussion system employs user recommendations to manage the problems of large-scale interactions without losing the benefits of such participation. The increased participation in many-to-many interactions and the development of technological tools to help manage those interactions have fostered new forms of participation that were not possible in the offline environment. However, characteristics of computer-mediated communication that undermine these benefits need to be managed in order to realize these innovative interactions. Ratings systems, as used by Slashdot, help to manage large-scale online interaction without constraining them such that you lose the benefits of scale.

In this dissertation I have shown that comment ratings are useful to readers of content, and that they have some effect on participation outcomes for new users. I have also shown that ratings on Slashdot are sufficiently and fairly applied. However, I have also pointed to several areas of improvement that are possible for ratings in this system. Readers have some inherent resistance to changing the viewing interface, and would benefit from more dynamic changes based on ratings. New members are also affected by how they observe content on the site, which has implications for designing new user experiences. Some comments on Slashdot were not as likely to get rated, so an interface that rewards moderators for finding valued content that might be missed would help the overall system.

Large-scale online interactions enable new forms of interaction with intriguing potential benefits. Bringing massive attention and resources to a shared environment allows for the accomplishment of complex tasks with reduced burden on individual contributors. Recommendation systems, like that used by Slashdot, may be an important tool in helping to realize the benefits of those interactions, but is not a panacea for all of the issues involved with coordinating online activity. As online interactions become more nuanced and sophisticated, the socio-technical tools we use to structure those interactions must keep pace. Hopefully, recommendation systems continue to develop as one of those mechanisms that support rich, unmediated interaction between people online.

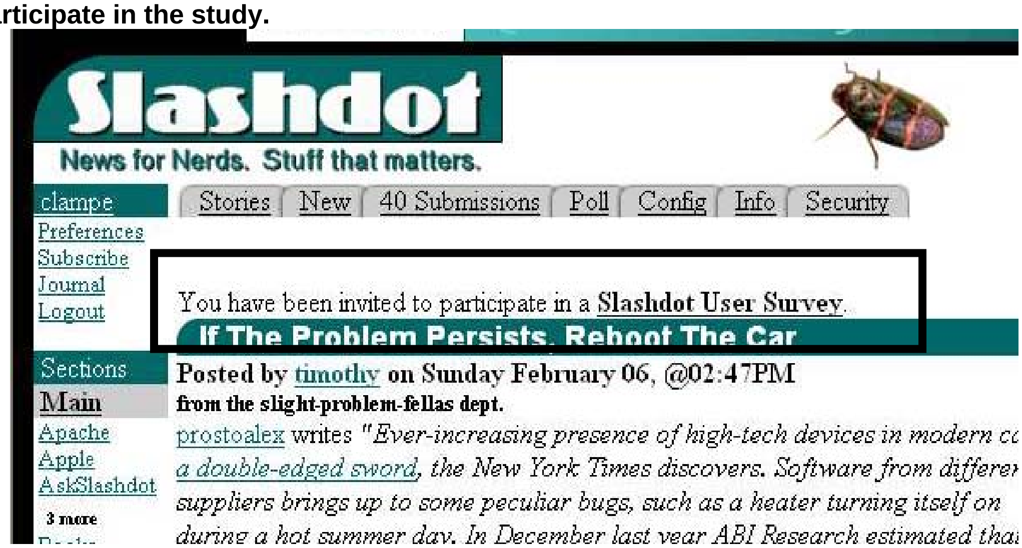
Appendix 1

Survey

To add context to the behavioral patterns contained in server logs, this project also relied on user characteristic information derived from two surveys of Slashdot users. One survey was targeted towards a group of new users of the site, and a second survey sampled the general population of Slashdot users.

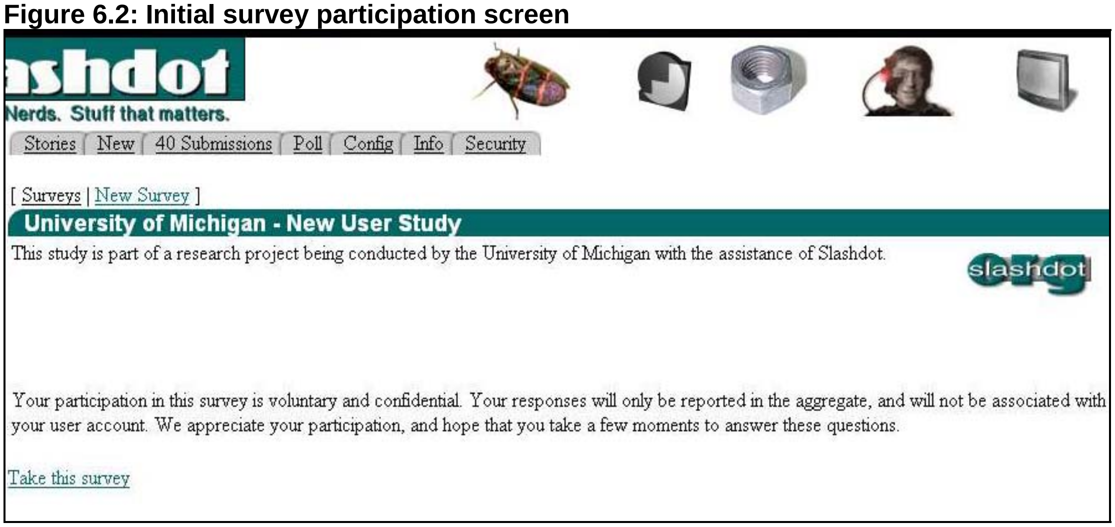
In each survey, users were notified of their invitation to participate through text at the top of the index page, as shown in Figure 3-7. This area is generally reserved for notifications of status changes like moderator or meta-moderator privileges, or invitations to subscribe to the site. The instrument itself was coded to operate on the Slashdot site, rather than taking users to another site to complete the survey.

