Comments Gone Wild: Trolls, Flames, and the Crisis at Online Newspapers

Daniel Drew Turner
School of Information, University of California, Berkeley
102 South Hall
Berkeley, CA 94720-4600
1.415.606.4811
ddt@ischool.berkeley.edu

ABSTRACT

As traditional media outlets add forums for community participation and feedback, they invariably attract trolls - unruly net denizens that spam, flame, and generally detract from the sense of community, free exchange of information, and overall readership. The actions of such trolls highlight a tension at the heart of new media business strategy: how to simultaneously encourage constructive participation, while discouraging abusive and destructive behavior by this population and the unintentional compounding by otherwise good actors. On one extreme, sites such as 4chan remove all bars to participation in selected forums; on the other, newspapers such as *The Washington Post* and over 100 others have shut down entire threads in the last year due to overwhelming flame wars sparked and fanned by trolls. In this paper, we discuss the financial and existential tradeoffs involved in this decision, and describe existing strategies that attempt to control abuse. Through an analysis of these, and the sociological underpinnings of synchronous and asynchronous computer-mediated communication, we demonstrate that existing technological counters to trolls, though elaborate and expensive, are not viable long-term solutions. We also make the case why simply abandoning the problematic feature of comments is not an option for newspapers. We conclude with a discussion of the motivations of those who troll and point to directions for future work.

1.INTRODUCTION

Over half the newsrooms that responded in a 2009 survey by the nonprofit American Society of News Editors (ASNE) reported having to shut down at least one online forum within the previous year. This represents the most dramatic failure point. Yet commenting and user interaction has (as we will show) too high intrinsic and extrinsic values to newspaper publishers for them to simply abandon these problematic systems altogether.

And even publishers who have not been driven to the point of shutdown are facing ongoing issues with online forums and story-connected comments at a crisis level.

As of 2006, approximately 80 percent of online sites for newspapers included some kind of comment system for news stories, or allowed users to submit user feedback of some kind [1]. In 2009 the ASNE surveyed over 1,000 newsrooms across the United States. Out of 276 respondants, 87.6 percent "invite online comments regarding specific stories" [2].

But 38.9 percent also reported shutting down at least one comment thread for a specific story within the previous year and 46.2 percent said they had to ban users within the same period. These are significant numbers, both at the macro (total number of troublemakers) and at the micro (negative effects to user experience) levels.

The immediate and ongoing result is damage to the user experience, and damage to the news outlet that unwittingly hosts such content.

"We've lost subscriptions over the comments," said Michael Freimann, the online editor of The Pantagraph in Bloomington, Ill, which saw "overwhelming" abuse of the system. "We've been bad-mouthed on the radio," he added. Freimann was quoted by Kurt Greenbaum in his online article "Reader comments online: have we lost control?"[1] . Soon after instituting an online comment system, the News & Observer or Raleigh, N.C. saw " the arrival of spam, profanity, harassment, and the need to spend time each day deleting inappropriate comments" [4]. An editorial for the Online Journalism Review outlined the editorial attitude of the Los Angeles Times towards reader comments as "shut up"[2]. The New York Times featured a story outlining how many news sites such as their own, washingtonpost.com, and others, were "rethinking" comments[3].

(Please note: For the purposes of this paper, the terms “comments", “comment systems”, "comment threads", and the like will be considered to include systems that allow readers to attach their comments to a single online story, or post comments in a forum. All present the same problems for publishers and exhibit highly similar user behaviors.)

In this study, we will analyze the state of and stake in comments on newspaper web sites, survey and analyze existing approaches to controlling abusive users and flame wars, and apply sociological research on community and identity towards recommendations for future commenting systems and research.

---

1 http://tae.asne.org/StoryContent/tabid/65/id/123/Default.aspx

2 http://www.ojr.org/ojr/stories/070817niles/

3 http://www.nytimes.com/2010/04/12/technology/12comments.html

2.PROS AND CONS OF COMMUNITY PARTICIPATION

Newspapers have a deep investment in publishing readers' voices. The long history of Letters to the Editor has made community and exchange part of the definitional identity of a newspaper. The sense of being able to communicate with the writers and editors also is a key factor in user trust, which is a prime selling point of any news source. Including readers' comments can also be key to the small-d democratizing existential purpose of the modern newspaper.

2.1Remaining True to the Cultural DNA

"Trust is, at its heart, an *interpersonal* phenomenon," wrote Daryl Koehn in *The Journal of Business Ethics*. In this context, we are combining Worchel's 1979 concepts of **situational trust** (people adjust their tendencies in response to situation cues, such as the quality and amount of communication between parties) and **learned trust** (the experience parties have gained from past situational trusts) to obtain a fairly everyday sense of the word, and to avoid the pitfall of requiring that trust be between human actors only – in this case, human users can certainly feel there is agency in the other party (the newspaper), or at least can see that there are humans behind it.

In the case of web sites for established newspapers – the scope of this paper – there is a great deal of learned trust based on an individual paper's reputation (such as, "All the News that's Fit to Print") that carries over to the web version, though a new set of situational cues can alter that base.

Part of this trust grows from some form of bottom-up communication with the newspaper, whether in print or online; there the aspect of "quality and amount of communication" grows in importance. Offering interactive comments serves to boost an online news outlet's credibility with readers. "Interactivity and transparency have vast implications for the elevation of credibility on Internet news sites," wrote Martha Stone, a 2001 Poynter Ethics Fellow and Co-director of an Online News Association Web Credibility Study.

Modern newspapers, at least in the scope of this paper, define themselves as small-d democratic – that is, not an organ of established power. A quantitative study quoted by Christopher Kedzie, who worked for the Ford Foundation in Moscow, found that the "correlation between interconnectivity and democracy is positive" [4]. What can be called this democratizing influence is not limited to electronic media – television, radio, and the printed word have served. But a single post on the web has a much lower barrier to entry: it can be generated with a hand crank and a satellite phone, from an anonymous source, and appear attached to a front-page article on nytimes.com. This lends strong support to a connection between comments being open to users and the mission of news organizations.

2.2The Sociological Research of Community and its Application to our Scope

It may not be the first thing one thinks of when looking at, say, an international newspaper, but media outlets rely on both actual community (anyone can write to the editors, anyone could appear in or tip off reporters to a news story) and that the readers perceive and act with a sense of community – especially in the realm of online comments. In this section we will examine existing sociological research on the topic of community: how awareness of social capital, the nature of ties, and virtual communities can influence how users react when commenting online.

The study of what people exchange online and why has a strong trail in sociological theory. Perhaps this is a function of a consumerist society, but any interaction, even trivial speech, can be seen as a transaction in theory (note that it is called "exchanging pleasantries"), and these transactions can form a kind of economic basis for a virtual community. This basis is social capital, which is as a term has had some range of definitions, but will be considered in this paper the value of an individual's social networks and how much people within those networks are inclined to do things for each other.

