Comment Space

by

Geoffrey D. B. Glass

M.A., Simon Fraser University, 2008

B.Sc., University of Toronto, 1995

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Communication
Faculty of Communication, Art and Technology

SIMON FRASER UNIVERSITY
Fall 2015

Approval

Name:	**Geoffrey Glass**
Degree:	**Doctor of Philosophy (Communication)**
Title:	***Comment Space***
Examining Committee:	**Chair:** Frederik Lesage Assistant Professor
Andrew Feenberg Senior Supervisor Professor	
Stuart Poyntz Supervisor AssociateProfessor	
Maria Bakardjieva Supervisor Professor Faculty of Communication and Culture University of Calgary	
Richard Smith Internal Examiner Professor School of Communication	
Greg Elmer External Examiner Professor Radio TV Arts Ryerson University	*by video conference*
Date Defended/Approved:	November 18, 2015

Abstract

Reader comments are an online communication format defined by their marginality relative to a primary news or blog article. To investigate their distinctive technical features and social dynamics, I studied large-scale discussions in response to articles about two stories: the death of Aaron Swartz, and the outing of Edward Snowden as the NSA leaker. Using frame analysis and drawing on Hannah Arendt's theories of judgment and public action, I describe how the these comments by ordinary people give meaning to political action and define a space of political legitimacy.

Acknowledgements

I wish to thank Drs. Andrew Feenberg, Maria Bakardjieva and Stuart Poyntz for their patience and insight. This research was assisted by a grant from the Social Sciences Research Council of Canada. Above all, I think my wife, Cindy Xin, for her brilliance and support, and my son Julian for many good conversations. Thank you.

Table of Contents

List of Tables

List of Figures

Glossary

Article	A post on a blog or news site about a story.
Comment	A post, typically short, written in response to something else.
Discussion (of reader comments)	All the reader comments attached to a given article.
Hierarchical threading	A feature of some comment systems that represents threads as a visual hierarchy, with each reply indented below its parent. Some systems permit a multi-level hierarchy, others limit hierarchy to just two levels (top-level comments and their replies).
Overall rating	A derived rating calculated based on votes, usually by subtracting down-votes from up. Sometimes I refer to this as a comment's "score."
Post	Generic term referring to an article, comment or other piece of written content published on the Web.
Rating	The aggregate number of votes for a comment. Some systems may include both up and down ratings. Sometimes referred to as votes (plural).
Reader comment	A comment posted in response to an article, enabled by the article's host site. Article authors and reader comment writes belong to different social groups.
Score	See overall rating.
Story	An event or topic of discourse, usually written about in multiple articles.
Thread	A comment and its replies, their replies, and so forth. A thread can include other threads.
Top comments	Comments with the most number of up-votes. Sometimes I use "high-rated" comments instead. For the purposes of analysis, I usually refer to the top 50 comments by up-votes as the top comments in a discussion.
Vote	A point, thumbs up or down, or "like" given to a comment by a reader. The practice of granting such a point. Some systems support both up- and down-votes.

Chapter 1: Introduction

We're all insignificant. That's our importance.

— reader comment in *Guardian* discussion about Edward Snowden's NSA leaks

Reader comments have become a common fixture on blogs and mainstream news sites, including prominent publications like the *New York Times*, the *Guardian*, *CNN*, and *CBC*. A 2010 Pew Internet study (Purcell, Rainie, Mitchell, Rosenstiel, & Olmstead, 2010) found that more people had "commented on a news story or blog item about news that they read" than had linked to a story through a social networking site or Twitter (25% versus 17% and 3% respectively). Marissa Nelson at CBC reports that "about a third of our audience have commented at some point, but . . . three-quarters read the comments" (Tremonti, 2014). As of February 2013 the CBC site was receiving up to 20,000 comments per day—double the volume six months earlier. Other sites handle even more comments: some *CNN* articles receive over 10,000 comments in just a few hours. Reader comments have rapidly become a recognized and prominent part of journalistic practice. Sites like *CBC, CNN* and *Ars Technica* (a technology new site) publish articles about reader comments in response to previous stories, while the *New York Times*, the *Washington Post* and Mozilla are collaborating to develop a new open source reader engagement and comment system (Farhi, 2014).

Although reader comments are a part of everyday life online, little existing research accounts for their unique characteristics. While some reader comment discussions resemble other online discussion formats (e.g. forums), some do not, taking place on a very large scale, with hundreds or thousands of comments over a short period of time. This is particularly true in discussions on high profile news sites and blogs where the significance of reader comments is perhaps greatest. Studies of political discourse in discussions like these are dominated by deliberative theories for which many such discussions are a poor fit. There is a lack of research that describes what participants in reader comments do and say on their own terms, while considering the particular technical and social features of the format.

I argue that marginality is the foundational defining characteristic common to all reader comment discussions: each is subsidiary to some primary article or document.

Each comment exists in the space of another. Not only have comment discussions emerged as secondary to the content of the primary article, their marginality is also a material fact: each discussion is one click or link removed or quite literally in the bottom or side margin of the page.

I address a number of questions about the large-scale reader comment discussions in my study:

- What are the consequences of this marginal condition for the content and dynamics of large-scale reader comment discussions?
- What else, beyond responding to a primary article, characterizes and unites them?
- How do participants perform and act? What do they say?
- What methods can be developed to get a grip on the tremendous volume and diversity of reader comment technical and social practices and discussion?

I studied cases of a number of large-scale reader comment discussions about two stories: the death of Aaron Swartz, and the revelation that Edward Snowden was responsible for National Security Agency (NSA) leaks. While not typical of reader comment discussions, these are exemplars that illustrate two critical discourse moments (Chilton, 1987) that make visible political exclusions and challenge the assumptions of hegemonic discourse. In these instances, reader comments challenged the norms of dominant discourse on two levels: first in explicit debate about the legitimacy of the actions of these two men; second in the clash of how these events were understood and framed, among comments and in contrast to statements by privileged journalists and officials to which they responded.

In Chapter 2, I survey existing studies of reader comments, define them and explore some logical consequences of their marginality. I characterize their distinctiveness as a format compared to other online media, explain their genealogy and technical features, and consider the dynamics of the groups who use them and how they are perceived.

Chapter 3 develops a theoretical framework for understanding reader comments as a space for public political discussion. Participants act contingently in a space they do not control, in view of a public they do not know. Their comments are rarely strategic

interventions in a political debate: rather they are opportunistic responses to the actions of others. I draw on Warner's (2002) conception of publics, Arendt's (1959) theory of action, and de Certeau's (1984) distinction between strategy and tactics. Following Warner and Arendt, I see reader comments as exercises in world-building. Commenters judge what they comment on, they imagine a public they address and the shared world in which they live; in doing so they help to bring it into existence. The performance of commenters is not simply the rational debate of democratic deliberation. It is also an exercise in shared imagination: one that often explicitly or implicitly rejects assumptions and expectations of the primary article.

Through my early attempts to analyze reader comments I developed approaches for analyzing them effectively. These efforts are the topic of Chapter 4. I present not only the techniques I arrived at, but a narrative of how I got there, of what worked and what didn't. I include some worthwhile results about copyright discussion online, and additional analysis of collected eulogies of Aaron Swartz excluded from the main body of the study because they are not reader comments as I use the term; their differences nonetheless provide useful points of comparison.

Chapter 5, Method, provides an overview of existing approaches to analyzing reader comments. The discussions I examine collectively encompass tens of thousands of comments. The challenge for my method is how to analyze large-scale discourse lacking in authorial or editorial unity, characterized by widely varying content, popularity and representativeness (reflected in reader ratings) across sites with different audiences, technical features, and social norms. I explain how and why I pick cases, and describe how I use frame analysis (Gamson, 1992; Goffman, 1974) to study what participants actually say. Frames capture assumptions and expectations of participants in discourse: about political matters, about the relationships among participants, about the expectations of discourse. I also describe techniques I use for taking advantage of the particular technical and social features of the discussions in my study.

Chapter 6 gives a detailed description of the technical design of each of the comment systems in my study, and describes some of the implications of their design features.

Chapters 7 and 8 zoom in to apply my method to the specific discussions in my

study. These are attached to articles about two news stories at particular moments in time: the death of Aaron Swartz (Chapter 7), and the revelation that Edward Snowden was the NSA leaker (Chapter 8). I first explain what each story was about and why I chose it. Then, on a discussion-by-discussion basis, I give a description of the discussion and the associated article. I use frame analysis to structure my effort to attend closely to what participants have to say in each case (discussion). My goal at this point is to provide a rich and textured exploration of my cases.

In Chapter 9 I synthesize data from all discussions in order to develop more general understandings of reader comment discussion. I present the legitimacy of political action as a major theme of discussion and propose judgment by readers as a unifying feature of comments in my cases. I argue that reader comments occupy and shape physical and conceptual spaces, describe the relationship between the two and explain the implications for public discourse.

I conclude with Chapter 10, followed by a discussion of the limitations of my research and opportunities for future work in Chapter 11.

Through this study, I hope to develop an understanding of reader comments that can serve as a basis for further research. Reader comment discussions are largely the province of ordinary people who are neither politically disengaged nor mobilized, few of whom are elites, recognized experts or members of the media. They sometimes appear to be radical, ignorant, hateful, empathetic, insignificant. Yet contrary to the scorn of many journalists, participants and observers, I find they often live up to Gamson's (1992, p. 4) optimistic claim that "people are not so dumb." They are an important avenue of public participation and expression, and a resource for studying public discourse on many topics. Their marginality does not render them insignificant: on the contrary, it enables them to judge, providing meaning and context where journalists do not, and helping to define a space for politics.

Terminology

Several of the terms I use are potentially ambiguous. For example, *post* might refer to a reader comment, or to the blog article to which the comment responds. Unless the meaning is clear from context, I use the term *comment* for the former, *article* for the

latter. *Story* is not a synonym for article, nor is it a text: it is an event or other topic of discourse reported on in one or more particular articles.

The collection of reader comments attached to a single article is a *discussion*, whereas a *thread* is a comment and subsequent replies within a discussion. I never refer to a discussion as a whole as a thread. Some sites support hierarchical threads, in which threads are represented as a visual hierarchy. Others are flat; participants in these can nonetheless construct threads by replying to and possibly quoting one another.

The reader comment systems I examine allow readers to *vote* comments up and down. In general terms, the sum of votes a comment has received is its *rating*. Some sites display an *overall rating* calculated by subtracting the total number of down-votes from the total number of up-votes. If a site does not permit down-votes, the overall rating is the number of up-votes. I often cite a comment's rating as a pair of numbers (e.g. +12/-0) or as a single number (e.g. +9) if only up-votes are permitted. To help ratings stand out from the text, and to distinguish up-votes from down, they always include the sign, plus or minus, even when a total is zero. When I refer a *top* or *high-rated* comment I am concerned only with up-votes. A comment with a rating of +200/-800 and an overall rating of -600 might nevertheless be a top comment. I justify this choice later.

Chapter 2: Reader Comments

> Please use the comments to demonstrate your own ignorance, unfamiliarity with empirical data and lack of respect for scientific knowledge. Be sure to create straw men and argue against things I have neither said nor implied. If you could repeat previously discredited memes or steer the conversation into irrelevant, off topic discussions, it would be appreciated. Lastly, kindly forgo all civility in your discourse . . . you are, after all, anonymous.
>
> — Barry Ritholtz's (n.d.) advice to prospective commenters on his blog

To begin, I outline existing studies that address reader comments, then explain how reader comments differ from other communication formats like forums and networked social media. I approach this from multiple perspectives, including the genealogy of the form, the relationship of reader comments to a primary article, how that relationship shapes participation, technical design, the character of reader comment discussion, and perceptions of reader comments as a distinct format.

Here I establish the foundation for the rest of my study. Some of what I say resembles conclusions rather than groundwork, begging questions that I ask later. In fact, the understanding I describe here is rooted in over fifteen years of personal experience reading and participating in reader comments. Long before I conceived of studying reader comments, I read them as a matter of course, often skipping past the articles to which they were attached just to read the comments. My claims here are also strongly influenced by results of my preliminary study, described in Chapter 4, in which I experimented with different ways to approach and understand reader comments. Later, I will return to some of these claims to test them against my data.

Defining Reader Comments and Existing Research

Compared to discussion forums and social media like Facebook and Twitter, reader comments seem to have attracted less scholarly attention. In particular, there is relatively little research that examines the meaning and content of reader comments *per se*. This is likely due in part to the status of comments. Reader comment discussions are attached to mainstream news articles, blog posts, and just about every kind of user-produced media. Their defining feature is their marginality. Every reader comment

discussion is a response to some primary article or document. For example, Figure 1 shows some comments responding to an article in my study.