Figure 6.1: The notification area where registered Slashdot users received their invitation to participate in the study.



Couper (Couper 2000) describes this type of survey as an "intercept survey", with the sampling frame limited to site visitors. He identifies the two main concerns with intercept surveys as timing when to invite users to participate in the study and the higher possibility for nonresponse bias. The timing issue involves selecting at what point in browsing a site the user is invited to participate in the survey, with Couper recommending that potential participants be invited at the beginning of their time on the site. In line with this advice, this study placed the invitation on the first page that users would see when logging onto the site, provided they did not follow a link to a section that bypasses the main index page. Nonresponse bias is the potential error in findings introduced by the likelihood that participants who responded to the survey have different characteristics than those who did not. Response rates for intercept surveys range between 15% and 30% (Couper 2001).

Figure 6.2: Initial survey participation screen



The sampling frame for the survey of new users was a list of all users who had created an account between November 1, 2004 and the day the survey first entered the field, December 6, 2004. Survey invitations appeared for a two day period, during which time only 3,341 users identified in the dataset visited the site. 233 new users responded to the survey, meaning this part of the study had a response rate of 7%. In this case, the short study period and the type of users being targeted may have had an adverse effect on overall response rate.

The new user survey was designed to elicit demographic information about new users, as well as their beliefs about the moderation system. In particular, this data collection was designed to ask about previous experiences that might predict how users participate on Slashdot. Table 3-2 shows the questions that were asked in the new user survey, clustered by previous experience demographics and attitudes towards the moderation system, and beliefs in their own abilities to detect and create highly rated comments.

Table 6.1: New user survey questions

<u>Previous experience</u>
How much experience do you have participating in online discussion sites?
How expert do you consider yourself with computers?
What is your level of education?

<u>Attitude towards moderation system</u>
Compared to other sites you may have participated on, how would you rate discussion on Slashdot?
How concerned are you about the scores you receive on your comments?
How much do you agree with the following statement?
Moderators should use moderation to support particular viewpoints?
How much do you agree with the following statement?
The moderation system is important for fostering good discussion on Slashdot.
On average, how deserving are comments that receive very high scores?

<u>Ability to detect/create highly rated comments</u>
How sure are you that you know what a good comment is on Slashdot?
How confident are you that you understand why other's comments are usually scored very highly?
How clear are the expectations for making a quality comment on Slashdot?
How confident are you that you could write a comment that scored very highly?

The second online survey conducted on Slashdot was The sampling frame for this survey was the list of registered Slashdot users, based on the Slashdot assigned unique identifying number. User identification numbers were compared against IP addresses to assure that multiple numbers were not held by single individuals. Each day a script chose 10% of the registered Slashdot users to receive an invitation to participate in the survey.

Between June 15 and June 20, 2005, 8121 respondents participated in the study. The overall response rate for the study period was 19.1%, with some variation per day. This is a conservative calculation of response rate for several reasons. First, it doesn't account for users who were eligible to receive the survey, but did not happen to log in to the site during the study period. Secondly, users who did access the site might have missed the invitation to participate in the survey, as it was surrounded by a large amount of similar-looking text. Crawford et al (Crawford, McCabe et al. 2002) compared nonresponse in Web and phone surveys, and found that the largest reason given for not responding to Web surveys was that users did not remember being invited to do so, or did not notice the invitation in the first place. They also found that those users were willing to participate when they realized they were eligible.

This survey was designed to elicit two types of information: characteristics of Slashdot users and attitudes towards the moderation system. User characteristics can also be divided into demographic information, and other online conversation technology usage patterns. Table 3-3 lists the questions asked of respondents.

Table 6.2: Questions in Slashdot user survey

User demographics
How expert do you consider yourself with computers?
What is your level of education?
How old are you?
What is your gender?

Online conversation usage
How much experience do you have participating in online discussion sites?
On average, how many other news and discussion sites per day do you visit besides Slashdot?
How often in the past week did you engage in sending and receiving email?
How often in the past week did you engage in instant messaging or IRC?
How often in the past week did you participate in a Usenet newsgroup?
How often in the past week did you particpate in Web discussion forums?

Attitudes towards Slashdot and moderation
How much do you agree with the following statement?
It is easy to find highly rated comments in each story.
How often do you click through a story to read comments?
How much do you agree with the following statement?
The moderation system is important in identifying good comments.
Compared to other sites you may visit, how would you rate discussion on Slashdot?
How important do you think the moderation system is in fostering discussion on Slashdot?