Social capital in online contexts that value information (whether it be factual, opinion, or recommendations) can simply be the public demonstration of knowledge, and the sharing thereof. Ridings and Gefen [5] synthesized previous studies (Binik, Cantor, Ochs, & Meana, 1997; Hiltz & Wellman, 1997; Rheingold, 1993a; Sproull & Faraj, 1997) to be able to say, "Knowledge and information are, in general, a valuable currency or social resource in virtual communities."

Similarly, Donath [6] said, "Individual recognition is important in many newsgroup" (remember that above, we showed that we can apply Usenet-centric studies to this paper's scope) and, "On-line status is recognized and there is deferral to respected members" based on those members' previous display of knowledge." This display is "those members'" building of social capital.

Much of what has been written about exchange and social capital has been within the context of goods and services transactions, which can prove problematic to the public goods context of this paper (or even impossible to translate). But the idea of social capital seems to be a reliable measure in both contexts, motivating those participating in an eBay trade as well as a comment thread.

Also related to this context is the idea of **strong** and **weak ties**.

Mark Granovetter, in his seminal paper, devised the idea of relative measuring of interpersonal ties, dividing "strong" into basically "people you really trust" while "weak" are "merely acquaintances." While this alone seemed commonsense, Granovetter went on to show how weak ties were "indispensable to individuals' opportunities and to their integration into communities" [7].

The relevance here is that almost all relationships within the scope of this paper – that is, individuals interacting under usernames/ aliases/online personae through comments – form attachments that are almost the definition of weak ties. In fact, the relative anonymity of online comments, while leveling much of social, regional, sex, and age differences, may strengthen their relationships. In fact, "the reduction of social cues makes it far more difficult to develop the intimacy and confidence necessary to deepen relationships. Therefore, the Internet is more conducive for the development of weak ties rather than strong ties (e.g., Bargh & McKenna, 2004; Blanchard & Horan, 1998; Haythornthwaite, 2002)" [8].

Ridings and Gefen [5] also stated that what "makes virtual communities special in this regard as compared, for example, with

traditional social groups is the magnitude and impact of "weak ties," i.e., relationships with acquaintances or strangers to obtain useful information through online networks (Constant, Sproull, & Kiesler, 1996). A virtual community can be an ideal place to ask relative strangers about information."[4]

Best and Krueger also said that "some argue that these online social interactions meet the conditions necessary to facilitate the production of social capital (e.g., Ester & Vinken, 2003; Hill & Hughes, 1997; Rheingold, 1993)… the Internet offers opportunities for users to develop personal ties with others, even a shared sense of collective identity (Rheingold, 1993; Walthier, 1995)" and "individuals more actively pursuing and maintaining weak ties typically possess greater levels of social capital than those limiting their interactions to strong ties" [8].

And their empirical evidence, they claimed, did in fact "offer generalizable empirical evidence in support of the positive view of online relations; indicators of social capital positively relate to the level of interaction with people met on the Internet. … Although online social interactions likely do not produce strong connections that elicit intense loyalty, these results do suggest that they foster connections critical to expanding networks and producing residuals such as generalized trust" [8].

So news stories online provide a qualitatively different and more social capital building environment than topic-centric Usenet newsgroups, as people of many social and attitudinal stripes will visit and comment on a news story (overlooking the "edge cases" highly homophilic and/or partisan news and opinion sites, which are not in the scope of this paper).

Social capital is also strongly related to what Ridings and Gefen call in virtual communities "the social support that the community can provide. Social support is 'the degree to which a person's basic social needs are gratified through interaction with others (Thoits, 1982, p. 147). Social support may also be linked with individual motivation to join groups because of the sense of belonging and affiliation it entails (Watson & Johnson, 1972…. House (1981) offers a more specific definition of social support: a flow of emotional concern, instrumental aid, information, and/or appraisal (information relevant to self-evaluation) between people (p. 26)" [8]. We take this "appraisal" to be, in part, the fact that other members of the community "listen" to the user's comments.

Supporting that, is that "Herring [(1996)] found that the freedom to express views and to receive social support were the main reasons individuals joined and used virtual communities. Her study of two email distribution lists found that people participated to exchange opinions, beliefs, understandings, and judgments though a social interaction with others, but where the pure exchange of information took on a secondary role" (Ridings and Gefen).

In addition, in a study of online news media in China, Hong Kong, and Taiwan during the SARS crisis of 2002-2003, Alice Lee showed that online news sites can build community and interconnection amongst readers, though to different degrees in the three countries. Lee found the sites "were also capable of providing social linkage and social amusement" and could "facilitate communication and emotion sharing among members of the community" [10].

From this, we can conclude: most active users in online comments are motivated to increase their own social capital and maintain a network of weak ties. Feedback on these efforts, or lack of efforts, should be quite evident to the user: as capital and networks grow stronger, the more responses the user should see to comments and requests. How to entice flamers to "play nice" is a question that has not been studied in this context – but it is nontrivial to discover that these metrics of social capital and weak networks, which can and often are measure quantitatively, can be used to identify and potentially isolate abusive commenters.

2.3 It's Not Just a Good Idea, it's the Law: Legal Perspectives on Anonymous Comments

Free speech in comment threads on newspaper sites may not be protected speech and protected by the First Amendment – the sites explicitly discussed in this paper are not government organs[5] – but there is a strong legal tradition in the U.S. supporting the right to anonymous communications.

In 1995 the Supreme Court ruled, in *McIntyre v. Ohio Elections Commission*:

"Protections for anonymous speech are vital to democratic discourse. Allowing dissenters to shield their identities frees them to express critical, minority views . . . Anonymity is a shield from the tyranny of the majority. . . . It thus exemplifies the purpose behind the Bill of Rights, and of the First Amendment in particular: to protect unpopular individuals from retaliation . . . at the hand of an intolerant society."[6]

And in the 1997 case *Reno v. American Civil Liberties Union*, which struck down anti-indecency provisions of the Communications Decency Act, the Court reinforced its findings that anonymous speech was of value on the Internet, where any user could be "a town crier with a voice that resonates farther than it could from any soapbox."[7]

Aside from legal issues, the evidence so far suggests that users, even non-abusive ones, value some degree of anonymity (or pseudonymity).

---

4 This is why I feel we can put aside the question of Dunbar's Number.

5 " Private website operators and hosting services can control what kind of speech appears on their site and servers." – from http://www.citmedialaw.org/legal-guide/legal-issues-consider-when-getting-online

6 http://www.law.cornell.edu/supct/html/93-986.ZO.html

7 http://caselaw.lp.findlaw.com/cgi-bin/getcase.pl?court=us&navby=case&vol=521&invol=844

In one case, *The Buffalo News* announced that of August 2, 2010, it would no longer post anonymous comments. This received "fast and furious"[8] responses from readers, who were overwhelmingly opposed; many threatened to cancel subscriptions.