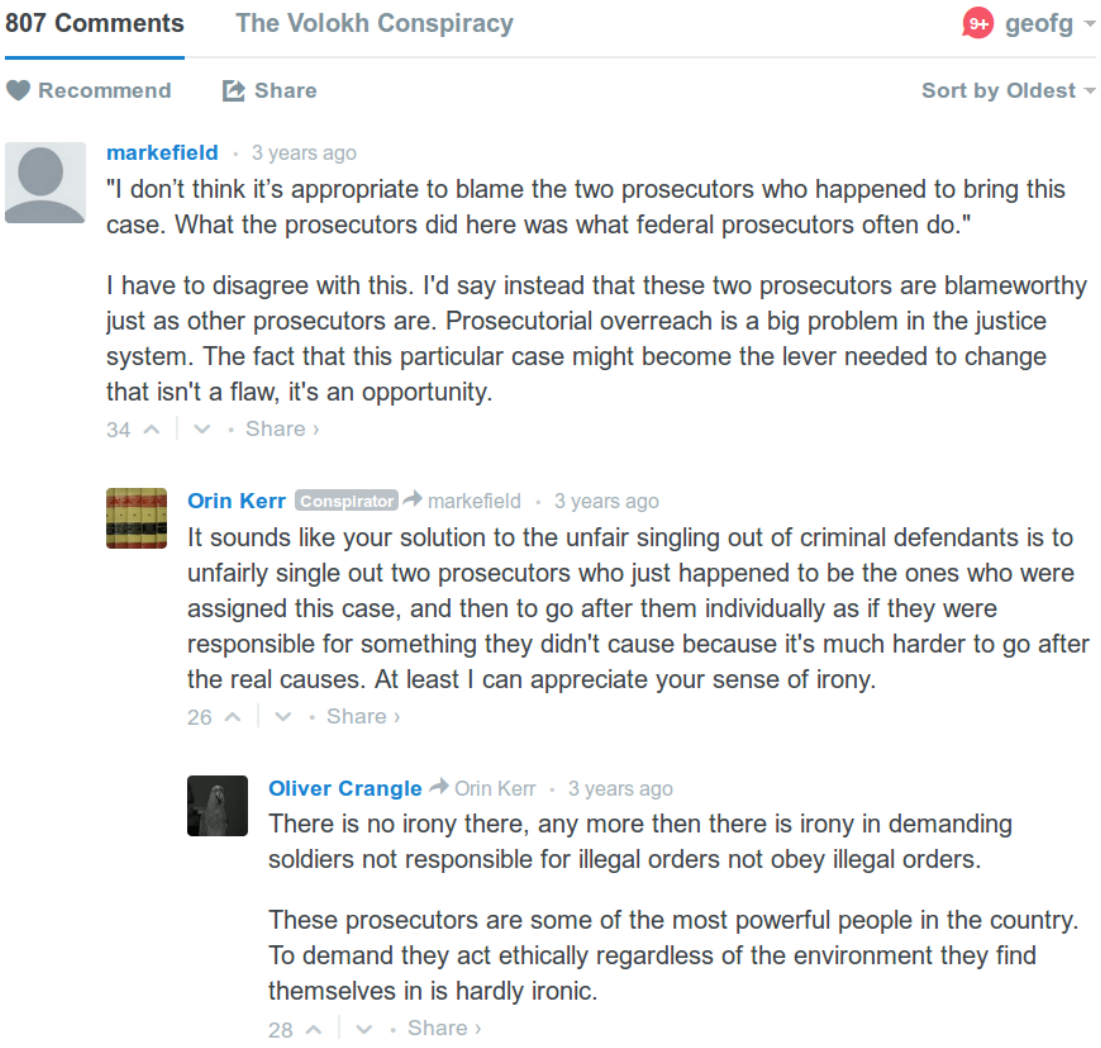

Figure 1: Reader comments about the death of Aaron Swartz (Kerr, 2013b)

Reader comments have a low profile. As a secondary feature for organizations whose sites host them, they lack the public (or financial) profile of a Facebook or a Twitter. The business of the *New York Times* is news and advertising, not comments. Comment systems are diverse, with no dominant implementation or representative, and no clear definition distinguishing them from other media such as discussion forums or social media. Not only are comments themselves marginal: they have often been

addressed indirectly by researchers examining some other phenomenon, or treated simply as instances of online forums.

Most of the early scholarship dealing with reader comments is concerned with them as a secondary feature of a page or site. They are closely associated with blogs, where they are widely used and whence they appear to have originated. Many studies of blogs have accordingly given them some consideration. For example, Herring, Scheidt, Bonus and Wright (2004), proposing that blogs constitute a distinct genre of web communication, describe how some enable discussion through comments on individual entries. They focus on posts on blogs that are "typically updated several times a week." Comments in active discussions are posted much more frequently, but they are subsidiary: "Visually and rhetorically, comments are behind the scenes," write Nardi, Schiano and Gumbrecht (2004, p. 228). Gumbrecht (2004) quotes an educator about the use of comments:

> comments are . . . very clearly rhetorically subservient . . . it starts a particular relationship to what you have to say, to what your comments have to say, which is very different than . . . forms like chat rooms or threaded discussions or Usenet-type discussions. (2004, p. 4)

Another blogger described comments as "the *heart* of the blog medium . . . a big part of making it publicly available is to have responses" (Gumbrecht, 2004, p. 4). Yet the blog is still primary. Nardi et al. (2004) and Gumbrecht (2004) argue that it is important for bloggers that they retain control over what happens on their blogs, including the comments. Blog authors and readers have "asymmetrical communication rights—the author retains ultimate control over the blog's content" (Herring et al., 2004, p. 10).

In the substantial body of research applying computational and artificial intelligence techniques to the analysis of large volumes of online communications, several studies have analyzed comments computationally as indirect indicators of the content of blog posts (Balasubramanyan, Cohen, Pierce, & Redlawsk, 2011; Kehoe & Gee, 2012; Mishne & Glance, 2006; Park, Ko, Kim, Liu, & Song, 2011). Balasubramanyan et al. (2011) use machine learning techniques to predict positive and negative and positive sentiment in a blog post based on the content of associated comments; they find differing sentiment in different blog communities. Kehoe and Gee

(2012) developed a corpus of 222,000 blog posts and 2.2 million associated comments for diachronic linguistic analysis. Their analysis finds that the occurrence of words in comments was a strong indicator of the content (the aboutness) of the associated post.

For scholars studying online journalism, comments are typically a secondary phenomenon, one that adds interactivity to news sites (Chen and Berger, 2013; Domingo, 2008; Kenney, Gorelik and Mwangi, 2000; Schultz, 1999). Here, the content of comments is of less concern than the relationship to news articles and the perceptions of journalists and news editors. Diakopoulos and Naaman (2011) examine factors contributing to comment quality using survey and interview data from journalists and comment readers and writers. Hermida and Thurman (2007) and Nielsen (2012; 2014) survey how journalists perceive and comments. Weber (2014) looks at factors influencing comment volume.

For computational research focusing on technical or structural features of reader comment discussion, the comments are central: yet the content of those comments is assessed quantitatively, not qualitatively (Gómez, Kaltenbrunner and López, 2008; Lampe and Resnick, 2007). Other studies examine the dynamics of reader voting systems used in reader comments and other online media (Cheng, Danescu-Niculescu-Mizil, & Leskovec, 2014; Danescu-Niculescu-Mizil, Kossinets, Kleinberg, & Lee, 2009; Sipos, Ghosh, & Joachims, 2014).

In addition to such quantitative studies there is a growing body of research that focuses on the political content of reader comments (Bakker, 2013; Gonzalez-Bailon, Kaltenbrunner and Banchs, 2010; Manosevich and Walker, 2009; Ruiz, Domingo, Micó, Díaz-Noci, Meso and Masip, 2011; Trice, 2011; Weber, 2014). Here, as with studies of online discussions in general (Bakardjieva, 2008), theories of political deliberation dominate (Toepfl and Piwoni, 2015). This is the approach taken by several studies of *Slashdot*, one of the first sites to feature reader comments (Gonzalez-Bailon, Kaltenbrunner and Banchs, 2010; Halavais, 2001; Ó Baoill 2000; Poor, 2005).

Another line of research examines problems with comments, such as trolling and spam, and their impact on reader perceptions of articles. Some address the failure of reader comments to live up to deliberative standards of civility or rationality. Anderson, Brossard and Scheufele (2012), for example, examine how readers of media stories and

blog articles about controversial science are influenced by reader comments; they conclude that negative comments influence the perceptions of infrequent readers of reader comments. Scheufele, Anderson, Brossard and Xenos (2013) find that a lack of civility polarizes readers' perception of the risks of controversial technologies. Lee (2011; 2012) also finds that reader comments on news stories influence readers' perception of wider public opinion about those stories. Buckels (2014) links trolling to sadism, psychopathy, and Machiavellianism.

Much of this analysis is rooted in deliberative theories of online discussion. On these terms, actual reader comment discussions often come up short. Bakardjieva (2008, p. 292) takes aim at this emphasis on deliberation in online discussion: emphasis on "a predetermined set of normative criteria focused on rational deliberation misses the importance of other communicative forms, and . . . may prevent analysis from assessing online political forums on their own merits." Trice (2011) argues that important dialog may take place in reader comments even though it may not reach the level of deliberation. Freelon (2013) notes the overemphasis on deliberation, applying communitarian, liberal individualist and deliberative norms. Toepfl and Piwoni (2015) propose that reader comment discussions can be seen as counterpublic rather than deliberative spaces.

In sum, there is a good deal of scholarship that considers comments and that confirms the volume and significance of the medium. There is very little, however, that attempts to understand the political significance of reader comment discussion on its own terms, or to integrate that with an understanding of the distinctive characteristics of comment systems and technologies. Such analysis, I believe, would help scholars use reader comments to understand what participants have to say about other issues.

One study that does use reader comments as a window into another matter is by Lovell and Lee (2011). Their interest is not in comments at all, but in patients' perceptions and experiences of health care. To this end, they examine fifty-four comments on a *Globe and Mail* news story. The journal editor prefaces the article with a remark justifying publication:

> Although one could argue that this small sample of patient comments from an online newspaper doesn't qualify as "research," it represents an

> interesting methodology . . . Although care must be taken not to draw firm conclusions due to this study's limitations, this is just the type of information that would likely never see the light of day in traditional medical journals . . . (Lovell & Lee, 2011)

This points to both the significance of what people say in reader comments, and the lack of a body of work developing convincing methods for doing so (the study takes "a general inductive approach" to analyze the comments which does not consider the specifics of the medium).[1] Appropriately, readers of the web version of the journal article have responded with comments.

Reader Comments as Distinct Format

Are not reader comments just one among a diversity of essentially similar online discussion formats? There have been extensive studies of web forums, email, Usenet, and social media platforms. What is to be gained by quibbling over technical and social details that may not even be consistent differentiators between these formats?

The answer, I believe, is quite a lot. Although they may not immediately be obvious, the distinguishing characteristics of reader comments really are significant. Reader comments are one of the most accessible, frequently encountered spaces in which serious political discussions take place among individuals who may not be particularly committed to formal political engagement through parties, activist groups or the like. They are an important space of ordinary political discourse. On prominent news sites they offer a space for political discussion before a large and unknown general audience that is scarce among more focused communities of interest online.

Studies of Internet communication are often about people and society. How are people using the Internet in everyday life? How does it transform our relationships with one another? Who is determining the shape technology takes and how it affects us? These are important questions. My analysis, however, is not about the experience of people. I am far more interested in discourse, publicness, politics and the technology itself. I believe none of these can be understood independently of the others.

---

1 Although I argue for the need to develop methods that take into account the particularities of reader comments as a medium, I should say that the focus Lovell and Lee's study—individual experience—is little affected by the issues I deal with and seems to be well served by the approach taken by its authors.

Many of the most distinctive features of reader comments are technical, from thumbs-up voting buttons to hierarchical threading and filtering mechanisms. Understanding the kind of publicness that reader comments afford entails examining technical details like these. Analysts of the Internet sometimes treat technology as a simple product of social or economic factors that determine it. This is a mistake. It is all the more important to get to grips with technology as technology *and* as social medium when it is in flux, when it has not yet settled into stable forms, for this is when critique offers the most potential for influence. If we wait until the technology stabilizes, we will only be able to write history.

