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

Cliff Lampe, Paul Resnick
School of Information
University of Michigan
304 West Hall
Ann Arbor, MI 48109-1092, USA
cacl@umich.edu, presnick@umich.edu

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, 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.

Author Keywords
Computer-mediated communication, collaborative filtering, recommender systems.

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.

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 [9]. 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[11,17]. 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 [4, 8], one indicator of user dissatisfaction.

CHI 2004, April 24–29, 2004, Vienna, Austria.

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 [13]) 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 [15].

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 [6], 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 [12,14,16].

A theoretical investigation of incentives for provision of evaluations [1] 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.

The commercial website 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.

SLASHDOT

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. Paid editors select about two dozen news stories each day, 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. The median number of comments per story in 2003 was 257, although some received 1000 or more. 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.

Part of the ethos of Slashdot is that posts are not deleted from the database, though they may not be shown to all readers. The site creators mandated that anonymous posting be allowed: *"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." [10]* To cope with the behavioral problems that occur in large scale conversations, especially given anonymous posting, and to help readers avoid information overload, Slashdot developed a moderation system to rate the worth of comments. To make the system more "democratic" and to relieve burden on centralized staff, Slashdot distributed the moderation system to its user base.

Each posted comment has a current score, from –1 to +5. Initial scores range from -1 to +2, with the default set at +1. Posts from Anonymous Cowards start at 0. Users achieve reputation, or "karma", though a number of activities, including moderating comments, reading comments and posting comments that get high or low scores. Comments from users with especially high karma can start with a score of +2, and comments from users with especially low karma can start at 0 or –1.

A moderator reads as he or she normally would but can click to moderate any comment up or down from its current score. A moderator chooses from a list of descriptors for the comments, such as "Offtopic", "Troll", "Insightful", "Funny", or "Overrated", each corresponding to a -1 or +1 moderation. The official guidelines encourage moderators to "*concentrate more on promoting, rather than on demoting*." [10]

Slashdot users achieve moderator eligibility by having high karma. A moderator is given five moderation points at a time, to be used within three days. Slashdot assigns moderation points based on the number of comments in the system, so there is some scarcity of moderation points available and not all comments can end up with +5 scores. Paid staff editors have an unlimited number of moderation points.

To "*remove bad moderators from the M1 (moderator) eligibility pool and reward good moderators with more delicious mod points*" [10], Slashdot developed a meta-moderation system. Meta-moderators are presented with a set of moderations that they then rate as either "fair" or "unfair". 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.

Readers can use the scores associated with comments to guide their reading in several ways, including sorting and filtering. Slashdot's default presentation of content is as a threaded list, showing all top-level comments rated +1 or above and response comments lower in threads being displayed if they are rated +4 or above. Users may change these defaults in their preferences, change them dynamically for a single session, or click to see responses to particular comments even if they are below the threshold.

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 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.

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.

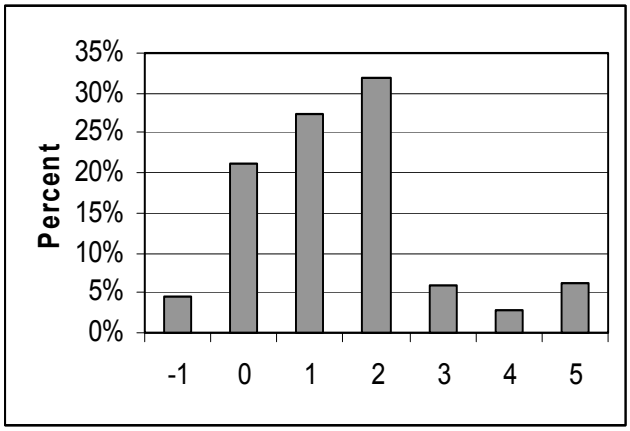

Figure 1: Distribution of final comment scores.

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.)

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

Table 1: Time to reach benchmark scores.

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 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.

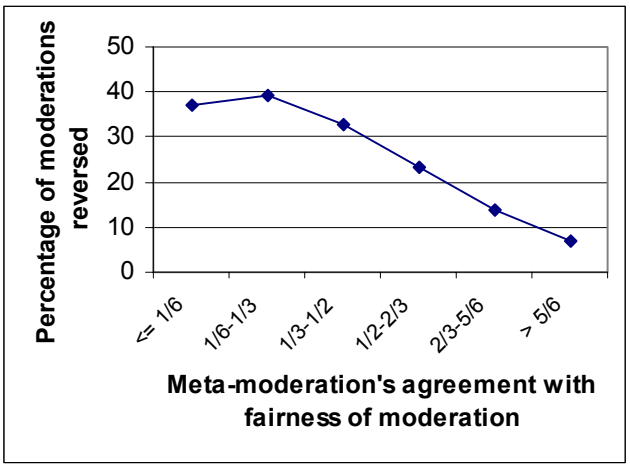

Figure 2: Moderation reversals

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 [1]. 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 [3, 2].