Though no studies have been done to catalogue and quantify the reasons, some responses by users can be seen as reasonable and representative.

The game company Blizzard announced mid-2010 that it would force users of their official online forums to use their real names. Almost immediately a widespread protest arose from users, many arguing that it would open game players (who, remember, were facing off in virtual battle) to harassment and stalking in real life – an issue especially acute to female gamers.[9] Blizzard quickly withdrew the policy.

Similar concerns hold for commenting systems, the users of which may not want to be vulnerable to being tracked down after a heated exchange of messages.

2.4 Comments Equal Cash: The Economic Case for Maintaining Commenting Systems

Sites also have a vested economic interest in keeping comment systems alive and active.

First, and most simply, adding a comment system to a news site generates revenue. Almost all newspaper sites have a revenue model that includes online ad displays; rates charged advertisers are higher on sites with more page views and unique visitors (so far, subscription and membership fees have not proved viable[10]). At the most immediate level, a user adding a comment forces a reload of the site's page. The more users posting comments, the more page views from the same size audience. And it is standard practice that higher viewership drives higher ad rates.

In the case of washingpost.com, the ombudsman Andrew Alexander wrote in his April 4, 2010 column that "The growth [in online comments] is critical to The Post's financial survival in the inevitable shift from print to online."[11]

Second, allowing comments can foster a sense of community among users, and between the users and the publication. Ridings, Gefen, and Arinze define "virtual community" as "groups of people with common interests and practices that communicate regularly and for some duration in an organized way over the Internet through a common location or mechanism" [11]. And Herring uncovered that the freedom to express views (as in: leaving a public comment) was one of the two main reasons people joined virtual communities; the offering of comment threads is a service of value provided to the users, who can "seed" threads, encouraging more users, which in turn drives site revenues.

Because the service of offering comment threads is of value, sites can ask for an exchange from the users. Many, as terms of service, ask from users their email address[12], sex, age, and/or ZIP code, and sell this aggregated, anonymous data ("In addition to personally identifiable information, we also collect certain non-personally identifiable information through technology and tools, including cookies, Web Beacons and log data. We aggregate and analyze this information in order to learn more about how our Web sites are used."[13]).

And Ash, Hettinga, and Halpern found in a 2009 quantitative study that "[i]f a news Web site wants readers to enjoy visiting the site, and continue to return, allowing comments is a good way to promote this sense of enjoyment" [12]. Oddly, that they also found that the presence of comments seemed to reduce the perception of quality of the journalism; they hope to study this further.

2.5 The Practical Problems

Donath characterized trolls as employing "pseudo-naîve" tactics. The troll would join a newsgroup or other online discussion. pretending to ask obvious and perhaps stupid questions, waiting to see who would take the question at face value, then ridiculing those who fell for it. However, Internet activities and opportunities have expanded since this 1980s definition – and in newspaper sites, trolls often are attracted to topical political stories, and use the above plus abusive tactics against those they see as on the opposite side. Their actions serve as much as a derailing as a mocking[14].

Shachaf and Hara, in their study of how trolls affect management of the Hebrew edition of Wikipedia, categorized trolls as being "engaged in intentional, repetitive, and harmful actions" and work through "hidden virtual identities." That is, they take advantage of lack of any tie to a real-world identity to repeatedly introduce abuse and/or derail conversation. This built on how Herring et al. [12] defined trolls: as aiming "to disrupt the ongoing conversation", usually with provocation, insults, and comments that may be deceptive but not blatantly so. These types of comments are more difficult for human or automated systems to catch.

In a 2009 study in four Norwegian municipalities, Winsvold compared comments and forums in online newspapers to the Letters to the Editor in their print counterparts. He found the online comments to be of "low quality" and the commenters to have dubious motivations while they seemed have "participated for fun or because they liked a good fight" [13].

In terms of quantity, Winsvold also found that "[a]nother major challenge to the position of online communication arenas was the

---

8 http://www.buffalonews.com/editorial-page/columns/margaret-sullivan/article89363.ece

9 The web comic Penny Arcade had its own and NSFW take on the issue: http://www.penny-arcade.com/comic/2010/7/9/

10 http://www.niemanlab.org/2009/04/paying-for-online-news-sorry-but-the-math-just-doesnt-work/)

11 http://www.washingtonpost.com/wp-dyn/content/article/2010/04/02/AR2010040202324.html

12 Greenbaum, Kurt. "You've got revenue." American Editor 82.4 (2007): 24. Academic Search Complete. EBSCO. Web. 28 Apr. 2010

13 http://www.nytimes.com/ref/membercenter/help/privacy.html

14 Schwartz, M. (2008, August 3). The trolls among us. The New York Times.

huge volume of contributions resulting from their inclusiveness.“ [13]. We will later discuss this in terms of moderation, but the relevance here is that traditional methods of dealing with potentially damaging user comments – that is, individual, in-house editorial consideration simply does not scale, especially for the traditionally resource-tight news industry.

In the more granular data obtained by the 2009 ASNE survey, we can see the primary reasons for shutdowns of story-related discussion threads. The top reported reason was "Discriminatory comments involving race, ethnicity, gender or sexual orientation", followed closely by "Hurtful comments not discriminatory in nature" and "Obscenities, profanities, foul language." Legal, fairness, and accuracy issues were far behind, suggesting that on-topic signal (user-generated content) is far overwhelmed, in terms of what is problematic, by off-topic noise.

Paradoxically, since the likelihood of abuse and flaming rises as the number of commenters goes up ("fruitful cooperation has proven to be difficult to sustain as the size of the collaboration increases" [15]), this means that the more attractive a comment system is, the less attractive it will be. "Why do people choose to join a virtual community? The most frequently cited reason in the literature is to access information (Furlong, 1989; S. G. Jones, 1995; Wellman et al., 1996), which is also a reason for group membership cited often by social psychologists (Watson & Johnson, 1972)" [5]. Too much abuse in the comments, the more noise and less signal, means a lower ability to access information. It's the Yogi Berra effect: nobody goes there any more – it's too crowded.

Abusive and non-abusive comments may not carry the same impact per comment. Cory Doctorow has speculated that one willfully contentious exchange can consume vastly more cognitive time and resources than any number of kind ones[15]. This ups the ante for controlling abuse.

In addition, Sudweeks and Rafaeli asserted that text-based computer-mediated communication can spur disagreements more often, lead to more extreme views expressed on hot-button issues, and be a bar to consensus, as compared to face-to-face communications. Imagine how heated a dinner conversation on the news of the day can get, then add this layer.