Sometimes legitimate concerns about the technocratic domination of human beings leads to attempts to play down technology and put it in its place. That impulse may be tempting, but it is dangerously mistaken. Latour (2005) treats technological things as actors on par with humans: technology is not simply a consequence of human choices, but must be understood in terms of how it mediates and transforms human efforts. The power of that mediation is the danger and the promise of technology. If we are to contest the place of technology in human life—if we are to make technology human, or at least if we are to make the Internet a thing that improves the world rather than degrading it—we need to take technology seriously. I do not wish to claim too much for reader comments, which hardly represent a panacea for or the end of democratic discourse. But I do believe that my analysis can at least illustrate the importance of technology as technology. My chief concern remains with public political speech and action by ordinary people. To get there, however, I must go through technology, not around it.[2]

I think it is important, therefore, to draw distinctions between online media. There is a critical tradition of analyzing the Internet as a single medium whose social significance can be assessed by considering the system as a whole. I myself would

2 I will not make the argument here in favor of technology for technology's sake. Were I to do so, I would begin by combining Dewey's instrumentalism with Arendt's (1959) theory of action to code understood as speech, drawing for illustration on Johan Söderberg's (2008) argument that the threat of free software to capitalism lies in its playfulness. It is my feeling that a model of technology in its industrial role has infected many critiques of technology of everyday life—which is not to say "consumer" technology, for despite often being described thus, this technology is perhaps most notable for its potential to engage people as producers rather than consumers. Applying an industrial understanding to it can play into the hands of those who would limit its democratic potential.

argue for the importance of key features like hierarchy, singularity and heterogeneity. Jodi Dean's (Dean, 2005) critique of communicative capitalism takes such an approach. She argues that the capitalist political economy of the Internet has reduced online communication to commodified contributions. Hers is an important argument that captures some significant dynamics of online communication. Yet such abstracted analysis of the Internet—of indeed any technology—risks essentializing it, leaving little room for other possibilities (good and bad), for change, or for the tactical activities of users. Close attention to technical detail and social practices helps to account for agency and to contextualize and assess the significance of such broad critiques.

The Internet is particularly ill-served by generalization. In many respects it is more plastic and more heterogeneous than other media; moreover, it is relatively recent and still subject to rapid change. It is difficult to gain a perspective somewhere between essential claims about the Internet as a whole and particular analyses of or reactions to the latest fads, innovations or businesses. To stereotype eras, forty years ago the Internet was mainframe resource sharing; thirty years ago it was FTP, email, and Usenet; twenty years ago it was home pages and search engines; ten years ago it was blogs and Google; today it is Twitter, YouTube, and mobile apps. In principle, the programmability of the Internet allows it to be the basis for an unlimited diversity of communication formats and media. In social terms it is subject to what Pinch and Bijker (Pinch & Bijker, 2001) call interpretative flexibility. Nevertheless, over time a relatively small number formats have stabilized, become widely used and understood. Email was the first of these; others include blogs, wikis, and discussion forums. I suggest that reader comments are becoming another.

To define them briefly, reader comment discussions are secondary discussions associated with an article or other primary document, usually a news story or blog post. They are enabled and subject to moderation by the owner of the primary article's host site. The reader comment discussions I will focus on are also, for reasons I will explain, characterized by high volumes (hundreds or thousands of comments) over a short span of time (on the order of a day) by participants who need not be members of a community or organization and who are able to vote for individual comments, contributing to aggregate ratings. This combination of features is distinctive to reader comment discussions. It also represents the majority of systems used by mainstream news sites.

Explaining why these characteristics cluster together requires that I develop a more complete understanding of reader comments, paying attention to their genealogy, their technical characteristics, the social groups involved, and the interpretations of participants.

Genealogy

Commenting is an older and more diverse practice than reader comments. It can be carried out through a variety of media, from discussion forums to email. As a medium, the category of reader comment discussion refers more narrowly to systems that are technically and socially designed specifically to link discussion with a primary document that discussion participants did not write. Commenting has much in common with annotation. As a distinctive Internet medium, comments emerged on the Web in the late 1990s in parallel with and in response to the popularization of the Internet.

Online discussion long predates the Web: mailing lists, dial-up bulletin board systems, and systems like Usenet and FidoNet were prevalent in the 1980s, often on non-Internet networks. At a minimum, participants were among the few private individuals who possessed a computer and modem, or they had access through affiliation with an institution. Restricted to particular places or topics, these environments tended to feature relatively small, homogeneous populations of users. Studies of online discussion from this period before the commercial Internet often remarked on a sense of community closeness.

In the 1990s this picture was turned upside down with the advent of a commercial, popular and largely Web-oriented Internet. What had in many cases previously been cozy communities of like-minded people were opened up to large, diverse user bases and to a flood of money-making schemes. These were given a boost by the appearance of Google in 1998, whose link-counting PageRank algorithm transformed hyperlinks into commodities. Attempts to gain clicks by link spamming was already becoming widespread; it now no longer mattered whether a link was ever clicked on by a human being—simply being indexed by Google made it valuable. This created a tremendous incentive to spam sites in order to boost profiles, page views, and profits.

Usenet, which had been a mainstay of vibrant online discussion, did not adapt

well to the influx of users and commerce. Many newsgroups became unmanageable; by the early 2000s service providers were starting to drop Usenet support. For most people, Usenet as a separate application has faded into Internet history, a forum of which few users are aware; even if they are, they probably access it over the Web via its integration into Google Groups. A guide for Usenet moderators from 1997 hints at what was to come:

> Since 1995 many proposals have been made to moderate previously unmoderated groups in the wake of increased volume of both on-topic and off-topic posts, widely cross-posted trolls, and a general perception that the signal-to-noise ratio of Usenet has been dropping. (McKeon, 1997)

While various discussion formats predate the Web, there is no clear antecedent that linked a digital document to comments by readers, as previously there was no widespread adoption of a networked software application with that capability. Before the Web there was a limited set of standard applications for discussion. Usenet was not simply a discussion format: it had to be realized as news reader software. Each form of discussion corresponded to a technical standard and application software that participants would need to install: email, news, IRC (Internet Relay Chat), FTP (File Transfer Protocol), Gopher. A text-oriented command line tool like Telnet could allow for a unique service with its own features, but except for bulletin board systems and perhaps MUDs (Multi-User Dungeons) these did not achieve any kind of widespread adoption. If someone had developed a reader comment system, they would need to deploy software supporting it to sufficient users to make it worthwhile; lacking critical mass, a new application and the discussion it enabled was unlikely to ever take off. In most cases it was more practical to adapt existing communications media, like email.

An important example of an approach to commenting is the RFC (Request for Comments) process undertaken by the Internet Engineering Task Force (IETF) in order to develop technical standards. The term RFC refers to actual documents, the first of which was essentially meeting minutes; Steve Crocker called it a Request for Comments “to avoid sounding too declarative” (Bardini, 2000, p. 185). The acronym RFC has come to indicate the standards themselves (termed “recommendations”), but RFCs begin life as proposals to which interested parties could respond with comments. Discussion itself

was (and is) carried out using email mailing lists. The following advice, for prospective participants in IETF deliberations, is from their current website:

> Beware: if the WG [Working Group] is in an active phase of discussion, you may well receive tens of message a day from each list. You need to be using a mail program with a good method of automatically sorting incoming mail into multiple inboxes. Once you have that working, simply read the mail threads daily, and read any draft documents that they refer to. (Force, n.d.)

This advice seems anachronistic. “Tens of messages a day” pale beside the hundreds or thousands with which many Web-based reader comment systems contend. (To be fair, IETF participants are expected to read most messages, while participants in Web-based reader comment threads seldom do so.)

Before the advent of the Web, online discussion was thus carried out using a relatively small collection of standard technologies embodied in specific application software. Compared to email or Usenet news, the Web presented a flexible and standard platform. It was possible to develop and deploy new applications without requiring users to download new software. The use of fonts, images and layout also placed design in the hands of authors, where previous applications had standardized utilitarian interfaces.[3] There was a flood of experimentation as developers produced Web-based versions of formats from discussion forums to email.

The wheel was reinvented many times for the Web. With no existing precedent (at least no popular one; it is likely someone somewhere had experimented with something similar), reader comments, however, were a novelty. Web-based reader comments, in which per-document discussion and the software to enable it are combined on a single site (often on a single page) appears to have emerged not as a translation of existing online practices to the Web, but as a new way to engage with a new format: the blog.

The Web made it practical for many people to post articles online for viewing by a

3 This outcome was contrary to design: HTML was originally intended to convey only content and structure (“semantics”), not formatting, allowing readers to decide for themselves how information should be presented. Browsers still have settings for default typeface, text and background colors. This vision of user empowerment is still influential in some technical circles, where the desire of designers to control design is sometimes considered destructive to communication, accessibility, and innovation.

large audience. The blog format stabilized in the late 1990s (Nardi et al., 2004); it presents posts by an individual or group in reverse chronological order. As the popularity of blogging exploded in the early 2000s, proponents often described it as dialogical, taking place within a "blogosphere" where bloggers responded to one another on their respective blogs. Popularity, a primary indicator of success, meant being acknowledged through hyperlinks, which PageRank had made the currency of the Web. The egalitarian ideal of blogging was participation by all; the reality, as many critics pointed out, was an "A-List" of bloggers who benefited disproportionately from network effects and attracted the vast majority of readers. For blogs farther down the popularity curve, building a sizable audience was very difficult and a huge bone of contention. Over time it also became apparent that the frequency of updates required to run a popular blog was a time-intensive activity that few could afford to sustain.

Regardless, adding the capability of readers to comment on a blog post was a logical extension of the ideal of dialog—not to mention a representation of the attention craved by bloggers. Several blogs had integrated comment systems as early as 1998. *Slashdot*, an important site I include in my study, was among them.[4] These typically allowed readers to post remarks beneath a blog post. Between roughly 2000 to 2002, comment features were integrated into a number of popular blogging systems (or supported through third-party software) including *LiveJournal*, *Blogger*, and Movable Type.

As the money-making potential of the Internet continued to grow, these new comment systems began to encounter the same challenges faced by Usenet. Usenet had survived for many years by relying on social norms or human moderation. Now the Internet was big business, and it was open to a much more diverse population of users than previously. What had worked at one time in relatively homogeneous small communities did not work at scale. Open comment systems were vulnerable to being flooded with spam. It would be years before spammers targeted virtually every blog with comments, but for very popular blogs the problem struck early, along with the prevalence of trolling.

4 *Wikipedia* claims that Open Diary was the first, but the article it cites on the topic circularly refers back to *Wikipedia*. The sole surviving *Slashdot* comment from 1998 is from August, and is a reply to another comment. *Slashdot* was probably not the first, but it was early.

Slashdot was one of these. It was one of the first (if not the first) to develop a user-driven ratings system in an attempt to manage high volumes of comments without employing an army of moderators to censor inappropriate comments. There were other reasons for *Slashdot* being a target for bad behavior. First, it did not require users to register: to this day, it permits anonymous comments, does not pre-approve comments, and claims to not remove comments unless legally required to do so. Second, members of its technical readership were particularly prone to gaming the system or engaging in other off-topic play, such as writing first posts that often read simply “first post”, or even simply “fp” (O’Brien, 2004).

Rob Malda, who created *Slashdot*, had established the main elements of the site’s rating system by late 1998 (Malda, 1997). Over the years the system evolved into something quite complex compared to most other sites, but the essential features are representative of how most reader comment rating systems work. Readers vote posts up; a discussion can then be filtered to show only high-rated posts. Low-rated posts, including those that are offensive, off-topic, or spam, are hidden. Aggregate ratings generated collaboratively help to reduce the need for human moderators and manage comment volumes. (In *Slashdot*’s early years the site’s popularity was exceptional: when a story linked to a third party site, that site often crumbled, its servers unable to cope with the influx of readers. The phenomenon was sufficiently common that it became known as “the *Slashdot* effect.”)

In the late 1990s online journalists and editors were pursuing the influential myth of “participatory journalism” (Domingo, 2008). The blog phenomenon was one manifestation in this blurring of the lines between journalists and their audiences. Amateur “citizen journalists” appeared to challenge the authority and security of traditional journalists; debates raged about whether they should be considered legitimate journalists (Domingo, 2008, p. 688). Mainstream publications felt the need to respond.

The integration of comments can thus be understood as one move in a larger effort to engage with readers online. Measures ranged from the use of hyperlinks (external hyperlinks were rare in early online newspapers, probably in part due to fears that readers would click away and not come back), email addresses for columnists, forum discussions with editors, live chats, and so forth. One early innovation was the

addition of discussion forums to web sites. A 2000 study of 100 online newspapers (Kenney et al., 2000) found that 17% featured “discussion groups.” In 1999, Schultz (1999) studied 100 American online newspapers to see who they used four major modes of reader interaction: email, polls and surveys, chat rooms, and discussion forums:

> Some offered an almost overwhelming variety of forums. . . . 13 newspapers had more than 30 different discussion forums, and two papers had between 16 and 30 different forums . . . A wide range of topics was covered from “discussions” about cooking, movies, or sports, to public affairs, political, and economic issues. . . . Only rarely (seven sites) were forums linked to articles or Web sites that provided background information on the discussion topics. (1999)

If thirty was a large number, these cannot have been comments attached to individual articles. But in the early 2000s many papers started to experiment with setting up their own blogs. Blogging brought with it existing technologies and genre practices, such as RSS syndication—and comments. It seems likely that this was the avenue through which commenting spread from blogs to newspapers. The influence of blogging is made clear by Ted Vaden, a newspaper ombudsman, describing the difficulties of moderating online comments in 2007: “We’re trying to conform to the blogosphere culture, which is one of freewheeling debate and resistance to censorship” (Loller, 2007).