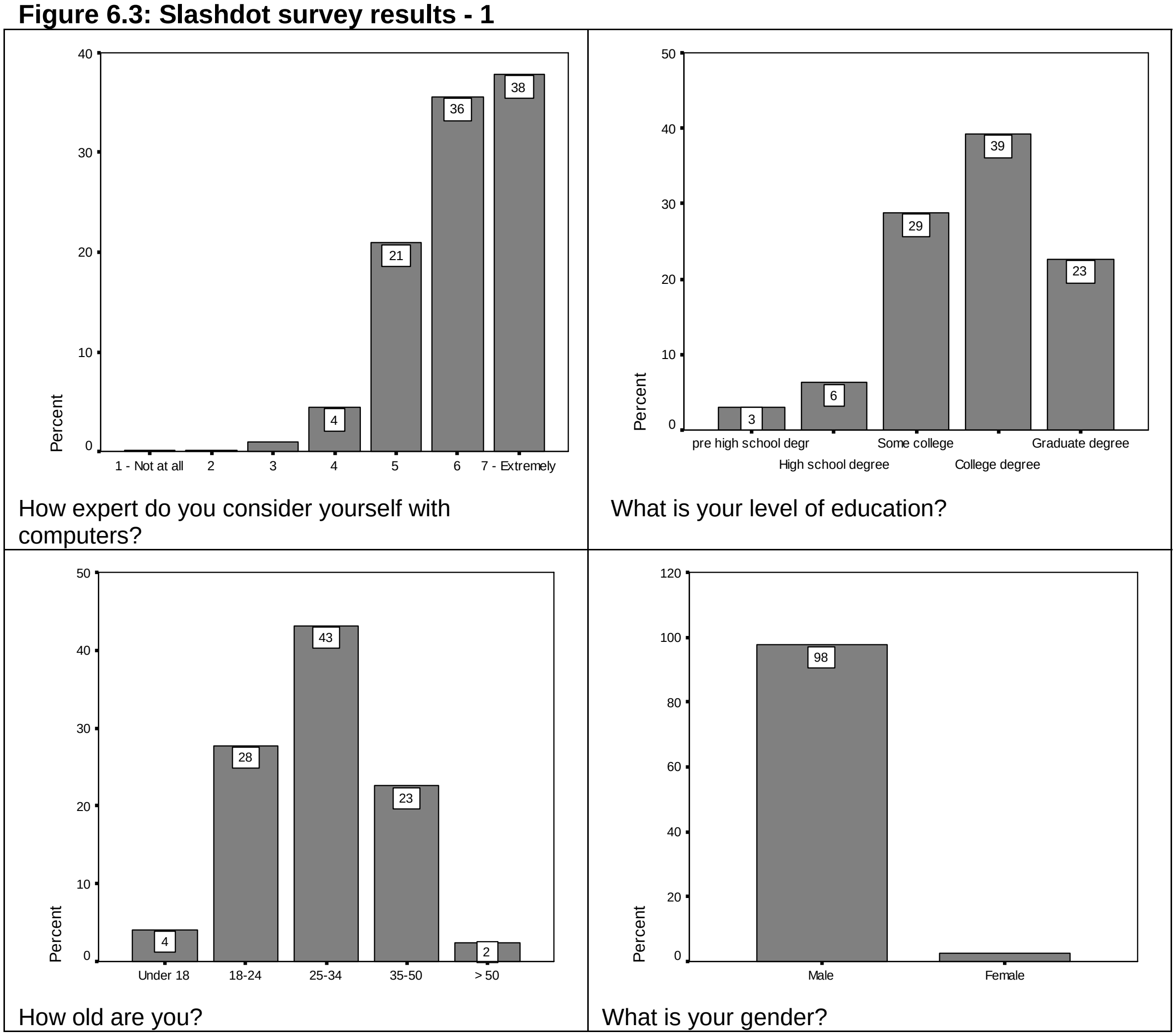
General population survey results

Although the new user survey is targeted to a specific group, the general population survey is a good approximation of the demographics and beliefs. Consequently, this section will describe the results of the survey, which will then be referred to in later chapters.

User demographics

Slashdot users have high levels of self-rated computer expertise, and tend to be highly educated. 43% of Slashdot users were in the 25-34 age category, skewing the Slashdot audience as younger than the general Internet population (Rainie and Horrigan 2005). 98% of Slashdot users reported being male, another significant difference from general Internet populations, which are currently nearly equally divided between males and females. These statistics paint a picture of a technologically savvy, highly educated, male population that mostly confirms popular impressions of the Slashdot audience.

Figure 6.3: Slashdot survey results - 1



How expert do you consider yourself with computers?

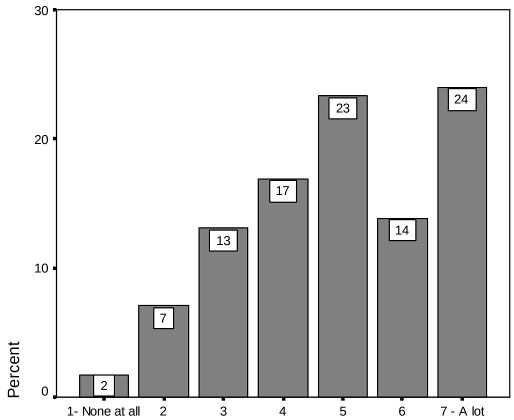
What is your level of education?

How old are you?