As we showed above, comment systems have high value in a range of contexts for newspapers. Yet, as the data from the ASNE survey indicate, many newsrooms have been forced to cut away the source of this value. And the content and frequency of industry press and seminars on the subject indicate that the newspaper industry is ready to pour scarce resources into this problem area (recently the washingtonpost.com and sfgate.com sites have stated they will undertake comment system overhauls, though they have not, as of this writing, finalized all the features).

3.SOLVING THE PROBLEM

As the web and comment systems are technological constructs, attempts at controlling or managing the problem have arisen mostly from a technological framework, with the addition of policy settings which are an outgrowth of past print newspaper social norms. The technological solutions include moderation, reputation/recommendation systems, and registration. We will show, however, that some of these are easily defeated, others are of the "shutting the barn door after the cows have escaped" type, others seem to have adverse affects on innocent users, and most are labor- and cost-intensive in an industry that cannot afford either.

(Some sites, such as message-board-based communities, use the idea of “splinter communities” – spinning off noisy or large threads – to deal with similar issues. However, as we are looking as comments related to news stories, this is not applicable.)

3.1Current Solutions

3.1.1Policy

Most, if not all, sites with comment systems have Terms of Service (TOS) that delineate not only what a user can expect from the site (conditions of privacy, etc.) but also what the user agrees to by using the site's features. It is key to note that the user need not sign anything, or register an identity at the site, or even have read the TOS in order to be held to them – it's an effective "you post, you've agreed" situation (for example, see the TOS for the Sacramento Bee's web site: http://www.sacbee.com/terms-of-service/). It remains a conundrum as to whether one is held to agreeing to the site's TOS if one visits the page listing the TOS in order to evaluate them.

The TOS for most newspaper sites include injunctions against much of what we above identified as a problem in online comments on their sites. The washingtonpost.com site demands agreement from users not to post "inappropriate" remarks, including those that are hateful or racist, or those that advocate violence. And here is a representative sample from the nytimes.com TOS:

"3. USER GENERATED CONTENT: SUBMISSIONS INCLUDING COMMENTS, READER REVIEWS, TIMESPEOPLE AND MORE

3.1 (a) You shall not upload to, or distribute or otherwise publish on to the Service any libelous, defamatory, obscene, pornographic, abusive, or otherwise illegal material.

3.1 (b) Be courteous. You agree that you will not threaten or verbally abuse other Members, use defamatory language, or deliberately disrupt discussions with repetitive messages, meaningless messages or "spam."

3.1 (c) Use respectful language. Like any community, the online conversation flourishes only when our Members feel welcome and safe. You agree not to use language that abuses or discriminates on the basis of race, religion, nationality, gender, sexual preference, age, region, disability, etc. Hate speech of any kind is grounds for immediate and permanent suspension of access to all or part of the Service.

3.1 (d) Debate, but don't attack. In a community full of opinions and preferences, people always disagree. NYTimes.com encourages active discussions and welcomes heated debate on the Service. But personal attacks are a direct violation of these Terms of Service and are grounds for immediate and permanent suspension of access to all or part of the Service."[16]

However, these policies have proven largely ineffective, even

–

1 le.jhtml;jsessionid=HOLITJ513XJC3QE1GHPCKHWATMY32JVN?

though all reserve the right, as in the nytimes.com TOS, to "terminate or suspend your access to all or part of the Service for any reason, including, without limitation, breach or assignment of these Terms of Service."

The reasons are twofold.

First, the sheer scale of incoming comments (over 320,000 a month at washingtonpost.com [17] is impossible to monitor by humans. The site's ombudsman wrote, "About 300 comments are deleted each day. But others slip through because The Post's staff of only a few monitors can't possibly scrutinize everything."[18] To date, there are no studies showing automated systems can effectively take up the slack; many problematic posts do not contain explicit keywords, or mask them by misspellings; it is difficult to judge when a commenter is enthusiastically and helpfully engaged and when harassing; most newspapers have the stance of preferring to err on the side of allowing speech.

Second, there is what Ford and Strauss called the "disposability" of online identity. As we will discuss more later, banned commenters can simply come back to comment systems, even ones that require "basic registration." Even if sites track IP addresses, it is not difficult for a dedicated troll to spoof or log on from another IP address.

As a result, TOS policies can give online publications justification for their efforts to police abuse, but these efforts can effectively be outmoded and insufficient.

These above are private policies, acting as if contract law between the publication and the commenter. As reported by Schwartz[19], various pieces of legislation have been proposed to combat online malevolence, including the Megan Meier Cyberbullying Prevention Act[20] (currently held at committee). Attempts have also been made to press federal fraud cases, but this is made rare both by questions as to whether these laws apply and by the unlikelihood of finding the real identity of abusers.

3.1.2All in Moderation

Moderators of online comments can be thought of as serving some, though not all, of the purposes "the editors" did in the context of Letters to the Editor in print versions. Online moderators should, at least in theory, vet submitted comments for problematic content and approve those that do not contain such while deleting those that do (and perhaps banning those commenters). Beyond theory, salon.com co-founder and former managing editor Scott Rosenberg wrote in 2010, "Show me a newspaper website without a comments host or moderation plan and I'll show you a nasty flamepit."[21]

Herring et al. cited Korenman and Wyatt, who in 1996 studied potentially contentious interactions within an email forum, and concluded that strong, explicit, and transparent moderation instituted from the inception of the forum can rein in many of the problems of trolling and flame wars. (Their study was relatively limited, and anonymity was not as difficult and complex an issue there as in our context.)

There are two types of moderators. "In-house" moderators are hired (full- or part-time) staff of the publication in question. They are trained and supervised by the publication, and responsible to that organization. User moderators are just that: users (perhaps veterans) of the service or publication who have earned, in some way, the right and responsibility to oversee the content of the comment system.

With in-house moderators, the process may not be transparent. That is, comments under review are not publicly displayed before approval; the in-house moderation process is not obvious to the average user, or to the person who made the comment under review. This has the advantage of not allowing all content into the public space by default, so those who would intentionally start flame wars do not have power to do so by default. The nytimes.com site is an example of this kind of system. One drawback for the user is the significant delay in the comment going up, and the even longer delay in seeing responses.

This delay in comments appearing could break, or be only a very weak formulation of, Rafaeli's "interactivity": the delay in comments appearing, which may depend arbitrarily on moderator workload, reduces the opportunity for messages to be in sequence and relate to each other. In addition, the experience of these delays could prove a disincentive to users, who may find the process not worth their time.

Diana Chung pointed to Rafaeli's 1997 definition of interactivity[22] as "responsiveness": "Rafaeli studied computer-mediated groups and the communication exchanges among asynchronous multi-participant public discussion groups. … He defines interactivity as 'the extent to which messages in a sequence relate to each other and especially the extent to which later messages recount the relatedness of earlier messages'"[15].