Vaden cautioned that commenting “seems to provide an opportunity for racists and various other kinds of unpleasant comment” (Loller, 2007). He was not alone in his concern about the behavior of commenters. Hermida and Thurman (2007, p. 14) document worries among newspaper editors about control, quality, possible legal liability and the impact on their brand. Yet there was also a sense that offering opportunities for reader participation was simply something that publications were compelled to do: “editors jumped on board because of the perceived need to be offered greater levels of interactivity with readers.”

Despite misgivings, reader comments have become established and valued features on many major news sites and blogs. It is difficult to pin down why so many news outlets adopted reader comments at roughly the same time, but my recollection is that the spread of reader comment systems took place in the mid- to late-2000s. This was the period following the launch of YouTube in 2005 with its comments and its thumbs-up/thumbs-down voting system—a system very similar to those that appeared

on many news sites. The successful use of reader reviews by sites like *Amazon* may have had an impact too. Commenting was no longer a practice associated with blogging, but with the Web in general.

The use of commenting is still in flux as sites experiment with new systems and approaches. In 2013, *CBC* transitioned to a new system, *Huffington Post* required the use of real names through Facebook, and the *New York Times* tweaked their system. The use of external services like Disqus or Livefyre to provide the technology has been widespread. Many sites have also implemented other forms of online reader feedback, such as Facebook pages and Twitter streams in stories. Reader comment technology has not stabilized. Implementations vary significantly from site to site as experimentation continues. Studies like this one could help inform the practice and technical design of reader comments.

Relationship to a Primary Article

Subordination to a primary article is the defining feature of reader comment discussion: the characteristic that makes them *comments*. Articles are written in a space controlled by their authors; comments are written in a space controlled by someone else. A reader comment discussion is thus the product of two different groups: the author of the article and the participants in discussion. Much of the distinctiveness of reader comment discussion can be traced to the differences between these groups and to the relationship between discussions and the articles to which they respond.

This division between different groups is not universal in comment discussions. *Facebook*, *Google+* and *Reddit* all feature comments among peers. The same is true of many forum discussions initiated by a substantial post to which others respond. These are best described simply as comments, not *reader* comments. As I use it, the latter term defines comments in relation to a central activity of reading. Emphasizing that commenters are readers implies that while they write, they are not Writers. This distinction accords with common use: forum, Facebook and Reddit posts are hardly if ever referred to as reader comments, while comments on blogs and news stories often are.

Reader comments are subject to two kinds of authority proceeding from their

relationship with the host site. The first proceeds from ownership: the owners of the website on which discussion take place make and enforce rules about participation in discussion. The second is textual. Comments are subordinate to the primary article to which they respond; they are invariably less prominent, either following the article or accessible only as a result of user action (e.g. clicking a button or scrolling down). The article sets the agenda for discussion and provides the initial interpretation of whatever topic is at issue; it is the "primary definer" (Hall, 1978, p. 58). The site also sets the tempo of discussion: as new articles with attendant comment discussions are published, existing discussions are sidelined.

Online reader comments are consistent with the older tradition of annotation. Prior to falling out of fashion in the late 19th century, the hand-written annotation of books was a widespread and often social practice (Jackson, 2001). Annotations were valued and shared; individuals would annotate a text specifically in order to share their notes and reactions with others. Books were commonly given as gifts complete with personal annotations written specifically for the recipient. Annotations also presented a challenge to the authority of the printed text. This applied to handwritten notes: but also to printed marginalia. The margins of pages were sites of conflict over religious orthodoxy in the 16th and 17th centuries (Tribble, 1993). The practice was invited by the wide margins of Gutenberg's Bible. Later printers published different translations of the Bible; they often also included margin notes instructing their readers how to interpret the text. These were political interventions in an era when a challenge to the authority of religious orthodoxy was often also a challenge to the authority of the state. The lack of margins and minimal annotations in the King James Bible were part of a deliberate effort to limit such diversity of interpretation.

Like annotation, reader comments present a challenge to the authority of an article. They are becoming one of the main channels of public response to stories in influential news media, many of which include reader comment discussions with their stories. Journalists have begun to use these discussions as resources for further reporting. The CBC radio program *The Current*, for example, reads a sampling of reader comments on-air; the CBC website sometimes publishes stories about reader reaction to previous articles. The *New York Times* includes excerpts from reader comments in some articles, as does *Slate*. I have seen reader comments printed in the Letters section of a

local paper, the *Burnaby News Leader*. The technology news site *Ars Technica* has run stories drawing on reader comments, both to assess opinion (e.g. in the case of Edward Snowden) and when commenters revealed additional information about a story (separately, *Ars Technica* published a profile of Snowden drawing on his writings in reader comments on that site). One *Guardian* story (Arthur, 2014) about Microsoft's mobile strategy quoted extensively from comments on an *Ars Technica* article. Journalists are thus selecting comments to elevate to the center of discourse from the margins.

The authority of a primary article is magnified when it is produced by a major respected news organization like CNN, the *New York Times*, CBC, or the *Guardian*. Publications like these play a special role in public discourse. Reader comments are a place where that authority can be questioned and challenged before a wide and diverse public audience. Even if reader comments on such sites were in other ways undifferentiated from discussion forums and social media, their placement—in broad public view, but also in the margin—would make them significant in a way that discussions oriented around particular interests could not be.

Surges of Attention

The authority and prominence of a primary article also in large part determines who participates in reader comment discussion and structures the relationships among these participants. The logic of attention of authored content (primary articles) is different from the logic of dialog. Association with a primary article thus has a major influence on the reader comment discussion that is attached.

Original content operates in a winner-takes-all world of popularity and attention, of hits and also-rans. Popularity compounds: the more popular something is, the more people talk about it, link to it, and so forth—and the more popular it becomes. Theorists studying network dynamics refer to this phenomenon as preferential attachment (Barabási, 2003): a node with many inbound links is likely to attract new links at a higher rate than a node with few links. The result is a long-tail distribution, in which the most popular works are vastly more popular than the next-most popular works, and so on

down the curve to a much greater number of numerous virtually unknown works.[5] Similar dynamics produce the unbalanced non-normal distributions of various social phenomena, from music popularity to individual wealth to the number of inbound links and visits to web sites. The larger the potential audience and the easier it is for them to communicate, the faster and more dramatically attention is likely to slosh from one target to another. The Internet is particularly prone to long-tail distributions.

The pursuit of such popularity is a major motivation for media sites and blogs. The business models of the former depend on traffic; the latter have prized attention since the early days of the A-list. Articles on blogs and news sites are first-class content on the Internet: usually corresponding to a unique URL, they are referred to and passed around online. In general, while most discussion sites pursue a certain level of activity, they are not driven by a quest to maximize popularity. Beyond the level of participation required to sustain worthwhile conversation, popularity may actually be a bad thing for forums.[6] Nor are links to individual discussion threads and posts disseminated widely. Even when the content of a forum post becomes extremely popular, the attention usually falls elsewhere, on blogs and news sites that report on or excerpt from it. Discussion forums, like most forms of online discussion, seldom experience the scale of attention surges directed at news and blog articles.

The audience draw of the primary article channels into comments a larger number of users who might not otherwise come. While this is obviously true of mainstream news stories, it also applies to personal blogs which occasionally experience large spikes in readership (and commentary) when they publish widely-propagated articles. The posts about Aaron Swartz by Lawrence Lessig and Orin Kerr in my study are examples of this. Many of these visitors have probably never visited that blog before; many will never do so again. When this happens, any existing community on the site is likely to be overwhelmed by strangers reading, posting and voting.

---

5 I avoid the term “power law,” as it has a specific mathematical definition. Many distributions claimed to be power law distributions do not meet the definition. For my purposes the descriptive but less precise term long-tail distribution is adequate.

6 I have observed on several occasions online communities of interest that split into a dominant forum and competing secondary ones. In these instances participants in the smaller forums sometimes criticize the dominant forum, typically for its moderation practices and scale and express a preference for the less popular forum to which they subscribe. I wonder whether there is a logic that drives communities to split in response to the pressures of excess scale.

Put another way, one can draw an analytic distinction between two groups of reader comment participants driven by distinct logics. The first are regulars who come for the comments; they are attracted by consistency: a social milieu with which they are familiar. In aggregate, like participants in a discussion forum, their numbers are fairly consistent. The second group are those who come for the novelty, originality or popularity of the article attracts them. Some of them are regulars too—but on occasion many of them are not. Their numbers can fluctuate wildly. Having come to read, some may convert into commenters, often only temporarily. For a popular post, the logic of attention for the primary article can overwhelm that of regular discussion. When the group of novelty-seekers is large, enough may convert into commenters to overwhelm the regulars. Variance in the former group is low; if they seek a particular community of discussion, then this is the only place they can find it. Variance in the latter group, however, is high: the novelty they seek is in constant motion in the winner-takes-all attention economy of original work. For sites whose profile is high enough to benefit from significant network effects, the volume of reader comments may often be dominated by converted readers rather than by returning commenters. This transforms the character of discussion for everyone.

Such popularity surges throw together in discussion people who may share few existing relationships, social norms, or interests. In general, in comparison to forums and other discussion formats, the diversity of participants in reader comments is likely to be relatively high. I say relatively because sites do develop reputations for certain audiences, such as *Fox News* for right-wing commenters.[7] That said, many news sites attract such a large general readership that the subset of readers interested in a particular article are in many cases likely to be mutual strangers. The key point is that when it is the content of the article that attracts commenters one should expect the diversity to be higher than when participants come (or in most cases return) for the discussion or community.

Given the dynamics of long-tail distributions one might expect article-driven attention surges to be the exception, not the rule. In my experience, the number of

7 I recall reading a comment on a *National Post*, a right-leaning Canadian paper, by a commenter who rather plaintively asked critics from the left to please stay away so that conservatives like himself could have a space of their own.

participants in many discussions is indeed small, most of them regulars, engaged in discussion that may be indistinguishable from what takes place in a discussion forum. This may well be the case for most reader comment discussions. Yet while the number of small-scale discussions may be large, *the discussions themselves are still small*. Much of reader comment participation takes place in large-scale discussions. These may not be the *typical* cases, but with respect to the distinguishing features of reader comment discussion they are the *characteristic* ones. They take place on influential web sites, about issues of interest to a wide audience, among a more diverse public. From the perspective of public political discourse by ordinary people with little political commitment, they are likely to be the discussions that matter most.

High-volume discussion represents something of an ideal toward which reader comments aspire (discussion forums, in contrast, likely to be damaged by popularity surges, seldom aim for unlimited volumes of comments). A reader comment discussion is one iteration in a series on the same site (corresponding to a series of articles), all of which make use of the same socio-technical system. Due to the long-tail logic of attention, scale is unpredictable. Many of the design and conventions of reader comments are responses to the challenge of and aspiration to volume, even though the majority of discussions may involve relatively few commenters (as my data in Appendix A illustrates).

Many of the social and technical features of reader comments systems can thus be traced to the challenge of dealing with high volumes of unacquainted users. Fragmented discussion, incivility and aggression are commonplace. Commenters often reply to those they disagree with, rather than those like them. They seldom address one another as unique individuals, instead focusing on what has been said or applying stereotypes. Participation in a discussion is typically intense but brief; often the vast majority of posts are made within 24 hours or less. Where in discussion forums it is not unusual to see a commenter directed to refer to something said earlier in a thread, in large-scale reader comment discussions there is relatively little evidence of collective memory among participants in discussion, or indication that participants have read the entire discussion. Moderator involvement on popular sites is minimal, usually limited to removing material that violates site policies; moderation tasks are often delegated to user-driven rating and filtering systems. Rating systems lower the barrier to participation,

giving occasional participants a greater voice. Ratings play a key role, both directly through voting and filtering comments, and indirectly because commenters are conscious of the presence and attention of a larger unseen public.

The dynamics of large-scale reader comment discussions thus attract a particular sort of participant. Novelty-seekers have *low commitment*. They are not activists or party loyalists. They are likely to have less connection to the topic at hand. All of these factors contribute to the distinctive character and significance of reader comment discussion.

Technical Design

Comment systems implement combinations of technical features seldom found in other forms of discussion. These include hierarchical threading, systems that allow readers to vote individual comments up (or down), variable sort order and filtering, only minimal information about participants, and heavy use of JavaScript. Although none of these features is unique to reader comments, and while not all systems incorporate all features, a combination of them is typical for reader comments while relatively rare for other discussion systems. Many of these features help deal with large comment volumes.

The Facebook and Google+ social media platforms incorporate comment systems of their own. Users of these services can post entries to which other users can respond with comments. These discussions are mostly carried out among friends or members of a community, but like blogs they can sometimes draw larger audiences. Some news sites (e.g. *Huffington Post* and the *Vancouver Sun*) use Facebook's comment system for their reader comments.

Voting is particularly characteristic of reader comments. Most reader comment systems provide some mechanism for filtering or promoting high-scoring comments so that readers can concentrate on those without reading whole discussion. Such voting systems are very rare in discussion forums, likely because competition for popularity could lead to incivility and fragmentation.