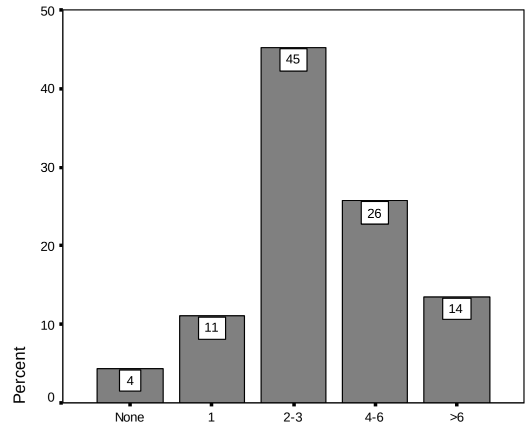
What is your gender?

Online conversation usage

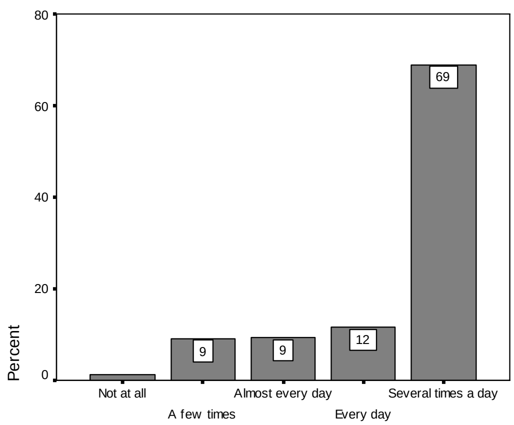
Figure 6.4: Slashdot survey results - 2



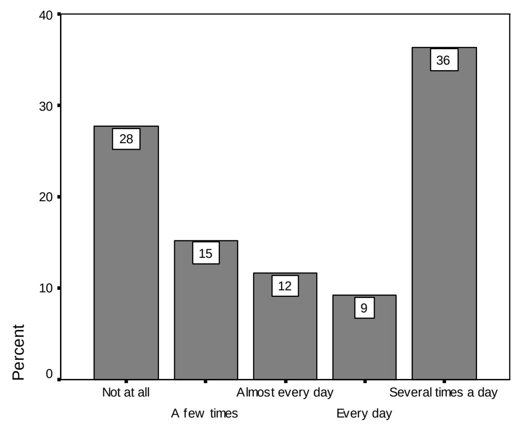
How much experience do you have participating in online discussion sites?



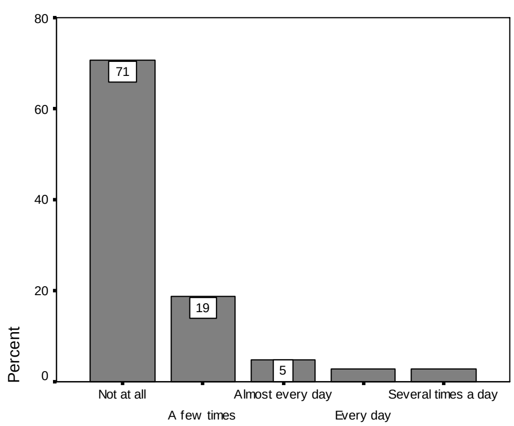
On average, how many other news and discussion sites per day do you visit besides Slashdot?



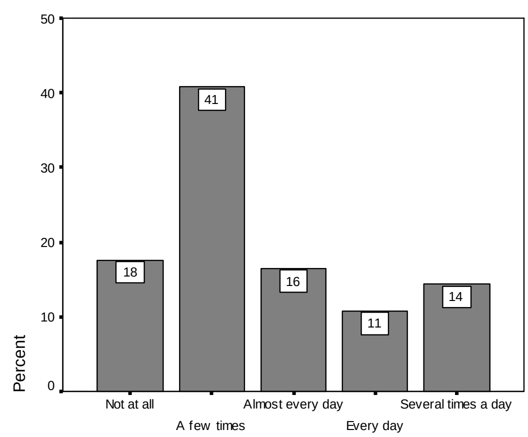
How often in the past week did you engage in sending and receiving email?



How often in the past week did you engage in instant messaging or IRC?



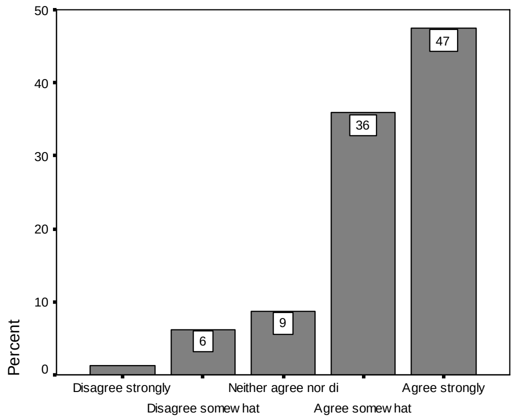
How often in the past week did you participate in a Usenet newsgroup?



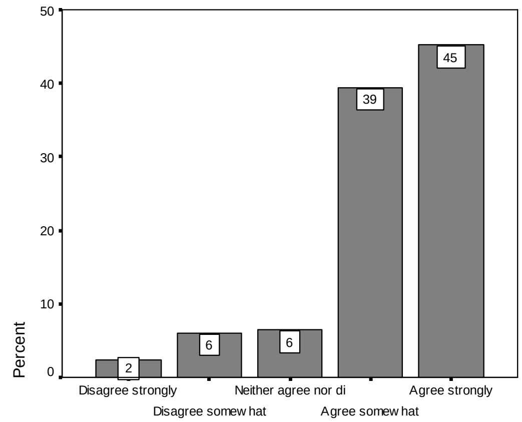
How often in the past week did you particpate in Web discussion forums?