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 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

		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

Table 2: Initial and final comment scores

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 3 shows correlations among the variables of interest. We then controlled for the potential confounds in regression analyses.

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 [16]. Anonymous posting is correlated with lower starting scores at Slashdot, since all anonymous posts start with a score of 0. As shown in Table 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.

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 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 [7] 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.

	Modded	Starting score	Final score	Karma	Short com't	Long com't	Anon user	Top level	Early in thread	Late in thread
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

Table 3: Correlations of characteristics and outcomes

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 4 and 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 reports a logistic regression predicting the binary outcome of whether a comment will be moderated: positive coefficients indicate higher probabilities. Table 5 reports an ordinary least squares regression predicting the final score: positive coefficients indicate higher predicted scores. 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.

The R-squared measure of fit for the predictions of final score was only .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.

First, we consider the delay until receiving the first moderation for a comment. Since this measure considers only comments that do receive a moderation, it should be independent of the quality of the comments and reflect only the amount of attention from moderators. Table 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.)

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

Table 4: Logistic regression predicting if a comment will be moderated.

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

Table 5: Ordinary least squares regression predicting final comment scores.

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

Table 6: Lower scoring comments took longer to receive first moderation

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 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).

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

Table 7: Errors were corrected less frequently for comments with lower scores

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, 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. 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.

ACKNOWLEDGMENTS

This work was made possible by support from the Kellogg Foundation, and NSF Grant IIS 0308006. We would also like to thank the Slashdot/OSDN team who helped provide and understand data: Rob "CmdrTaco" Malda, Jonathan "CowboyNeal" Pater, Jamie McCarthy, Nathan Oostendorp and Rob "Samzenpus" Rozeboom.

REFERENCES

1. Avery, C., P. Resnick, and R. Zeckhauser, The Market for Evaluations. *American Economic Review*, 1999. 89(3): p. 564-584.
2. Banerjee, A., A Simple Model of Herd Behavior. *Quarterly Journal of Economics,* 1992. 107(3): p. 797-818.
3. Bikhchandani, S., D. Hirshleifer, and I. Welch, A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades. *Journal of Political Economy, 1989*. 100(5): p. 992-1026.
4. Butler, B., When is a group not a group: An empirical examination of metaphors for online social structure. 1999, Carnegie Mellon University: Pittsburgh.
5. Friedman, E.J. and P. Resnick, The Social Cost of Cheap Pseudonyms. *Journal of Economics and Management Strategy,* 1997. 10(2): p. 173-179.
6. Goldberg, D., et al., Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 1992. 35(12).
7. Grice, H.P., Utterer's meaning and intentions. *Philosophical Review,* 1969. 78: p. 147-177.
8. Jones, Q., G. Ravid, and S. Rafaeli. An empirical exploration of mass interaction system dynamics: Individual information overload and Usenet discourse in *35th Hawaii International Conference on System Sciences. 2002*.
9. Kollock, P. and M. Smith, Managing the Virtual Commons: Cooperation and Conflict in Computer Communities, in *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*, S. Herring, Editor. 1996, John Benjamin: Amsterdam.
10. Malda, R., Slashdot FAQ, . 2003, http://slashdot.org
11. Pfaffenberger, B., A Standing Wave in the Web of Our Communications: Usenet and the Socio-Technical Construction of Cyberspace Values, in *From Usenet to CoWebs: Interacting with Social Information Spaces,* C. Lueg and D. Fisher, Editors. 2002, Springer Verlag: New York, NY.
12. Resnick, P., et al. GroupLens: an open architecture for collaborative filtering of netnews. In *ACM conference on Computer Supported Cooperative Work. 1994.* Chapel Hill, NC.
13. Sack, W., Conversation map: An interface for very large-scale conversations. *Journal of Management information Systems, 2000. 17*(3): p. 73-92.
14. Shardanand, U. and P. Maes. Social information filtering: algorithms for automating "word of mouth" in *SIGCHI conference on Human factors in computing systems. 1995.* Denver, CO.
15. Smith, M.A. and A.T. Fiore. Visualization components for persistent conversations. in SIGCHI *conference on Human factors in computing systems.* 1991. Seattle, WA: ACM Press.
16. Terveen, L. and W. Hill, Beyond recommender systems: Helping people help each other, in HCI in the New Millennium, J.M. Carroll, Editor. 2001, Addison-Wesley: New York.
17. Whittaker, S., et al. The Dynamics of Mass Interaction, in *Proc. of Computer-Supported Cooperative Wor*k. 1998. Seattle Washington: ACM.