Facebook and Google+ have like "like" buttons (or Google's "+1") button for posts (articles) as well as comments. From the point of view of Facebook and Google, however, these serve a very different purpose. A vote for a reader comment is about the

comment itself: its quality, popularity, and so forth. Its primary use in discussion is to aggregate opinions and filter comments based on quality. A like button is much less about the content itself than it is about the person who clicked it. Unsurprisingly, most reader comment votes are anonymous, while Facebook likes and Google +1s are prominently associated with the users who made them. Facebook is more interested in the *who* than the *what*. Likes provide Facebook with data about users and Facebook's customers with attention; they are not primarily a mechanism for filtering content (though they may be used to filter users). A Facebook like is thus primarily about the person who clicked the button. That information is valuable, as is the attention produced: but the quality of the like—whether it was clicked thoughtfully or in good faith—is irrelevant to Facebook. Of course Facebook has no interest in a mechanism for voting something down.

On a newspaper site or blog, in contrast, votes are used to filter comments and manage attention. For most of these sites, unlike Facebook, producing content is part of their core business (regardless of whether that content is the means to deliver eyeballs to advertisers). The quality of the votes and comments matters to the quality of a site as a whole; the identity of the people who made them is relatively unimportant.

Hierarchical threading is also much more prevalent in reader comment systems than it is in discussion forums. Most forums feature a flat sequence of posts, earliest first. These are typically grouped into higher level threads, but there is usually little sense in which different threads, although they may be logically related, are part of the same discussion (though participants may be members of the same community). An individual thread is generally a sequential coherent conversation from beginning to end. If a significant digression occurs, it is likely to be curtailed by moderators or split off into a separate thread. In contrast, a hierarchical system allows in-context digressions at the cost of a more fragmented flow of discussion overall.

Filtering and sorting controls further break up the coherence of reader comment discussions. Options offered by many sites include: oldest first, newest first, top-rated by users, and editors' picks. A participant in discussion cannot assume that other participants are seeing the same posts in the same order.

All of this complexity is typically supported by dynamic behaviors implemented

with JavaScript in the browser; users can then sort, filter, or expand and collapse comments without reloading the discussion. Long discussions are usually not loaded all at once: a reader must click to load more comments into a given thread or at the end of the discussion. Discussion forums, by contrast, are not usually so reliant on JavaScript; instead of loading more comments in-place, they break a discussion into multiple pages.

Scripting comes with a cost: search engines index static content. When that content is loaded only at behest of users interacting through scripting, it is invisible to them, as it is to archive sites (e.g. the Internet Archive) or to most web crawlers. Reader comments are a significant part of the online landscape, but most of them are absent from indexes and archives. When sites encounter problems or upgrade their technology old comments are often lost to history. This is a problem shared with social networks, albeit for a different reason: it is in the interest of Facebook and the like to lock-in users by making it difficult to extract content from their systems. Discussion forums using regular HTML pages do not suffer from this invisibility. Unless they require a log-in to view or deliberately exclude web crawlers by implementing the robots.txt exclusion protocol, discussion posts are easy to find through search engines and present in Internet archives. Discussion forums often provide expert advice and wisdom to non-participants who encounter them through searches. Reader comments almost never do.

Another use of JavaScript is to outsource reader comment management. It takes a lot of work to implement and maintain a secure, spam-free, full-featured comment system handling large volumes of comments. Many blogs and news organizations instead integrate with an external service like Disqus, using JavaScript to embed commenting into web pages. These have the additional effect of making participation easier (and thus attracting even more low-commitment users), as users can often log in with an existing login profile rather than creating a new one on each site.

Outsourcing is unsurprising, as with rare exceptions (e.g. *Slashdot*), reader comments are not the core business of the sites that have them. If they are not worth the cost or effort, they can be jettisoned. Austin Frakt explained why he disabled comments on his blog:

> we all just kind of put our heads together and realized we weren't enjoying moderating comments. We thought the costs were higher than

the benefits, so we decided to shut them off by default. (Tremonti, 2014)

For a medium devoted to discussion, like a forum, this is not an option. The only alternative to adequate moderation is to shut down entirely. Forums therefore often feature substantial rules of conduct and vigorous human moderation to enforce them. Bans, suspensions and warnings about user conduct are commonplace. Not so in large-scale reader comments. On sites with high comment volumes moderator intervention is typically rare, blunt and impersonal: a moderator simply removes offending material without comment (banning users may not be worth it when community bonds are weak and so many visitors arrive with popularity surges). Technical features are widely used to transfer moderating tasks to users: buttons for reporting bad behavior and voting and filtering systems free site maintainers from many of the fine-grained interventions common on forums. Many news sites, it seems, perceive little marginal benefit to improvements in moderation. Moderation needs to be just aggressive enough to make reader comments worthwhile; beyond that, participants are usually left to their own devices. The *CNN* discussions in my study appear to be examples of such lax moderation.

Fragmentation & Difference

Most modes of communication are characterized by unity and coherence. Consider newspapers, face-to-face communication, and discussion among friends on Facebook, for instance, which feature persistent relationships among participants who recognize each other as individuals apart from what they say. When more than a handful of people are involved, some take on leadership roles for guiding discussion; these leaders are recognized as individuals transcending the discussion itself. There is a scarce resource to be shared (time in the case of face-to-face discussion, space in a newspaper) and an expectation that it will be shared rationally or fairly. There are strong social pressures to adhere to norms of participation and towards achieving some sort of common ground. Participants who fail to respect these norms are subject to censure, exclusion or editing. The result of is a degree of narrative unity: discussion may digress or go in new directions, but each point of departure is connected to the tissue of what came before.

In large-scale online comment threads social pressures are weaker. Most participation is effectively anonymous; emphasis tends to be on what is said rather than who says it. This is reflected in personal attacks ("CONservative," "lieberal", "repuglican") that target stereotypes rather than particular individuals. Dialog is thus primarily among comments rather than individuals, each of whom is not granted the same time, space, or respect for what he or she says.[8] On many sites norms are weak or the consequences for violating them are minor. There is little expectation that consensus will be achieved; one rarely finds commenters expressing a change of heart.

This is supported by studies of *Slashdot*. Halavais (2001) argues that *Slashdot* tends to bring together diverse perspectives on a topic. Gómez, Kaltenbrunner and López (2008) found that it lacked a "complex community structure": users showed only a slight preference for writing replies to others to whom they were related by previous reply exchanges (Gómez et al., 2008). They suggest that replies most often the result of differences of opinion.

Reading a comment thread one encounters what appears to be a relatively homogeneous stream of text; in fact, each comment is a separate fragment of individual expression. The "rolling present" (Xin, Glass, Feenberg, Bures, & Abrami, 2010) of the forum, in which discussion progresses and is focused on the latest post, is absent from hierarchically-threaded reader comments in which each thread replies to the article rather than to the comments that have gone before. Coherent exchanges are separated by random remarks. Even taken together, the comments in a discussion represent but a fraction of the other readers who have passed that way, raters among them. It is therefore a mistake to treat comments as roughly equivalent atoms of discussion or as parts of an organic whole. Comments are neither all independent, nor are they equal and separate. They are interdependent and unequal. There is a structural hierarchy in the formatting of the discussion and its threads, and an emergent hierarchy that arises from the support given by other comments and reader ratings. The significance of comments varies radically: some reflect the views of no-one, not even their authors; others are broadly representative.

The structure of an authored text both reflects and determines the significance of

8 Trolling to attract attention may sometimes reflect an attempt to achieve the attention that in other modes is enforced through norms of reciprocity.

the elements that make it up. It is characterized by unity. The ordering of stories on a television news program, for example, reflects editorial judgments about what does matter and about what should matter. A comment thread is not designed by a single mind or organization. Each comment is authored, but the discussion as a whole is aggregated. While structure influences how it is interpreted (e.g. early comments have more opportunity to attract attention), and while that structure is not entirely arbitrary (comments respond to one another), it is only one influence among many.

Different media have different relative biases: towards common ground in the case of face-to-face discussion, perhaps less so for newspapers. Relative to most forms of communication, the bias in large-scale reader comment discussion is towards heterogeneity.

All communication entails a tension between the identification that makes understanding possible and the difference that impels it (Burke, 1969). This dialectical relationship is necessary for the spread of new ideas, as described by Rogers (1995). Most communication takes place between people who have much in common: “When two individuals share a set of similar characteristics, common meanings, and mutual value positions, communication between them is likely to be effective, which is rewarding, thus encouraging homophilous communication” (1995, p. 127). Yet the diffusion of new ideas requires difference. “The information-exchange potential of dyadic communication is related to the degree of heterophily between the transceivers,” he writes (1995, p. 142), defining heterophily as “the degree to which pairs of individuals who interact are different in certain attributes” (1995, p. 142). From the perspective of information exchange, discussion degenerates when either tendency dominates.

The contrast between high-volume reader comment discussions and more community-oriented forums throws this tension into relief. Where community or affinity is the basis for participation, discourse itself may be sufficient to establish discursive convergence. Large-scale reader comment discussions, with their wider publics and heterophilous participants, require additional mechanisms to prevent diversity from tearing them apart. Reader comment systems negotiate divergent argument and convergence in part through their technical designs. Threading systems promote difference in the pattern of dialectical replies, while rating systems aggregate votes to

highlight convergence. These exist, in a sense, on different planes: difference is manifested qualitatively through the text of individual arguments, while convergence is represented in quantitative ratings. The success of such measures depends not only on technical design, but also on social norms and the character of the community.

Yet communication is not only about information exchange. As performance, for example, consensus or divergence may not be problems. A dialogic model of communication, moreover, is a temporal and linear model: it assumes a process that progresses and develops over time. Only then are convergence or divergence possible. But many reader comment discussions are not linear: they are broken into threads, digressions, and isolates. Many comments address the primary article without regard to what has gone before. As dialog proceeding through time, such discussions are fragmented. As discussion unfolding in space, however, they cohere around repeated themes and practices, comprising a multi-dimensional terrain.

Interpretation

> It's as though when you order a sirloin steak, it comes with a side of maggots.
>
> — Gene Weingarten (2010), *Washington Post* columnist, on reader comments

Pinch and Bijker's social construction of technology theory (SCOT) emphasizes the impact of different interpretations of technology by different groups. Latour (2005) argues for the importance of paying attention to categories defined by actors themselves. If reader comments really are a distinct format, one would expect evidence of this in the interpretations of users.

Terminology provides one indication. Sites and participants consistently distinguish reader comments from other forms of discussion by calling them "comments." Nearly every site with comments uses the term, in links, to label the section where they are shown, when stating how many there are, in pages and articles discussing comment changes and policies. By contrast, forum discussions are almost never referred to as comments; the word "forum," or more rarely "discussion" or "board," is most commonly used in headings, link text, and so forth. The consistent naming of comments as

something different than other forms of discussion is an indication of a stable and distinct interpretation. A weaker distinction is drawn between reader comments, which respond to a primary article by a different group, and a broader category of comments embracing responses to friends on social media and the like. Although the former are sometimes referred to as reader comments, the latter almost never are.

Criticism of reader comments provides more evidence that they are perceived differently. While Weingarten's assessment above is harsh, similar sentiments are common. Brossard and Scheufele (2013, quoted subsequently) compare reader comments to the Biblical Fall of Man. Barry Ritholtz's sarcastic instruction to commenters on his blog that they should demonstrate ignorance and dishonesty shows that even those who operate reader comment sections often express low opinions of their worth.

To be fair, reader comments, like other social media, host some of the most vile invective on the Internet. It is not difficult to find misogynistic, racist, and hateful reader comment threads on mainstream web sites. In her survey of journalists, Nielsen (2012, p. 98) found that many were concerned about the level of bigotry in comments. Sexism is particularly rampant, notably in relation to technology and video games (as 2014's "gamergate" controversy made clear). For example, Anita Sarkeesian's criticism of sexist tropes in video games attracted "a flood of violent comments and emails" (Robertson, 2014). *Jezebel*, a web site oriented towards women, had "gifs of violent pornography" posted in reader comment discussions (Jezebel, 2014).

Criticism often focuses on reader comments as a format. Other formats, such as discussion forums, mailing lists and blogs, or friend-based social media are seldom the targets of the same kind of broad criticism. While critics may attack a *particular* forum (e.g. for moderation policies or the behavior of users), criticism of reader comments is often of the format in general. There seems to be something different about reader comments and how they are perceived.

Weingarten's simile indicates what that might be. Comments are like maggots alongside the article that is like a sirloin steak: not only are they bad in themselves, but they accompany and detract from something far superior. This contrast with the primary article is nearly ubiquitous in criticisms of reader comments. Implicit or explicit is the

challenge that that comments pose to (legitimate) authority and authorship.

Several scholars have studied how comments influence the perceptions of the primary article to which they are attached. In an experiment, Kareklas, Muehling and Weber (2015) found that comments strongly influenced research subjects perception of a health public service announcement (PSA); when the commenter was perceived as an expert, the comment had a greater impact than the PSA itself. In one study, Anderson, Brossard, Scheufele and Xenos (2012) found that negative comments influence the perceptions of infrequent readers of reader comments; in another, Scheufele, Anderson, Brossard, Xenos and Ladwig (2013) found that a lack of civility polarizes readers' perception of the risks of controversial technologies. This latter result is the basis for their comparison of reader comments to the Fall of Man.