Attitudes towards Slashdot and moderation

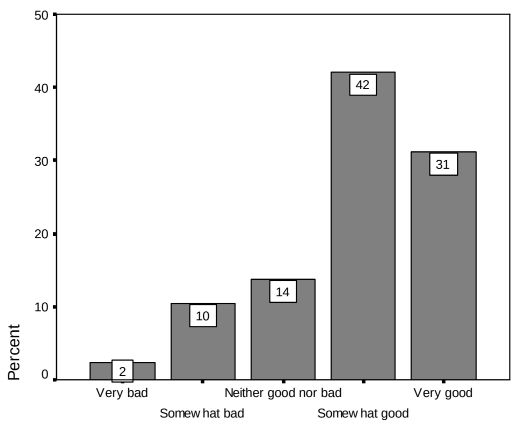
Figure 6.5: Slashdot survey results - 3



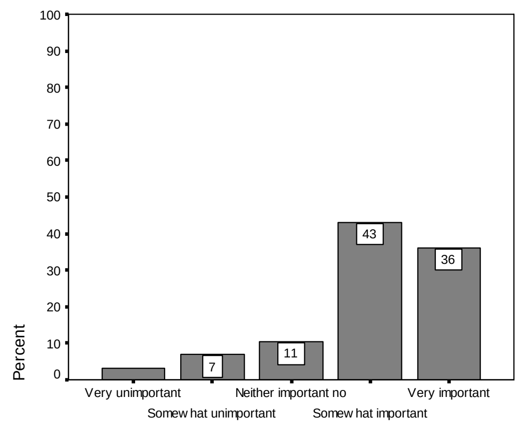
It is easy to find highly rated comments in each story.



The moderation system is important in identifying good comments.



Compared to other sites you may visit, how would you rate discussion on Slashdot?



How important do you think the moderation system is in fostering discussion on Slashdot?

Bibliography

Abbate, J. (2000). Inventing the Internet. Cambridge, The MIT Press.

Ackerman, M. S. and D. W. McDonald (1996). Answer Garden 2: Merging Organizational Memory with Collaborative Help. Proceedings of the ACM Conference on Computer-Supported Cooperative Work, ACM.

Ackerman, M. S., A. Swenson, et al. (2003). I-DIAG: From Community Discussion to Knowledge Distillation. Conference on Communities and Technologies.

Adamic, L. and N. Glance (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, WWW2005, Japan.

Ali, K. and W. vanStam (2004). TiVo: making show recommendations using a distributed collaborative filtering architecture. Conference on Knowledge Discovery in Data, Seattle, WA.

Avery, C., P. Resnick, et al. (1999). "The Market for Evaluations." American Economic Review **89**(3): 564-584.

Balabanovic, M. and Y. Shoham (1997). "Combining Content-Based and Collaborative Recommendation." Communications of the ACM(March, 1997).

Bandura, A. (1973). Aggression: a social learning analysis. Englewood Cliffs, NJ, Prentice Hall.

Banerjee, A. (1992). "A Simple Model of Herd Behavior." Quarterly Journal of Economics **107**(3): 797-818.

Bianchini, M., M. Gori, et al. (2005). "Inside PageRank." ACM Transaction on Internet Technology **5**(1): 92-128.

Bikhchandani, S., D. Hirshleifer, et al. (1989). "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades." Journal of Political Economy **100**(5): 992-1026.

Bikhchandani, S., D. Hirshleifer, et al. (1998). "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades." The Journal of Economic Perspectives **12**(3): 151-170.

Bruckman, A. (1997). MOOSE Crossing: Construction, Community, and Learning in a Networked Virtual World for Kids. Program in Media Arts and Sciences, School of Architecture and Planning, Massachusetts Institute of Technology**:** 231.

Bryant, S., A. Forte, et al. (2005). Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. ACM-GROUP, Sanibel Island, FL.

Butler, B. (1999). When is a group not a group: An empirical examination of metaphors for online social structure. Pittsburgh, Carnegie Mellon University.

Butler, B. S. (2001). "Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures." Information Systems Research **12**(4): 346-362.

chromatic, B. Aker, et al. (2002). Running Weblogs with Slash, O'Reilly & Associates.

Clark, H. H. (1992). Arenas of Language Use. Chicago, IL, University of Chicago Press.

Collins, M. P. and Z. L. Berge (1997). Moderating Online Electronic Discussion Groups. the American Educational Research Association, Chicago, IL.

Cosley, D., S. K. Lam, et al. (2003). Is Seeing Believing? How Recommender Interfaces Affect Users' Opinions. Computer Human Interaction (ACM-CHI), Minneapolis, MN, ACM Press.

Couper, M. (2001). The Promises and Perils of Web Surveys. The Challenge of the Internet, UK, Association for Computing Machinery.

Couper, M. P. (2000). "Web surveys: A review of issues and approaches." Public Opinion Quarterly **64**: 464-494.

Crawford, S., S. McCabe, et al. (2002). From Mail to Web: Improving Response Rates and Data Collection Efficiencies. International Conference on Improving Surveys, Copenhagen, Denmark.