And this practice of in-house moderation is resource-intensive in a resource-starved industry. A top paper such as *The New York Times* can get away with posting only select comments moderated "by the authors and editors working with a blog, supplemented by a team of paid moderators whom we train and supervise,"[23]. However, this is not the case for most publications. Saundra

---

17 http://www.washingtonpost.com/wp-dyn/content/article/2010/04/02/AR2010040202324_2.html

18 ibid.

19 Schwartz, ibid.

20 http://thomas.loc.gov/cgi-bin/query/z?c111:H.R.1966:

21 http://www.salon.com/news/feature/2010/04/13/newspaper_online_comments_moderation_open2010

22 http://www.usc.edu/dept/annenberg/vol2/issue4/rafaeli.sudweeks.html

23 "long-time editor" for the Times, personal communication, May 9, 2010

Keyes, a journalism professor at the Reynolds School of Journalism, University of Nevada, Reno, and a columnist for ASNE.org., said, "Few newsrooms, of course, can devote the resources the Times devotes to achieving that goal through moderation"[24]. And as noted above, washingtonpost.com has "only a few monitors." (For more on the nytimes.com moderation policies, see http://www.nytimes.com/ref/membercenter/faq/comments.html and the wonderfully titled blog post "The Top 10 Reasons We Deleted Your Comment" at http://cityroom.blogs.nytimes.com/2007/11/15/the-top-10-reasons-we-deleted-your-comment/.)

Though it is beyond the scope of this paper, we did a casual survey of what online newsrooms offer in terms of in-house moderation. Most at the national and regional level, are along the lines of washingpost.com at best; a few have dedicated (if sometimes part-time) staffers monitoring comments at they come in, stepping in only to delete comments and ban users that violate their TOS. This still has the weakness of exposing troll content that could instigate reaction noise, but proactive moderating (blocking content before it appears online) reduces transparency, could raise censorship questions, and could break interactivity of comments.

In contrast, user or distributed moderation systems allow any user to "score" any other user's comments. This is best seen on the site Slashdot.org, which we will talk more about in terms of reputation systems. Lampe and Resnick studied the distributed moderation system of that site and found that there was a general success in "floating" the best comments and "burying" the worst, though they recorded that "much of a conversation can pass before the best and the worst comments are identified" [17]; this is due to the human time it takes for a significant amount of Slashdot.org users to read and evaluate each comment.

We would add that this type of system also does not address at all a critical problem of comment abuse. In a user or distributed moderation system, all comments are posted to the public – they have to be, for users to moderate them. Though they may eventually be buried, a skilled or willful troll can still pollute the discourse, and rapidly.

4chan.org, a notoriously noisy (intentionally so) message and imageboard site, has experimented with a Perl-based "moderator bot" called Robot9000[25]. Identity-linked controls are not available for 4chan, as the site is known for keeping a policy of anonymity for posters.

Robot9000 "mutes (-v) chatters for a period after every violation. The mute time starts at two seconds and quadruples with each subsequent violation, so you have five or six tries to get the hang of it. Your mute-time decays by half every six hours (we're still tweaking the parameters). When looking for matches, the bot ignores punctuation, case, and nicks." Impact was not well-defined; the threads Robot9000 oversaw even had more noise, as users started meta-discussions about the rules of the bot, and some tried to get their nicknames at the top of a public scoreboard – which was, ironically, posted in an attempt to shame repeat violators. A more manual analog is the SalonTrollBeGone script that relies on Firefox and the Grease Monkey addon; users who install it can manually add usernames to a kill or feature list. This, like the post-hoc user moderation systems seen above, still requires trolls to act and be identified, and does not give them feedback that they are not reaching their targets.

"Disemvoweling" is another technique that has been used by some sites to discourage trolls, though the larger context of shaming. Noted by *Time* magazine in 2008 as #42 of its "Best Inventions" of that year[26], the practice takes out vowels from the comments, so that, for example, "You are an idiot" becomes "Y r n dt". (Habitual texters may be able to make out some of this, but maybe not "dt".)

This relies on at times killing the meaning of a trolling post, as well as showing other commenters that the affected commenter was perceived as violating site conduct standards. Also, it's not quite moderating in the sense of agency, but the function of moving down a suspect comment in the hierarchy is similar.

However, there are shortcomings, ones perhaps insurmountable. Though this process can bar trollers from posting complex insults or lures, basic abuse can still get through even in vowel-free form. The action, whether by hand or automated, distances users from each other and from editors. There can be legal issues[27]. And counting on a troll being moved by peer disapproval is a long shot by any measure.

So, though existing moderation paradigms offer some limited promise and victories, they each have failed, as implemented, to stem the crisis.

3.1.3Reputation and Recommendations

Reputation and recommendation systems have become commonplace on web sites that specialize in enabling b2p, b2b, and p2p sales of goods and services (for the purposes of this paper, we will treat the two as functionally interchangeable; recommendations and ratings combine to form a reputation). Ratings and reputation metrics can be attached to users themselves (think of buyer and seller ratings through the Feedback Forum on ebay.com, retailer rankings on Google Product Search) or to user submissions (such as "Was this review helpful to you?"). Dan Byler pointed out that a common factor to commercial and non-commercial uses of recommender systems is that they serve to personalize and customize a user's online experience by taking feedback and narrowing the display of information [17].

Eric Goldman, Director of the High Tech Law Institute at Santa Clara University (http://law.scu.edu/hightech/), has defined **reputational information** as "information about an actor's past performance that helps predicts the actor's future ability to

---

24 http://www.igreenbaum.com/2009/04/19-qs-and-as-from-asnes-story-comment-webinar

25 Information at http://blog.xkcd.com/2008/01/14/robot9000-and-xkcd-signal-attacking-noise-in-chat/

26 http://www.time.com/time/specials/packages/article/0,28804,1852747_1854195_1854185,00.html

27 http://blog.timesunion.com/baumgartner/a-e-i-o-u-and-sometimes-why/1641/

perform or satisfy preferences"[28]. This functional sense of reputation is distinct from and unrelated to any shame or gossip connotations.

Two types of reputation systems come into play online, Goldman said.

First are **unmediated (or distributed) systems**, where users comment or rate directly, in a "word of mouth" or p2p mode. There are various formats for this online already. Slashdot.org has a mature system that assigns reputation scores both to users and individual comments. Registered users can rate the comments of other (registered or "Anonymous Coward") users, so visitors to the site can filter their view by minimum comment rating and see only the contributions that have been ranked to their standards. Users themselves are also scored for reputation. Starting with a "+1" score, users can increase their "karma" by posting comments that receive high scores from others, or by moderating comments. Or, in the case of ebay.com, "After a transaction is complete, the buyer and seller have the opportunity to rate each other (1, 0, or -1) and leave comments (such as 'good transaction,' 'nice person to do business with,' 'would highly recommend'). Participants have running totals of feedback points attached (visibly) to their screen names, which might be pseudonyms. Yahoo! Auction, Amazon, and other auction sites feature reputation systems like eBay's, with variations, including a rating scale of 1-5, several measures (such as friendliness, prompt response, quality product), and averaging instead of total feedback score" [18].