This research has been very influential. In September 2013 *Popular Science* decided to disable reader comments for the majority of their stories. The decision became a story itself, drawing criticism and praise from numerous cites across the web. There was clearly widespread concern about the format. Suzanne Labarre (2013), *Popular Science*'s online content director, cited Brossard and Scheufule when she argued that "Comments can be bad for science":

> commenters shape public opinion; public opinion shapes public policy; public policy shapes how and whether and what research gets funded . . . because comments sections tend to be a grotesque reflection of the media culture surrounding them, the cynical work of undermining bedrock scientific doctrine is now being done beneath our own stories, within a website devoted to championing science. (Labarre, 2013)

Such fears collide with the ideal of participative journalism. Marissa Nelson, director of digital media at CBC news, says of reader comments, "My job is to engage as many Canadians as possible with our content" (Tremonti, 2014). Kelly McBride with the Poynter Institute argues that news media that aim to "create civic engagement" have a responsibility to engage through comments, otherwise "they're failing their audience and they're failing to fulfill their journalistic duty" (Tremonti, 2014). Nelson suggests that the expectations of readers have changed, as have the responsibilities of journalists: "if they're talking about it they're much more engaged, they're more likely to be civically engaged and I think that that's part of the CBC's mission."

For Nelson, reader comments are important to public discourse. I have described what reader comments are: their marginality and relationship to a primary article, their history and technical design, how they are perceived. My task now is to lay a theoretical foundation for understanding what comments and commenters *do*.

Chapter 3: Theoretical Framework

> In the beginning, the technology gods created the Internet and saw that it was good. Here, at last, was a public sphere with unlimited potential for reasoned debate . . . Then someone invented “reader comments” and paradise was lost.
>
> — Dominique Brossard and Dietram Scheufele (2013)

Explicit in many of the studies of reader comments, and implicit in many of the critiques, is an ideal of public deliberation. While many reader comments are as thoughtful and rational as the articles to which they respond, reader comment discussion in general seldom lives up to any ideal of reasoned consensus. On the contrary, comments are frequently argumentative, even combative. Incivility and irrationality in comments threaten to undermine the authority of journalists and experts. When climate change deniers bombard a science report with long discredited denunciations or when partisans reduce every issue to an attack on their political opponents, reader comment discussions hardly seem to constitute rational deliberation, especially when contrasted to the journalism to which they are often juxtaposed.

But rationality is not the only possible standard. In what follows I aim to construct a theoretical framework for the understanding how reader comments operate as a public space in democracy. I begin by outlining Habermas’s (1991) theory of the public sphere. I then consider Arendt’s (1959) understanding of world-building through public action, Warner’s Arendt-inspired work on public spaces (2002), and de Certeau’s (1984) analysis of tactical action in space. I propose that comments can be seen as tactical public actions through which participants construct a shared but uneven space.

Democracy and Publics

Much of the thinking about public discourse and deliberation has been developed in response to Habermas’s (1991) *Structural Transformation of the Public Sphere*, in which he developed a historical conception of a space in which bourgeois individuals, their independence secured through their ownership of property, set aside their particular interests to discuss matters of the general good. This role was reserved for the bourgeoisie: others, due to relationships of dependence (on the state, on capitalists),

were not free to set aside their interests and so were incapable of taking a broader view. Property ownership secured the independence necessary for the bourgeoisie to act disinterestedly, while the liberal ideal of equality led them to imagine themselves as representatives of society at large. Not every man had achieved economic independence, but every man had that opportunity: an opportunity secured by the rights of property. The private interests of the bourgeoisie in defending their property rights thus coincided with the universal interests of all men.

Of course the liberal ideal was a myth; Habermas's analysis is in part an immanent critique of the tension between liberal ideals and liberal practice. Even at the height of classical liberalism state action played an essential role in securing the foundation of capitalist accumulation and inequality (Linebaugh & Rediker, 2000; Polanyi, 2001). Yet from this myth developed a very real political subjectivity among the bourgeoisie that led many of them to sincerely take on the role of pursuing the public good, bracketing their private interests in order to debate matters of general importance, to expose state action to publicity, and to transmit public opinion. These became embodied in institutions such as a free press, free speech guarantees, and democratic representation. Habermas's nineteenth-century bourgeois public sphere does not overcome domination, but it does propose democracy in which reasoned deliberation and debate, shielded from particular interests, is a guiding ideal for mediation between the state and private individuals.

Habermas examines the conditions that gave rise to the bourgeois public sphere, and those that led to its decline. Its foundation was the separation of discourse from political and economic interests. Without economic independence participants would be unable to set aside their private interests. In the twentieth century this distinction eroded: welfare state policies and Keynsian management of the economy entangled the state with private business and private life and public opinion was increasingly manufactured by private interests. The private realm, which had previously encompassed both economic activity and family life, collapsed down to the latter, while public discourse narrowed to focus on leisure and entertainment.

Surely the most influential critique of Habermas is by Nancy Fraser (1992). While characterizing his work as an "indispensable" starting point for thinking about public

discourse, she digs into flaws in his original conception. Focusing particularly on the exclusion of women, she argues that the exclusions of the bourgeois public sphere were not just incidental flaws: they are key to understanding it. For bourgeois men, to take on the role of universal man in the public sphere was to become a member of an elite club. They were, Fraser (1992, p. 114) writes, “preparing to assert their fitness to govern.” The status of the sphere itself was similarly defined by exclusions: for it to be *the* sphere for public discourse it had to exclude other competing publics. Fraser (1992, p. 116) writes, “exclusions and conflicts that appeared as accidental trappings” of the public sphere “become constitutive.” She (1992, p. 117) summarizes an argument by Geoff Eley: “This is the shift from a repressive mode of domination to a hegemonic one, from rule based primarily on acquiescence to superior force to rule based primarily on consent.”

Participants in the bourgeois public sphere were supposed to leave their particular differences and inequalities at the door in order to deliberate as equals. Fraser argues that this in itself was a form of exclusion. In practice, status cannot be bracketed out. The self-presentation as a universal man was far from neutral: it took bourgeois culture and manners as its standard, compared to which other modes were inappropriate and intrusive. Moreover, setting inequalities aside made them invisible. The poor cannot protest their poverty while obliged to conceal the reality of their everyday lives. Explicitly, of course, the bourgeois public sphere was bourgeois and it was male: most people were not welcome there.

Against this bourgeois space, excluded groups engaged in what Fraser calls subaltern counterpublics; in these, the inequalities and differences of participants were often defining features. Even if they lacked the influence of *the* public sphere, they enabled participants to develop their own identities and discourses in contrast to the dominant one. These alternative spheres, Fraser argues, are important staging grounds for difference and opposition to develop, from which they can emerge to challenge and influence dominant discourses and the state.

The point of Fraser’s critique is not that Habermas is blind to the existence of the inequities and biases of the bourgeois public sphere. He has no illusions that its participants really were the universal men they imagined themselves to be. The more substantial challenge is contrasting understandings of discourse in general. Habermas

idealizes the development of understanding and common ground, something that is seldom apparent in reader comments, while Fraser describes how they can actually conceal and propagate hegemonic domination. In advocating *competing* publics as a democratic response to inequality, she shifts from a vision in which deliberation is adequate to achieve understanding and consensus towards one in which inclusion is fostered by agonistic competition and difference.

Space and Action

Reader comment discussions seldom achieve understanding or consensus, unless they are united in opposition to something. Yet unlike Fraser's (1992) competing publics, antagonism is often internal. Hannah Arendt (1959) argues for a different agonistic politics oriented not around democracy or the domination of elites, but freedom and meaningful human life, which for her derive from the unique potential of every human being for unpredictable action. She agrees with the ancient Greeks that private life is privation compared to the freedom and equality of life in public. The liberal elevation of the sovereign individual, for whom public discourse is a means to negotiate and pursue private interests, thus represents a threat to human freedom.

For Arendt, freedom is not a condition: it is *action*, a practice to which she contrasts labor and work. Labor is the realm of necessity. All human beings need food, shelter and so forth; the purpose of labor is to secure these things. Labor does not make us different from one another, nor even from animals. Work is making: it is effort towards achieving a predetermined end, whether that be crafting a television commercial or achieving social justice. While the things we make may distinguish us from animals and from one another, they do not define us. We are always more and other than the sum of our works; we are never reducible to them; they do not reveal who we are.

What makes action and freedom possible is plurality, the birthright of every human being: each of us has the capacity to initiate something new. When the pursuit of ends does not guide our actions, we make choices and perform deeds that disclose who we are. The action we undertake is unpredictable, unique and original to the individual who begins it. Moreover, it is only a beginning: for action to be meaningful, it must be performed in view of others. It then sets in motion chains of consequences that exceed

the survey of the actor herself. Only others can perceive the full implications of our actions. Without their witness and memory, human action is ephemeral. We are the agents of our own stories, says Arendt (1959, pp. 163–4), but not the authors.

Acting in view of others is insufficient, however: for action to reveal who we are, for it to be remembered, it must be recognized as such (Zerilli, 2005). The realization of human freedom is to act *and be seen to act*. The recognition of action requires judgment. For recognition of action is not a necessary consequence of the deed: it cannot be reduced to reason; its novelty cannot simply be subsumed under an existing category or concept. Judging is therefore an active practice. Zerilli writes,

> Spectators do not produce judgments that ought to serve as *principles* for other judgments or for action; they create, rather, the *space* in which the objects of political judgment, the actors and actions themselves, can appear, and in this sense alter our sense of what belongs in the common world. If the world, as Arendt argues, 'is the space in which things become public,' then judging is a practice that alters that world. (2005, p. 160)

Action is more than an individual performance before others: it is an interaction with them, in which their capacity for unexpected action confronts one's own. Where Habermas's communicative interaction is oriented around common understanding, Arendt celebrates agonistic struggle in which there is no goal of consensus. Nonetheless, this agonistic interaction does give rise to something common: it discloses a shared world. "To live together in the world means essentially that a world of things is between those who have it in common,", she writes, "as a table is located between those who sit around it; the world, like every in-between, relates and separates men at the same time" (Arendt, 1959, p. 48). The world is not simply an objective given. Public action defines it and makes it real:

> For us, appearance—something that is being seen and heard by others as well as by ourselves—constitutes reality. Compared with the reality which comes from being seen and heard, even the greatest forces of intimate life . . . lead an uncertain, shadowy kind of existence unless and until they are transformed, deprivatized and deindividualized, as it were, into a shape to fit them for public appearance. . . . The presence of others who see what we see and hear what we hear assures us of the reality of the world and ourselves . . . (Arendt, 1959, pp. 45–6)

As I have argued, reader comments are marginal: there is an unequal relationship between comments and the article to which they respond, and between comments and the actions of others that may be described in the article. This asymmetry is present also in Michael Warner (2002), who develops Arendt's insight to develop a theory of publics and publicness, encompassing actors and spectators.

Warner (2002, p. 67) describes a public as a self-organized "space of discourse" brought into being by the mere act of addressing it. When one addresses a public, one does not speak to a known group of others, but rather to an ambiguous audience whose composition and extent are unknown and unknowable. Membership in a public is determined by each individual who chooses to grant his or her attention. The public is a relation among strangers: even if members of a public happen to know one another, that prior relationship is not what brings them together as a public. The address of public speech is therefore not to actual people, but to an imagined audience. The speaker knows this; so do the audience—but what draws them to pay attention (and the speaker to speak) is the imagination or expectation of some latent or emergent shared interest. The address of public speech, Warner says, is thus both personal (in that we give it attention because we identify with it) and impersonal (in the sense that it is directed at that identity, not at us as unique individuals).

Taken alone, these characteristics are sufficiently broad to encompass a film screening, a radio broadcast or a street performance. Warner emphasizes that participants in a public cannot be so cleanly divided between speaker and audience. The relation of publicness entails circulation: public speech is expected to circulated and to elicit responses. Publics thus have a temporality: different patterns of circulation are characteristic of different publics.

A public is not produced by the mere act of speaking, nor is it reproduced by the mere act of replying. In order to contribute to the circulation of a public, speech must anticipate the kind of public being addressed. One might say it must be felicitous. But public action does is not simply determined by the world it seeks to address: it also helps to realize that world, to bring it into existence. Public speech seeks its public; it also *creates* its public:

There is no speech or performance addressed to a public that does not

> try to specify in advance, it countless highly condensed ways, the lifeworld if its circulation . . . It then goes in search that such a public exists, with greater or lesser success . . . Public speech lies under the necessity of addressing its public as already existing real persons. It cannot work by frankly declaring its subjunctive-creative project. Its success depends on the recognition of participants and their further circulatory activity, and people do not commonly recognize themselves as virtual projections. They recognize themselves only as being already the persons they are addressed as being and as already belonging to the world that is condensed in their discourse. (Warner, 2002, p. 114)

Here is Warner's obvious debt to Arendt. Arendt's spectators judge action; in doing so, they alter the world. Warner's public chooses to attend; in doing so they affirm the world imagined in the address to which they respond. Here too is the implication that publics are capable of political action. A public can, at least a little, bring into being the world that it imagines and develop the subjectivity of the audience it addresses. This even applies to the idea of publicness itself: Warner sees the public as a modern innovation, one might say a cultural technology, that has spread around the world. In a sense, publics created a world of publics.