Cristiano, T. (1999). The Significance of Public Deliberation. Deliberative Democracy. J. Bohman and W. Rehg. Cambridge, MA, MIT Press.

Crystal, D. (2001). Language and the Internet. Cambridge, UK, Cambridge University Press.

Curtis, P. (1992). Mudding: Social phenomena in text-based virtual realities. Conference on Directions and Implications of Advanced Computing, Berkeley, CA.

Curtis, P. and D. A. Nichols (1993). "MUDs Grow Up: Social Virtual Reality in the Real World."

Cypher, A. (1991). Customizing End-User Applications. East-West HCI Conference, ICSTI.

Dave, K., M. Wattenberg, et al. (2004). Flash forums and forumReader: navigating a new kind of large-scale online discussion. CSCW: Computer-Supported Cooperative Work, Chicago, IL, ACM Press.

Dellarocas, C. (2003). "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms." Management Science **49**(10): 1407-1424.

Facca, F. M. and P. L. Lanzi (2005). "Mining interesting knowledge from weblogs: a survey." Data & Knowledge Engineering **53**(3): 225-241.

Findlater, L. and J. McGrenere (2004). A comparison of static, adaptive, and adaptable menus. Conference on Human Factors in Computing Systems, Vienna, Austria, ACM Press.

Finholt, T. A. and G. M. Olson (1997). "From laboratories to collaboratories: A new organizational form for scientific collaboration." Psychological Science **8**: 28-36.

Fiore, A. T., S. L. Tiernan, et al. (2002). Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap. CHI2002, Minneapolis, MN, ACM.

Fischer, G. and A. Girgensohn (1990). End-User Modifiability in Design Environments. Computer Human Interaction (CHI), ACM Press.

Fischer, G. and B. Reeves (1992). "Beyond Intelligent Interfaces: Exploring, Analyzing and Creating Success Models of Cooperative Problem Solving." Journal of Applied Intelligence **1**: 311-323.

Fisher, D. and C. Lueg (2002). Studying Online Newsgroups. From Usenet to CoWebs: Interacting with Social Information Spaces. C. Lueg and D. Fisher. New York, NY, Springer Verlag.

Fisher, D. A. (2004). Social and Temporal Structures in Everyday Collaboration. Information and Computer Science, University of California, Irvine**:** 214.

Forsyth, D. R. (1999). Group Dynamics. Belmont, CA, Wadsworth Publishing Company.

Friedman, E. J. and P. Resnick (1997). "The Social Cost of Cheap Pseudonyms." Journal of Economics and Management Strategy **10**(2): 173-179.

Galston, W. (1999). Does the Internet Strengthen Community?, Institute for Philosophy & Public Policy. **November, 2000**.

Gillmor, D. (2004). We the Media. Sebastopol, NC, O'Reilly Media Inc.

Goffman, E. (1963). Behavior in Public Places: Notes on the Social Organization of Gatherings. New York, NY, The Free Press.

Goldberg, D., D. Nichols, et al. (1992). "Using collaborative filtering to weave an information tapestry." Communications of the ACM **35**(12).

Grice, H. P. (1969). "Utterer's meaning and intentions." Philosophical Review **78**: 147-177.

Hafner, K. and M. Lyon (1996). Where Wizards Stay Up Late: The Origins of the Internet, Simon & Schuster.

Halavais, A. (2001). The Slashdot Effect: Analysis of a Large-Scale Public Conversation on the World Wide Web. Communications. Seattle, WA, University of Washington.

Herlocker, J. L., J. A. Konstan, et al. (2000). Explaining Collaborative Filtering Recommendations. ACM 2000 Conference on Computer Supported Cooperative Work, Philadelphia, PA.

Herlocker, J. L., J. A. Konstan, et al. (2004). "Evaluating collaborative filtering recommender systems." ACM Transactions on Information Systems **22**(1): 5-53.

Hill, W. and L. Terveen (1996). Using frequency-of-mention in public conversations for social filtering. ACM 1996 Conference on Computer Supported Cooperative Work, New York, ACM Press.

Hiltz, S. R. and M. Turoff (1978). The Network Nation: Human Communication via Computer. London, UK, Addison-Wesley.

Hiltz, S. R. and M. Turoff (1985). "Structuring Computer-Mediated Communication System to Avoid Information Overload." Communications of the ACM **28**(7).

Hollan, J. and S. Stornetta (1992). Beyond Being There. Conference on Human Factors in Computing Systems (CHI), Monterey CA, ACM Press.

Horn, S. (1998). Cyberville: Clicks, Culture, and the Creation of an Online Town. New York, NY, Warner Books.

Horrigan, J. B. (2001). Online Communities: Networks that nurture long-distance relationships and local ties. Washington, DC, Pew Internet & American Life Project.

Ilgen, D. R., C. D. Fisher, et al. (1979). "Consequences of individual feedback on behavior in organizations." Journal of Applied Psychology **64**(4): 349-371.

Jones, Q. (1997). "Virtual-Communities, Virtual Settlements & Cyber-Archaeology: A Theoretical Outline." Journal of Computer-Mediated Communication **3**(3).

Jones, Q. and S. Rafaeli (1999). User Population and User Contributions to Virtual Publics: A Systems Model. GROUP'99, Phoenix, AZ, ACM Press.