Second are **mediated systems**; in these, reputation information from other users is aggregated, transformed, and published, after the model of bond ratings or credit scores. This is the model that is adapted by Slashdot.org (users rate comments, highly rated users gain "karma" for metamoderating rating votes).

We would suggest that there can be a third type. In an **authoritarian reputation system**, publishers or managers of the web site could either publicly post their assessment of a user, a user's comments, a product, etc., or save this information "behind the scenes" for future actions against the user, such as banning or inviting him or her to participate in the site's management or creation.

Though some ratings and reputation systems are trivial (for example: where users can rate a comment thumbs-up or thumbs-down, and all users can see who voted which way), but some rely on complex algebra and/or data not accessible to users, such as how one user has been rated over the course of years. Opacity may be intentional, in an attempt to prevent users from gaming the system (as Google has strongly guarded its PageRank algorithm), or simply "security through obscurity." We will discuss in below the relevant issues with and research on this.

More and more online newspapers seem to be adopting, or considering adopting, some sort of reputation system. The washingtonpost.com site, for example, which sees over 320,000 comments per month, will go to a "tiered" system in mid-2010; the site's ombudsman described it as "commenters being assigned to different 'tiers' based on their past behavior and other factors. Those with a track record of staying within the guidelines, and those providing their real names, will likely be considered 'trusted commenters.' Repeat violators or discourteous agitators will be grouped elsewhere or blocked outright."[29]

It's critical to remember that a necessary component of these systems is persistent identity. Whether this comes from a system-assigned ID or a user-chosen username, without having some way of identifying and naming the subject of a rating, and having that subject remain the same person/company, no recommender or reputation system can provide useful information to other users. And so reputation systems inherit the same issues that apply to identity.

3.1.4Registration (is not Identity)

Many sites require some sort of **registration** for users to post a comment, or access other site features. For example:

> "NYTimes.com requires that you supply certain personally identifiable information, including a unique e-mail address and demographic information (zip code, age, sex, household income (optional), job industry and job title) to register. By using NYTimes.com, you are agreeing to our Terms of Service."30

However, this "basic registration" (which we define as requiring a valid email address but no other confirmable data – the other demographic information cannot be confirmed or disproven in the registration system) can easily be defeated. And in fact, usually is, even by well-intentioned users. It requires no confirmable identification information to sign up for a "valid email address" and sites like bugmenot.com do brisk business with not only the malicious but also the privacy-minded.

And this results in a mild barrier to entry to some, as it adds a step and asks for what some might consider personal data while asking for "trust" (which, as mentioned above, can be seen as requiring human-to-human interaction, as is not the case in a registration system) that no ill will come from this. Yet it is no barrier to the willful comment abusers, or a very low one.

The online site Huffington Post, which is not a newspaper site but contains many articles and commentaries about political issues, requires commenters to register, and like *The New York Times*, encourages users to use their real names. "That makes things more civil. It also creates a data trail of individual passions and preferences. Huffington says that data might be useful down the road."[31]

And some sites "simply do not allow their users to be anonymous. Social networking sites, for example, like Facebook, often require their users to act under their real names. Accordingly, you should consider a site's terms of service on this subject if anonymity is

---

[28] E. Goldman, "Regulating Reputation Systems" talk at UC Berkeley iSchool, April 14, 2010

29http://www.washingtonpost.com/wp-dyn/content/article/2010/04/02/AR2010040202324_2.html

30 http://www.nytimes.com/ref/membercenter/help/privacy.html

31 http://marketplace.publicradio.org/display/web/2010/08/02/pm-comment-cops-help-manage-websites/

important to you."[32] However, these registrations are also easily defeated; we have friends who use alternate email addresses to register on Facebook under fake names (sites such as 10minutemail.com even offer disposable email addresses) in order to separate out professional and personal profiles.

Some newspaper sites have used third-party commenting systems, such as Disqus, Pluck, and Intense Debate (with varying degrees of success). Such systems can accept or even require commenters to log in using OpenID, Facebook, or Twitter logins and identities. These were designed to be tied to "real-world" personal identities and as such would tie online actions to offline life, which has shown to be an effective counter to the effect of "on the Internet, nobody knows you're a dog".

There have been no studies of how requiring a real identity would affect participation, but there have been real-world examples. In mid-June of 2010, *The Times* (UK) demanded visitors to their web site register accounts; this did not entail any monetary cost to the user. Metrics firm Hitwise stated that they saw traffic to the site drop to half of what it was. The company saw that users who encountered the registration page mostly (by a factor of two to one) went to another news site or to Google. Though this could be only a brief dip, it's a scary prospect for any site that relies on traffic.

In addition, it is almost as trivial for a determined troll to create multiple and untraceable OpenID, Facebook, and Twitter accounts. It adds another step, but ultimately would not be much of a deterrent.

And even innocent and innocuous parties can react strongly and negatively to the requirement of a real-life identity. The game company Blizzard announced mid-2010 that it would force users of their official online forums to use their real names. Almost immediately a widespread protest arose from users, many arguing that it would open users (who, remember, were facing off in virtual battle) to harassment and stalking in real life – an issue especially acute to female gamers.[33] Blizzard quickly withdrew the policy.

And though an opaque identity is a vector for trolling and abuse, it can also be a shield for political dissidents (as in the case of a protest online comic in Iran[34]), whistleblowers[35], and at-risk populations. So there is a balancing act, a tradeoff.

We would contest that while registration for online comment systems does serve various purposes, current forms are not enough to prevent the crisis in commenting, and do not begin to consider why. In addition, the mere act of requesting registration, strong of weak, may have a negative effect on participation. To what degree has yet to be studied.

3.2Towards Building a Better Trolltrap

Though we can offer no definitive technological solution – all attempts would require controlled testing in the wild – we can present general guidelines for various approaches and identification of their challenges.

There are a number of approaches. One is through identity and banning; another is through "burying" abusive or flaming comments; another is an extension of reputation ranking.

For any system of the first type to have a significant impact against trolling, it would require the ability to recognize some sort of permanent identity connected to a known troll while not placing a high bar for regular users.

However, there are technical and sociological issue inherent in this approach.

Existing systems are losing the arms race against trolls, who find new ways to mask or spoof identifiers such as IP addresses, and there is no indication this will change in the future. This is perhaps as it should be, as online privacy should not be an ultimately permeable feature.