Unlike Warner, who embraces mediated publicness, Arendt rejects media, defining action as the only form of relation among people that can be direct and unmediated by things. While she leaves little space for technology, her ideal of non-instrumental action recalls Dewey's analysis of technology (Hickman, 1990). Both Dewey and Arendt are concerned with open-ended exploration and experience. Arendt is interested in discovering who people are; Dewey is concerned with investigation or learning more generally. Dewey's instrumentalism, despite its name, rejects final goals in favor of temporary ends-in-view. These ends-in-view are themselves instrumental: jumping-off points from which we continue our investigation by choosing subsequent temporary goals. For Dewey, even the ideas we have and the theories we construct are instruments for further learning. Likewise, technology can temporarily be an end rather than a means as part of that larger exploration. For Arendt, politics can occasion action; for Dewey, ends-in-view occasion investigation but are not its final objective. If technological mediation is accepted into Arendt's category of action, it is then not as a means to a final end but as an instrument of exploration.

Still, the distinction Arendt draws between work and action is not always clear.

Action is not driven by ends, but it can be occasioned by them. Had we no ends in mind at all, be they ever so transient, we would never act. It seems also that there is an element of action in the works we produce. Arendt's action takes place in a common world; the solitary work of the romantic author with no thought given to others is then clearly not action. Yet this is more romantic fiction than lived reality: our deeds and words are often embodied in works we make.

Refocusing the disjunction of work and action makes possible an alternative description, in terms of planning. This ability to plan and control is the basis of another asymmetric distinction drawn by Michel de Certeau (1984, p. xix), between strategy and tactics. The strategic actor, he explains, controls the space or institution from which he operates; with control comes the capacity to predict and plan. Tactical actors, lacking this control over space, are unable to plan. "The place of a tactic," writes de Certeau (1984, p. xix), "belongs to the other." Tactical action is thus opportunistic and unpredictable, taking advantage of circumstances as they arise. In our roles as readers and consumers our everyday lives are full of tactical reactions to and interpretations of a culture we have little personal role in producing.

The unpredictability of tactics grants a freedom that escapes the rationalized analysis and planning of the strategist. To borrow an illustration from de Certeau, the planner can build the city, but he cannot choose our route through it; he can trace the route we have taken, but what we have made of the experience—a smile to a passer-by, the scent of flowers in a garden—escapes his measurement. The diversity of individual experience eludes the homogenized traffic mapped by a planner.

Means and ends are bound by the temporal logic of cause and effect. The unpredictable, non-strategic practices that concern Arendt, Warner and de Certeau all take place in space. Publics (not least public *spheres*), worlds and in-betweens all organize space. Freedom is realized not in the certainty of strategy or the harmony of consensus, and not in a sequence of causes and effects, but in the plurality of contingent actions and judgments taken at the right moment.

What of politics? Warner does not dwell on the political implications of the worlds publics build. Jodi Dean (2002) accuses him of disavowing his publics as mediators with the state, for this would entail the kinds of exclusions and limits that he seeks to escape.

For Warner, it is the idea of publicness itself that makes actual publics possible. Dean argues that this idea is ideological: it serves elites by offering audiences an avenue for personal expression that promises significance but because of its separation from actual decision making never delivers. She recapitulates this in her (Dean, 2005) argument that the communicative capitalism of the Internet turns communication into contributions, encouraging people to redirect their political impulses into speech that is circulated but never heard.

On the one hand, the insistence by Arendt, Warner, de Certeau and Habermas on the inherent value of non-instrumental human action continues a long tradition of critiques of rationality that degrades human life and dignity. Goal-directed politics must not be its own justification lest human beings become means rather than ends. On the other, the world we construct through public action is the same world in which we conduct politics. Action of the sort Arendt and de Certeau celebrate is not driven by politics as it is often understood (in terms of the achievement of concrete policies or social outcomes), but politics of this sort is implicated by the practices of freedom they advocate.

Comment Spaces

Taken together, Arendt, Warner and de Certeau describe public action before spectators in a space the actors do not control. This action is not directed towards predetermined ends; rather it responds to the contingencies of the moment. Spectators are strangers, making the imagined audience real; moreover, their attention, judgment and memory realize the originality and freedom of the actors. Spectators are actors too: their responses to action make consequences unpredictable. The recursive circulation of actions and responses produces and confirms the shared world inhabited by all.

This shared world, however, is not a world of consensus. It never emerges as an objective entity; it is constantly subject to challenge and judgment. There is never a moment when actors perceive their agonistic divergent actions and judgments to have produced *the* world. The significance of this world-building is not what it accomplishes, but the performance of the participants. Nevertheless, the space it creates is one in which politics is carried out.

This, I propose, is a fair description of large-scale reader comment discussion. This takes place in a marginal public space. Surges of attention bring together mutual strangers, each of whom may read, rate or write. Each reader comment is an unequal and opportunistic response to the strategic actions of a blogger or journalist about actions by third parties in a world. It is thus tactical both in the sense of de Certeau (commenters do not control the space and are unable to plan) and of Arendt (few commenters are working towards produce a strategic end).

The tactical character of comments is revealed by social norms: many of the activities considered antisocial in online discussion, such as trolling or shilling for a particular interest, are strategic. According to the implicit norms of most discussions, participants are expected to represent themselves authentically, representing their views rather than acting in pursuit of an external goal or on behalf of a third party. Although authenticity might seem to be at odds with the anonymity of so much online discussion, there are different kinds of anonymity. Pseudonyms refer to stable identities; for those otherwise unacquainted with an individual there may be little difference. In discussion forums, the covert use of multiple pseudonyms (e.g., to support one another's arguments) is referred to as sock-puppetry and is usually grounds for censure. Trolling is another strategic violation of authenticity, for it is the deliberate misrepresentation of oneself or one's views in order to provoke others. The same is true when participants feign sincerity while acting on behalf of a third; such "shilling" is considered socially unacceptable.

These problems are more apparent in discussion forums where they are often grounds for action by moderators, and where pseudonymous identities form more long-term relationships. In the drive-by conversation of large-scale reader comments with their focus on what is said over who says it they are much less so, though criticisms still arise. The point I wish to make is that both of these formats tend towards tactical rather than strategic action by participants, although I also argue that the marginality of reader comments renders them even less strategic.

On its own terms, a given comment is usually a coherent work of authorship, in pursuit of something like Dewey's ends-in-view. In a larger context, however, it is unpredictable and only loosely connected to the many other tactical contributions of

other commenters. Rather than a progression of discussion there are many smaller threads and isolated comments responding directly to the original article.

Fragmentation is thus a consequence of the marginal status and tactical character of reader comment discussion. Yet to say reader comments are fragmented is to measure them against the linear development of dialog. Understood spatially, what may appear to be fragmented in temporal terms instead define a space surrounding and relating to the primary article. One might imagine the article as a central point to which comments connect, with threads radiating out in multiple directions. While it is uncommon to encounter an argument developed through dialog, large comment discussions almost always contextualize the article relative to divergent opinions, claims and above all *judgments*.

I am getting ahead of myself. I began this discussion with theories of deliberation because they dominate existing scholarship. Yet I was skeptical of it from the beginning; deliberation is an imperfect fit at best for much of what is said in reader comments. Were I to begin my research with too strong or too early a commitment to theory, I would have difficulty reading comments on their own terms. Yet one needs a standpoint from which to investigate. Reader comments are evidently public. Rather than embracing an ideal of a public sphere, I turned to the broader understandings of publicness described by Warner and Arendt. Judgment emerged from my analysis, and from the method that I developed through preliminary research.

Chapter 4: Preliminary Study

Qualitative research is a craft skill developed through experience. My research began with different subject matter and different methods from those I ultimately decided on. What follows is a description of my experience as I began gathering and analyzing data, including the mistakes I made and the things I learned. Scholars are biased towards positive results; when a hypothesis fails it is often not reported. But failures matter: they may not tell us what we wish to hear, but what they say may be just as important.

My initial intention was to analyze copyright discourse on *Slashdot*. My impression, based on over a decade of participation in *Slashdot* discussions, was that over time copyright discourse became more radical and increasingly articulated with other politics. I came to suspect otherwise: indeed that the opposite is happening. The same arguments are circulating over and over, but radical discussion has moved to other topics; meanwhile *Slahsdot* itself has become a shadow of its former self. The larger problem for my study was that my approach to the problem never resolved in a tight body of evidence that could make an interesting argument one way or the other.

First Attempt: Canadian Copyright Bill

I wrote my own analysis software, into which I could download reader comment discussions. It allows me to annotate and code passages of text, then query or filter a discussion, for example showing only comments with certain ratings, by certain authors or with certain codes assigned by me, always keeping comments in the context of the larger discussion. This allows for rapid exploration of very large amounts of data. Appendix C provides a more complete description of the software. Because the software is mine, I was able to experiment with new capabilities. Figure 2 shows the comments from Figure 1 being coded in my analysis software.

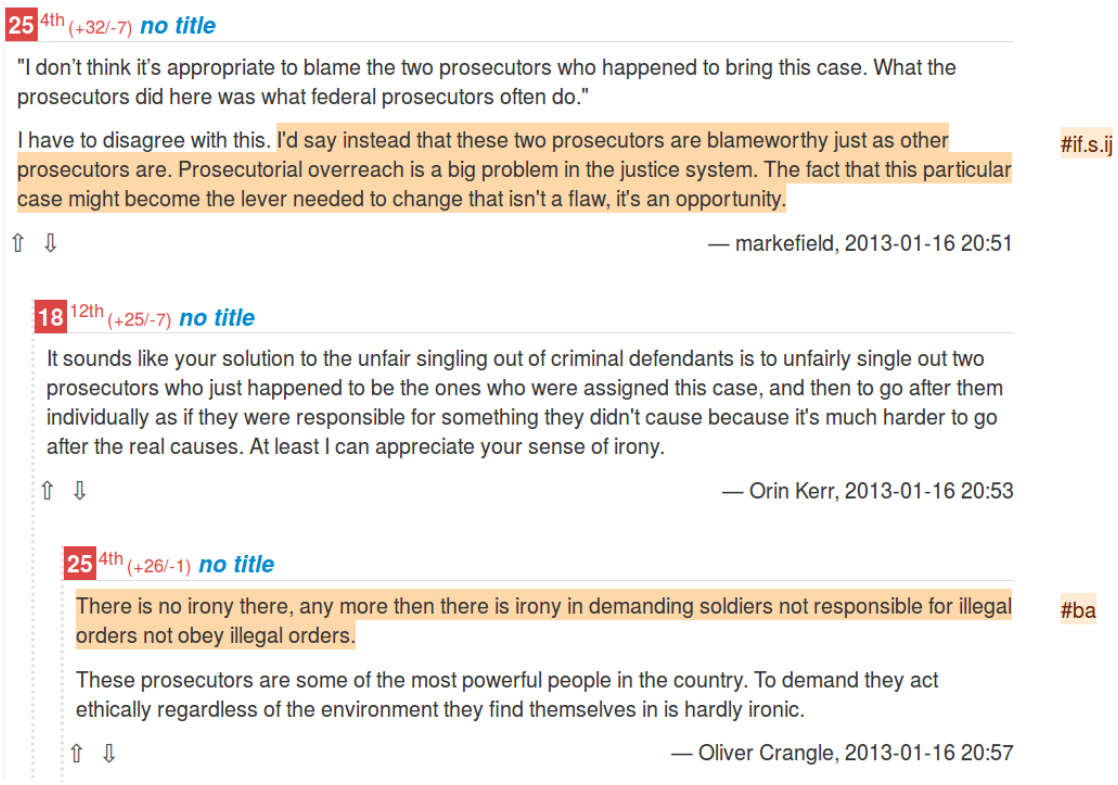

Figure 2: Comment coding in analysis software

I needed to try out my software and approach. In 2010, the Government of Canada introduced Bill C-32 (later passed unchanged as Bill C-11), which was to implement the World Intellectual Property Organization (WIPO) treaties, embedding in Canadian law provisions similar to those in the American Digital Millennium Copyright Act (DMCA). This was a key event in a multi-year struggle in which I had been an activist organizer.

Inspired by Gamson (1992), I decided to apply content analysis to count themes. I did not prepare a set of codes in advance. Instead, I read every comment in the discussion and developed codes as I went. This resulting set of over 30 distinct codes. I was curious about overlapping or related codes, so I enhanced my software to calculate co-occurring codes. The strongest relationship that jumped out was between the legitimacy of the law in Canada and criticisms of pressure from the United States. Although the number of co-occurrences was low (single digits), the relationship seemed potentially significant. (The matter of legitimacy—a topic that interested me greatly—

would stay with me through to my final research.)