Jones, Q., G. Ravid, et al. (2002). An empirical exploration of mass interaction system dynamics: Individual information overload and Usenet discourse. 35th Hawaii International Conference on System Sciences.

Jones, Q., G. Ravid, et al. (2004). "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration." Information Systems Research **15**(2): 194-210.

Karau, S. J. and K. D. Williams (1993). "Social Loafing: A Meta-Analytic Review and Theoretical Integration." Journal of Personality and Social Psychology **65**(4): 681-705.

Kelly, J., D. Fisher, et al. (2005). Debate, Division, and Diversity: Political Discourse Networks in USENET Newsgroups. Online Deliberation 2005 / DIAC-2005, Stanford, CA.

Kim, A. J. (2000). Community Building on the Web: Secret Strategies for Successful Online Communities. Berkeley, CA, Peachpit Press.

King, B. and J. Borland (2003). Dungeons and Dreamers: The Rise of Computer Game Culture, McGraw-Hill.

Kollock, P. and M. Smith (1996). Managing the Virtual Commons: Cooperation and Conflict in Computer Communities. Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives. S. Herring. Amsterdam, John Benjamin.

Konstan, J. A., B. N. Miller, et al. (1997). "Applying Collaborative Filtering to Usenet News." Communications of the ACM **40**(3): 77-87.

Lampe, C. and E. Johnston (2005). Follow the (Slash) dot: Effects of Feedback on New Members in an Online Community. International Conference on Supporting Group Work, GROUP '05, Sanibel Island, FL, ACM Press.

Lampe, C. and P. Resnick (2004). Slash(dot) and burn: distributed moderation in a large online conversation space. Conference on Human Factors in Computing Systems (CHI), Vienna, Austria, ACM Press.

Landauer, T. (1991). Let's Get Real: A Position Paper on the Role of Cognitive Psychology in the Design of Humanly Useful and Usable Systems. Designing Interaction. J. Carroll, Cambridge University Press. **60-73**.

Lave, J. and E. Wenger (1993). Situated Learning: Legitimate Peripheral Participation. New York, Cambridge University Press.

L-Soft International, I. (2002). List Owner's Manual for LISTSERV, version 1.8e, L-Soft International, Inc.

Mackay, W. E. (1991). Triggers and barriers to customizing software. Conference on Human Factors in Computing Systems, New Orleans, Louisiana, ACM Press.

MacLean, A., K. Carter, et al. (1990). User-tailorable systems: pressing the issues with buttons. Conference on Human Factors in Computing Systems, Seattle, WA, ACM Press.

Madden, M. and L. Rainie (2003). America's Online Pursuits: The changing picture of who's online and what they do. Washington D.C., Pew Internet & American Life Project.

Malda, R. (2003). Slashdot FAQ, Slashdot.org. **2003**.

March, J. G. and H. A. Simon (1958). Organizations. New York, John Wiley.

Matei, S. A. (2005). "From Counterculture to Cyberculture: Virtual Community Discourse and the Dilemma of Modernity." Journal of Computer-Mediated Communication **10**(3).

McPhail, C. (1991). The myth of the madding crowd. Hawthorme, NY, Walter de Gruyter, Inc.

Middleton, S. E., N. R. Shadbolt, et al. (2004). "Ontological user profiling in recommender systems." ACM Transactions on Information Systems (TOIS) **22**(1): 54-88.

Miller, G. A. (1956). "The Magical Number Seven, Plus or Minus Two: Some limits on our capacity for processing information." Psychological Review **63**(2): 81-97.

Miller, G. A. (1962). Information Input Overload. Conference on Self-Organizing Sytems, Washington, Spartan Books.

Mockus, A., R. T. Field, et al. (2002). "Two Case Studies of Open Source Software Development: Apache and Mozilla." ACM Transactions on Software Engineering and Methodology **11**(3): 309-346.

Morris, M. (1993). E-mail editors: gatekeepers or facilitators? Annual Meeting of the Association for Education in Journalism and Mass Communication, Kansas City.

Muramatsu, J. and M. Ackerman (1998). "Computing, Social Activity, and Entertainment: A Field Study of a Game MUD." The Journal of Collaborative Computing **7**: 87-122.

Olson, G. M. and J. S. Olson (2001). Distance matters. HCI in the New Millennium. J. Carroll. New York, Addison-Weasley.

Olson, G. M. and J. S. Olson (2002). Groupware and computer-supported cooperative work. The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. J. A. Jacko and A. Sears, Lawrence Erlbaum Associates, Inc.

Page, S. R., T. J. Johnsgard, et al. (1996). User customization of a word processor. Conference on Human Factors in Computing Systems, Vancouver, British Columbia, ACM Press.

Palme, J. (1995). The optimal group size in computer mediated communication. Electronic Mail. London, Artech House Publishers.

Perkins, D. N. and G. Salomon, Eds. (1992). Transfer of Learning. International Encyclopedia of Education. Oxford, UK, Pergamon Press.

Pfaffenberger, B. (2002). A Standing Wave in the Web of Our Communications: Usenet and the Socio-Technical Construction of Cyberspace Values. From Usenet to CoWebs: Interacting with Social Information Spaces. C. Lueg and D. Fisher. New York, NY, Springer Verlag.

Poor, N. (2004). "Mechanisms of an online public sphere: The website Slashdot." Journal of Computer-Mediated Communication **10**(2).