The above issue could be tied to requiring strong registration linked to some form of verifiable real (and real-life) identity. But such an action has, in the instances where it has been applied, showed a dramatic drop in all users. Research is needed into whether this was an isolated incident, related to other factors, or a generalized effect.

At least initially, this type of system would act post-hoc – trolls would be able to post, these posts would be visible by the entire community and have their negative effect until reported (by humans) or analyzed (through automation, perhaps by a natural language process that recognizes common patterns or words in abusive comments).

Perhaps such a system would act in the long run as a deterrent, if reaction to abusive posts were rapid enough to mitigate their effects and the identification system were strong enough to make logging in under another identity arduous enough. But again, potential implementers would have to study whether the long-run gain would balance out the short-term burden and repellent to average users – and whether this approach would work at all.

This approach also could possibly penalize otherwise valid users who get engaged by a troll, or other users, in a heated exchange or flame war.

And still, this approach would have to be studied in the wild to see if it would be at all acceptable to a significant amount of users. Anecdotally, this seems not to be the case.

Systems of the second type share with the first type a post-hoc nature, and would not prevent abusive postings (perhaps giving trolls all the rewards for their motivations.) But they do not rely on the tricky problem of identification of users.

However, they do require some sort of moderation, whether in-house, user, or automated. The technical question is whether the "burying" would be rapid enough to minimize or eliminate the damage, though the stronger this process is the more likely it will produce "false positives". And all the problems outlined above in the "All in Moderation" section apply.

The reputational approach can suffer all the issues associated with recognizing or assigning a personal identity to users, as in the first

---

[32] http://www.citmedialaw.org/legal-guide/legal-issues-consider-when-getting-online

33 The web comic Penny Arcade had its own and NSFW take on the issue: http://www.penny-arcade.com/comic/2010/7/9/

[34] http://www.huffingtonpost.com/2010/02/19/online-comic-from-iranian_n_469001.html

[35] http://www.journalismethics.ca/citizen_journalism/anonymity_of_reporters.htm

approach.

Some sites that use a reputation system, such as Slashdot.org, allow users to set a filter on how they view comment threads. For instance, a regular visitor to the site could choose only to see comments that have been rated 5 (on a scale of 0 to 5, 5 being best by a number of criteria). This does not prevent trolling, or flame wars, but in a way it requires users to "opt-in" to see them, as it is unlikely abuse would be voted up.

This is promising, but still faces some hurdles that could negatively users.

First, it does require some sort of registering, which seems to discourage some users, especially ones who are not regular visitors to that site (and as news sites link to each other more, and aggregators such as Google News become the first step for many readers, it is more likely news sites see a higher proportion of irregular visitors).

Second, it presents a higher bar for each comment to clear before it can engage in the "public" discourse. This could also make most discussions asynchronous and not meeting Rafaeli's definition of interactivity.

4.DISCUSSION

4.1What Drives Trolls?

Mattathias Schwartz's 2008 *New York Times Magazine* article "The Trolls Among Us" was a rare example of first-person voices of active, "professional" trolls. This is anecdotal and far from wide-reaching, but as Herring, Job-Sluder, Scheckler, and Barab pointed out, no substantial (if any) study has been done of that population, so we'll take unverified insights where we can. The article profiled a few notorious and enthusiastic trolls, giving them voice to explain their own (public, at least) motivations.

Jason Fortuny (of "The Craigslist Experiment"[36] and other pranks), "Weez", and a few even more anonymous sources went on the record as to the whys of their activities. (There of course is a caveat: Schwartz admitted that "I did everything I could to verify the trolls' stories and identities, but I could never be certain"; the same high levels of deception and secrecy that prevented Herring et al. from getting useful data (more on this below) were in play.)

"[Y]ou exploit their insecurities to get an insane amount of drama, laughs, and lulz," said one, and another said that he "keeps score" by lulz. This is consistent with "The object of recreational trolling is to sit back and laught at all those gullible idiots that will believe anything", as stated by "Andrew" on *The troller's FAQ*[37], cited in Herring et al. It should be noted that the trolls were aware that the laughs are all on their side, intentionally, even though their targets perceive it as sexual or violent harassment. Schwartz wrote Fortuny's attitude was that "no one is entitled to our sympathy or empathy", though Fortuny later expanded that actions were to "find people who do stupid things and turn them around," usually through an auto da fe of flaming.

Indeed, Orchard and Fullwood stated that research suggested that "the CMC environment can reduce personal and social restrictions that may occur in face-to-face communication," [19] and noted Suler's 2004 theory of the "online disinhibition effect". The six factors key to this effect are dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, and reduced authority. We've discussed some of these, such as anonymity and authority, but we would encourage further study of the other factors, if such research could be captured.

This dovetails with Donath's earlier study of trolling in Usenet groups. Donath labeled the tactics there "pesudo-naive": posting a "stupid" question, seeing who would respond, and overwhelming that responder. The idea was that only a newbie wouldn't recognize the trolling attempt, and this was their punishment until they learned or, in Fortuny's terms, "turned around", sussed the shibboleth, and joined the tribe of worthwhile citizens.

However, we think it would be unsupportable to extend this motivation to the world of flame wars on news sites. As described above, the contexts and actions of trolls on news sites do not lend to the goal of initiation but instead of intimidation; the end goal of Fortuny, if his statements are to be believed, are to expand his perceived community, which news trollers poison and close communities.

Herring et al. focused more on the disruptive nature of trolls, in the context of their disruptions of a "feminist web-based discussion forum". This forum shares, for our purposes, many features in common with newspaper online discussion spaces – a desire for signal easily confounded by trolling, a perception that one population is insiders and the other is outsiders (the overwhelming instances of news-centric trolling can be said to be a "rage against the machine", whoever the "machine" is seen as) – and so we can treat the findings as applicable to the newspaper site space. Perhaps this is a better match than with Donath's work on Usenet.

The context of Herring's study was an online feminist forum of over 4,000 members (about 200 actively participating). The study looked at a "single thread of 111 messages" among 41 individuals, 90 percent female, between March 13 and March 21 of 2000. The discussion was sparked by "a new male participant, Kent, [who] started posting messages that were intentionally antagonistic." He continued to disrupt the forum for almost eight weeks.

Herring et al. did not contact Kent; they attempted to analyze via his actions using grounded theory methods, assigning codes to his behaviors. They found his tactics included: outward manifestations of sincerity (presenting himself as a valid, rule-

36 http://www.wired.com/threatlevel/2006/09/craigslist/

37 http://www.altairiv.demon.co.uk/afaq/posts/trollfaq.html

following participant), flame bait (as per Donath's definition), and attempts to provoke futile argument (willfully misinterpreting, offering to change his actions if others would only "answer" his questions). In response, forum members: called for administrative banning, implored other members to ignore him (the "don't feed the trolls" tactic), took on his claims head-to-head, exchanging insults, and then trying to negotiate what is and isn't appropriate behavior in the forum. None of these responses were satisfactory.