Powazek, D. M. (2002). Design for Community: The Art of Connecting Real People in Virtual Places. Indianapolis, New Riders.

Preece, J., B. Nonnecke, et al. (2004). "The top five reasons for lurking: improving community experiences for everyone." Computers in Human Behavior **20**: 201-223.

Quarterman, J. S. and J. C. Hoskins (1986). "Notable Computer Networks." Communications of the ACM **29**(10): 932-971.

Rainie, L. and J. Horrigan (2005). Internet: The Mainstreaming of Online Life. Washington D.C., The Pew Internet & American Life Project.

Raymond, E. S. (2000). The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary, O'Reilly & Associates.

Raymond, E. S. (2004). The Jargon File. **2004**.

Reid, E. (1999). Hierarchy and power: Social control in cyberspace. Communities in Cyberspace. M. A. Smith and P. Kollock. London, Routledge**:** 107-133.

Resnick, P. (2001). Beyond Bowling Together: SocioTechnical Capital. HCI in the New Millenium. J. Carroll, Addison-Wesley.

Resnick, P., N. Iacovou, et al. (1994). GroupLens: an open architecture for collaborative filtering of netnews. ACM conference on Computer Supported Cooperative Work, Chapel Hill, NC.

Resnick, P. and H. Varian (1997). "Special Issue on Recommender Systems." Communications of the ACM **40**(3).

Resnick, P., R. Zeckhauser, et al. (2000). "Reputation Systems." Communications of the ACM(Dec).

Rheingold, H. (2000). The Virtual Community: Homesteading on the Electronic Frontier. Cambridge, MA, MIT Press.

Rheingold, H. (2003). Smart Mobs: The Next Social Revolution. Cambridge, MA, Perseus Publishing.

Ritchie, D. M. and K. Thompson (1974). "The UNIX time-sharing system." Communications of the ACM **17**(7): 365-375.

Sack, W. (2000). "Conversation map: An interface for very large-scale conversations." Journal of Management information Systems **17**(3): 73-92.

Sandusky, R. and L. Gasser (2005). Negotiation and the Coordination of Information and Activity in Distributed Software Problem Management. ACM-GROUP, Sanibel Island, FL.

Schafer, J. B., J. A. Konstan, et al. (2002). Meta-recommendation Systems: User-controlled Integration of Diverse Recommendations. 11th International Conference on Information and Knowledge Management (CIKM 2002), McLean, VA.

Schein, A. I., A. Popescul, et al. (2002). Methods and metrics for cold-start recommendations. ACM Conference on Research and Development in Information Retrieval, Tampere, Finland, ACM Press.

Shardanand, U. and P. Maes (1995). Social information filtering: algorithms for automating "word of mouth". Conference on Human Factors in Computing Systems, Denver, CO, ACM Press.

Simon, H. A. (1996). Sciences of the Artificial. Cambridge, MA, MIT Press.

Smith, A. D. (1999). Problems of conflict management in virtual communities. Communities in Cyberspace. M. Smith and P. Kollock. London, Routledge**:** 134-163.

Smith, M. (2002). Measures and Maps of Usenet. From Usenet to CoWebs: Interacting with Social Information Spaces. C. Lueg and D. Fisher. New York, NY, Springer Verlag.

Smith, M. A. (1999). Invisible crowds in cyberspace: Mapping the social structure of the Usenet. Communities in Cyberspace. P. Kollock and M. A. Smith. London, Routledge. **195-219**.

Smith, M. A. and A. T. Fiore (1991). Visualization components for persistent conversations. SIGCHI conference on Human factors in computing systems, Seattle, WA, ACM Press.

Sproull, L. and S. Kiesler (1991). Connections: New ways of working in the networked organization. Cambridge, MA, MIT Press.

Terveen, L. and W. Hill (2002). Beyond recommender systems: Helping people help each other. HCI in the New Millennium. J. M. Carroll. New York, Addison-Wesley.

Turkle, S. (1995). Life on the Screen: Identity in the Age of the Internet. New York, Simon & Schuster.

Viegas, F. B. and M. Smith (2004). Newsgroup Crowds and Authorlines: Visualizing the Activity of Individuals in Conversational Cyberspaces. 37th Hawaii International Conference on System Sciences, Honolulu, HI, IEEE.

Viegas, F. B., M. Wattenberg, et al. (2004). Studying cooperation and conflict between authors with history flow visualizations. Proceedings of the 2004 conference on Human factors in computing systems, Vienna, Austria, ACM Press.

Wasko, M. M. and S. Faraj (2000). ""It is what one does": why people participate and help others in electronic communities of practice." Journal of Strategic Information Systems **9**: 155-173.

Wellman, B. (1997). An Electronic Group is Virtually a Social Network. Culture of the Internet. S. Kiesler. Hillsdale, NJ, Lawrence Erlbaum.

Wellman, B. and M. Gulia (1999). Net Surfers Don't Ride Alone: Virtual Community as Community. Networks in the Global Village. B. Wellman. Boulder, CO, Westview Press.

Wenger, E. (1998). "Defining Communities of Practice." System Thinker(June).

Whittaker, S., L. Terveen, et al. (1998). The Dynamics of Mass Interaction. Computer-Supported Cooperative Work, Seattle Washington, ACM.