This Kent was a classic troll, in both Donath's narrower and our broader sense of including flame war initiators. Though he was targeting an overtly feminist forum, all the observed behaviors have been seen in almost every newspaper site, and have been (without reference to or knowledge of this study) described in "we have a crisis" news stories.

Herring et al. concluded that with targeted groups (such as feminists), it could simply be a dislike of that group. But they noted that it's not just "at-risk populations" suffering from trolls – apparently even forums about car racing and Internet researchers (!) get hit. The "common denominator", they said, was that trolls savor the attention they receive. (It was a long way around to get back to "don't feed the trolls".)

Shachaf and Harada, in their study "Beyond vandalism: Wikipedia trolls" agreed that the motivations they could draw out included "boredom, attention seeking, and revenge." [20]

However, their research ran into the same problems as that of Herring et al.: lack of sample. The latter drew conclusions from one troll, and Shachaf and Harada were able to get only eight sysops and four sysop-identified trolls, and were aware of the shortcomings of this. They said, "Despite the fact that trolling is a common online phenomenon, it has rarely been the focus of previous research, with only one exception" – that exception being Herring et al!

4.2 Could Comment Systems Learn Countering Tactics from Troll Motivations?

Though evidence of what motivates trolls, and how to cool tempers that flare up into flame wars, is anecdotal, it is not inconceivable that learning more about both could result in design recommendations for more efficient commenting systems.

Just based on the few available interviews and results outlined above, we can see that anything that serves to reduce the level of "drama, laughs, and lulz" would reduce the rewards certain types of trolls seek. This could include a time-based solution that would get rid of the abusive comment as soon as possible (identity-based or content-based recognition and deletion/banning, in-house or effective user moderation), or a way to suppress the visibility of the abusive comment independently of the user (as in requiring a high reputation to make the comment visible to anyone, or allowing reputation-based browsing).

The findings by Schwartz, Herring et al., and Shachaf and Harada all suggest that the examples they saw were motivated by conflict. This would suggest not posing the site itself as in opposition to the troll, possibly redoubling his or her resolve and moving the target from users to the site's structure or administration. Perhaps making whatever tactic suppresses the troll more distributed, as in an all-user-based system, would diffuse the target and give the troll no feedback, or "drama".

Finally, Suler's and Orchard and Fullwood's work suggest that there are qualitative differences between face-to-face, identifiable interactions and computer-mediated ones – namely, that the dissociating features of online interactions make users feel freer to act along the lines of trolls or flamers. Though we cannot suggest specific technical solutions (video comments? personal photos alongside handles?), anything that more solidly relates an online handle to a "real person" could serve to reduce the motivation for, or install societal barriers, to abuse.

5. CONCLUSIONS AND IMPLICATIONS FOR FUTURE STUDIES

The interest in this topic arose from one author's past experience working as a journalist and editor, and his concern for the health of the newspaper industry as well as civic discourse. The latter seems, as outlined in the Introduction, closely tied to the latter.

We found that existing commenting systems do not, and perhaps cannot, offer ironclad protection against even casual trolls and tend not to mitigate flame wars. Modifying these systems would require a careful examination of the tradeoff between protection and driving away well-intentioned users, who may not stomach requests for registration, infringement (even voluntary) of their online privacy, or want to make the effort to contribute moderation votes.

Towards supporting or disproving our conclusions, we strongly recommend further quantitative and qualitative research into the motivations or trolls, to see what factors could reduce their urges, or mollify their needs. We also encourage user testing to determine how what levels of registration (from weak, email-account-only, to strong, tied-to-real) affect user participation.

6. REFERENCES

[1] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (Nov. 1993), 795-825. DOI= http://doi.acm.org/10.1145/161468.16147.

[2] Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.

[3] Fröhlich, B. and Plate, J. 2000. The cubic mouse: a new device for three-dimensional input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands, April 01 - 06, 2000). CHI '00. ACM, New York, NY, 526-531. DOI= http://doi.acm.org/10.1145/332040.332491.

[4] Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.

[5] Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[6] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.

[7] Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology* (Vancouver, Canada, November 02 - 05, 2003). UIST '03. ACM, New York, NY, 1-10. DOI= http://doi.acm.org/10.1145/964696.964697.

[8] Yu, Y. T. and Lau, M. F. 2006. A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions. *J. Syst. Softw.* 79, 5 (May. 2006), 577-590. DOI= http://dx.doi.org/10.1016/j.jss.2005.05.030.

[9] Spector, A. Z. 1989. Achieving application requirements. In *Distributed Systems*, S. Mullender, Ed. ACM Press Frontier Series. ACM, New York, NY, 19-33. DOI= http://doi.acm.org/10.1145/90417.90738.

[10] Ridings, C., Gefen, D., & Arinze, B. 2002. Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information Systems, 11*(3-4), 271-295.

[11] Ash, E. , Hettinga, K. and Halpern, D. 2009-08-05. Effects of a trend: The influence of user comments on readers' perceptions of online newspapers. *Paper presented at the annual meeting of the Association for Education in Journalism and Mass Communication, Sheraton Boston, Boston, MA*. Retrieved from http://www.allacademic.com/meta/p375864_index.html

[12] Herring, S. C., Job-Sluder, K., Scheckler, R., & Barab, S. 2002. Searching for Safety Online: managing 'trolling' in a feminist forum, The Information Society 18(5) (2002) 371-383.

[13] Winsvold, M. 2009. Arguing into the digital void?. *Javnost-The Public*, *16*(3), 39-54. Retrieved from Academic Search Complete database.

[14] Wilkinson, D., and Huberman, B. 2007. Assessing the value of cooperation in wikipedia. *First Monday 12*, 4.

[15] Chung, D. S. 2004. Into interactivity? How news websites use interactive features. Paper presented at the annual meeting of the International Communication Association, New Orleans Sheraton, New Orleans, LA Online. Retrieved from http://www.allacademic.com/meta/p113336_index.html

[16] Lampe , C. and Resnick, P. April 22-24, 2004. Slash(dot) and burn: distributed moderation in a large online conversation space. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 543-550.

[17] Byler, Dan. 2010. Recommenders and the high-choice environment. Paper for INFO 203, University of California, Berkeley.

[18] Resnick, P., Kuwabara, K., Zeckhauser, R., Friedmanm E. 2000. Reputation systems. *Communications of the ACM*, *43* (12), 45-48.

[19] Orchard, L., and Fullwood, C. 2009. Current perspectives on personality and internet use. *Social Science Computer Review*. 28: 155.

[20] Shachaf, P., Hara, N. 2010. Beyond vandalism: Wikipedia trolls. Journal of Information Science, 36(357).