I had planned to look at several web sites. The *Slashdot* data was interesting and my approach was producing promising results; each additional site would require programming work to parse its comments plus some way to reconcile results. I decided to focus on *Slashdot* alone. I was greatly concerned when *Slashdot* announced the site's sale to Dice, a technical recruiting firm: if they revamped the site or took old discussions offline I could lose access to my data set. Without further analysis I downloaded over 600 articles. *Slashdot* features a chronological article index; I selected every article with "copyright" in the title up to early October 2012.

I had written my parser with verification code to make sure I correctly parsed all comments. I soon discovered that many were missing: up to 40% in some cases. It turned out that *Slashdot* suffered from a bug with the static HTML archive. I reported the problem but; in the mean time I downloaded everything a second time using *Slashdot*'s JavaScript-based discussion system. This data turned out to be complete, except for early comments *Slashdot* itself had lost. I never heard back about the bug. I took this as lesson to take extra care to retrieve correct and complete data.

Was my selection criterion (articles with "copyright" in the title) a good one? I would need to justify my choice. *Slashdot*'s article index includes dates, titles, URLs—and comment counts. I downloaded all 470 pages of the list, parsed it, and dumped it into a database where I could analyze it. The results of this analysis were fascinating in their own right. They confirmed that "copyright" was a good choice of search term; also that copyright discussions attracted an above-average number of comments. The article list proved to be very useful later when choosing other articles to analyze.

Second Attempt: Most Commented Copyright Articles

So far I had looked closely at only one discussion. From the article index I chose copyright articles with the most comments, reasoning that these were a) clearly of interest to *Slashdot* participants, and b) likely to feature a range of discussion and themes. It soon became clear that topics like science, religion and American politics drove several of these discussions, and probably accounted for their popularity. Intense discussion about the tension between religion and science persuaded me that

participants greatly value self-identify as rational and science-minded. (Most are technical people with science training; computer science in particular has struggled to establish itself as a *science* (Ensmenger, 2010; Mahoney, 2011).)

These popular discussions consisted of over a thousand comments each. Rather than read everything, I decided to look only at comments with ratings of 3 or higher. This dramatically cut down on the analysis, eliminated most one-liners and "bad" comments (trolling and the like), and the comments seemed to be among the better ones written. In my final study, I would focus on top-rated comments. At this point, initial coding produced about 80 codes, including *uncategorized*, a code I used to indicate I intended to review a passage and assign a better code later. But there were problems.

First, many of the comments were not about copyright, while many of those that were about copyright were only about copyright. Coding them all would have been extremely time-consuming while producing only a small range of frequently repeated positions and claims. The context of comments in discussion, however, often implicitly implied relationships between themes. I decided to attempt to take context into account when assigning frames.

Second, I encountered a proliferation of themes with overlapping codes. For example, I found pro- and anti-corporate positions in opposition to each other. Simply coding "corporations" did not properly capture these difference stances. I went back to my codes and developed compound codes that combined topics with stances, and altered my software to "roll up" subsidiary codes when counting codes. The copyright theme illustrates the standard stances I applied across multiple themes:

- copyright—Discussion of copyright.
- copyright.pro—A passage expressing support for copyright.
- copyright.anti—A statement that copyright is a bad thing.
- copyright.critic—A criticism of an aspect of copyright, short of rejection of copyright as a whole.
- copyright.concern—An expression of concern about copyright or its effectiveness, but still expressing general support.

A given post could combine these: For example, a rejection of a specific copyright measure combined with strong support for copyright would be coded copyright.critic and copyright.pro.

Third, the *Slashdot* article appeared to have a strong influence on defining the topic and framing discussion. I wished to focus on what participants said, but there was no escaping the relationship of comments to the article. This was an early indication of my later concern with seeking unity in discussions, and would eventually lead to my definition of reader comments as responses to a primary article.

Fourth, connections between copyright and other issues were often implicit. For example, one of the most commented-on discussions was about a claim by software firm SCO that it owned the copyright to the Linux operating system. Despite detailed discussion about copyright law, contracts and tactics, many of codes did not show up. It was taken for granted that a legal victory by SCO would be devastating for GNU/Linux and for the Internet. Everybody knew this, so there was no need to say it. Codes could not tell this story. Their value was limited compared to a careful qualitative analysis of the discussion and its context.

I took away from this attempt an increased appreciation of the need to use ratings to choose which comments to focus on, the need to compare comments with the primary article, and indications that I would need to focus more on qualitative analysis of the data.

Third Attempt: Revised Coding Scheme

I coded several more discussions, but found the process time-consuming and error prone due to a proliferation of codes. I was reading discussions repeatedly yet found I was often coding the least interesting aspects of them. Certainly the codes could represent many perspectives on copyright and how it related to other issues, but most interesting were the exceptions: of course they were not represented by codes. For example, in a discussion about American pressure on New Zealand's copyright law, a discussion about the ethics of sharing included a high-rated comment by SteveTheNewbie (2011) about the privatization of rain water in Bolivia in response to pressure by the World Bank, linking copyright to the market colonization of common

resources. This comment interested me in part because it was so unexpected. Surprising comments like this introduced some of the most critical claims into comment discussion: but the regularity of codes could not capture what made them unique.

The occurrence of codes was strongly influenced by the article. Early comments also had a strong influence, often spawning long threads and digressions. The focus of attention and argument was highly contingent on what happened to be posted first. For example, a discussion about copyright could easily begin with a long thread critiquing two-party American politics.

The most interesting high-scoring comments were often responses to low-scoring ones. Despite their low ratings, those comments could play a key contextual role (e.g. linking different issues). High-scoring comments tend to attract many responses. The dynamics of discussion seemed to be in the ebb and flow of threads, something not captured by coding individual comments. I concluded that I would have to study threads, using high-scoring comments as attractors drawing my attention to others. All of this made coding very subjective and undermined the value of quantitative measures.

Other topics could not fit within the framework or level of detail of my codes. I expected this. The problem was the significance of these anomalous topics. Non-partisanship, for example, emerged as a major theme. Numerous threads digressed into American politics, yet there was virtually no partisan name calling: the consensus view was “a pox on both their houses.” I had seen vicious partisan discussions in comments on other sites. Not here. (Supporters of libertarian candidates sometimes show up, but debates about their merits are tame.) Nor did critiques of American imperialism, ascriptions of identity (“we” vs “they”), or emotional expression fit well with the codes I had chosen.

I had intended to develop a proliferation of codes bottom-up, then aggregate them as trends became clear. I now tried the opposite approach: coding high-level categories (and accepting false positives) as markers for places to return and themes to investigate. By keeping their number down and their generality up I could hope to more reliably capture unexpected but more interesting phenomena. Coding would also be easier because I was less interested in ensuring a correct count than in seeking out interesting discussions. Using high-scoring comments to focus my attention would lend

support to their significance without the need for reliable numerical counts.

I knew that views on climate change had shifted in *Slashdot* discussion, from libertarian skepticism to support for the scientific consensus. Similarly, I detected a political shift to the left during the aftermath of the 2008 financial crisis. I therefore sought change in copyright discussion. Earlier, the frequency and strength of claims of American imperialism had surprised me. I selected stories about international copyright issues from different years. I thought perhaps the financial crisis of 2008 would be a turning point. I looked at international stories from 2001, 2002, 2004, 2007, 2009 and 2011, and other stories from 2003, 2005, and 2010.

By now important themes were emerging.

Non-partisanship was a major one. In the threads I encountered about American politics from 2004 on, there is very little support for Democrats or Republicans. The much stronger sense is that the U.S. two party system is consistently failing to represent citizens or their interests. This sometimes slides into critiques of the legitimacy of the current order, including claims that government is effectively owned by corporations. The 2002 discussion, in contrast, featured partisan arguments. This suggested to me that there might have been a shift around the time of the Iraq War. Slashdotters have always been strong proponents of civil liberties and critics of government spying. Maybe the Patriot Act contributed to a shift in sentiment.

There was also a sense of identity in many copyright discussions, contrasting a “we,” whose interests are being ignored, with a “they” whose lobbyists are directing lawmakers. Who “we” are is often assumed, along with a consensus that copyright legislation is broken. Despite this, beyond an agreement that copyright terms are too long, there does not seem to be a consensus about what copyright laws should be.

Criticism of American imperialism was frequent and harsh in the later years. In 2002, top rated comments argued for realpolitik in pursuit of pursuing American interests, often out-scoring critics. In later years, imperialism and colonialism were brought up repeatedly. There is no indication that most of these criticisms are from non-Americans. Rather, in a number of instances the claim is that the U.S. is acting on behalf of corporations, and that Americans are equally victims.

Injustice showed up several times, particularly as a correlate of imperialism. Injustice is seldom individualized: it was usually characterized as irrationality. Commenters more often complained that copyright policy was economically irrational than that it hurt their individual interests.

Two of Gamson's three factors for collective action frames were thus present in many discussions. The third factor, agency, was weaker. There were some cases of technical agency, e.g. suggestions that copyright restrictions could be worked around with technology, and of political agency, e.g. contacting political representatives. The paucity of such action in most discussions was thrown into relief by the surfeit of it in the discussion about the Canadian bill C-32.

Rationality was a key recurring feature of discussion. The idea that arguments are won by evidence and reason seemed to be universally shared. Emotional outbursts occurred, but they almost never overwhelmed the argument a commenter was making. Someone might say "I am really angry," but then explain why. Similarly, expressions of empathy were nearly absent. Empathy was sometimes rejected explicitly, though that didn't happen very often either. This emphasis on rationality was evident in a long thread about copyright and intelligent design, which focused almost entirely on arguing for science over intelligent design (individual commenters argued for the latter, but comment volume and ratings opposed it). Slashdotters were keener to defend science than to critique copyright.

More generally, copyright was used as an issue to illustrate particular political points—corporations are controlling government, writing laws and so forth—but copyright itself did not frame discussions. It played a supporting role for other dominant frames, particularly the failure of democracy, corruption by corporations, and (occasionally) critiques of capitalism and consumer society. Copyright might play a politicizing role, but it is not the end-point of politics: only perhaps a beginning.

At this point, I was starting to question my focus on copyright. Yet my work here helped me improve my method of analysis. I realized I could not look at comments in isolation: I had to somehow consider the dialogic back-and-forth of threads and replies. I also realized that I could not use quantitative analysis as my entry into the data. From here on I abandoned my attempts to code precise features of discussion: I turned

instead to the broadest codes I could think of, so that they would encompass and serve as a map to the interesting anomalies I came across.

Fourth Attempt: Aaron Swartz

When Aaron Swartz died in January 2013, I was surprised by the volume, passion and duration of reader comments. The story was not directly about copyright, but I found the story compelling and noticed that comments were echoing some of the themes I had been analyzing—particularly expressions of identity and solidarity. I decided to do an analysis just of the Swartz case. The results transformed my research.

I collected over a hundred links to news articles, blog posts and discussions in the weeks following Swartz's death. I then analyzed a few particularly interesting discussions. The *Remember Aaron Swartz* (2013) eulogy site was a stand-out. It was not a reader comment discussion, but at this point I was interested in copyright discussion by ordinary people, and had not narrowed my interest to reader comments alone.

Gamson's (1992) three elements—injustice, identity, and agency—are evident in many of the eulogies. Gamson's unit of analysis is the conversation. To determine whether a conversation expresses a frame, for example, he assesses whether it is expressed without provoking opposition. His conversations possess a unity: it makes sense to talk about them as a whole. This is not true for the collection of Swartz eulogies, nor for sprawling comment discussions.

I started to code the extraordinarily rich *Remember Aaron Swartz* site with Gamson's three categories. I read all of the comments, but found they blurred together in my mind. It was difficult to maintain adequate mental concentration to treat them consistently. Unlike comments in a discussion thread, eulogies are independent. They do not respond to each other and lack ratings. With discussion threads, I thought, a random sampling would not make sense. Here it would. I enhanced my software to display random samples.

The sample worked beautifully. I read the comments in my sample much more carefully, coming back them again and again until I had deep understanding. I discovered some strong patterns. Friends of Swartz talked about their memories; people

who had not known him talked about solidarity with his ideals. There were many expressions of identity and agency. Here was a body of people who seemed not to know one another personally, but who shared a collective identity and values.

My excitement faded when I tried to apply similar techniques to other discussions. The first challenge was finding good discussions: comments for the *Globe and Mail* story about Swartz were missing. I mistakenly thought the same was true of the *New York Times* (to discover later that I had Javascript disabled). I settled on *CNN*, expecting comments of low quality. I needed something to contrast with the memorial site; if similar themes were present here too, I might have an extreme case—i.e., a case in which something is found though it is least expected.

I approached the *CNN* discussion casually, focusing on comments that jumped out at me. I found what I was looking for: people defending Swartz, arguing for the legitimacy of his actions, protesters against bad laws. This was promising, but I couldn't just pick what I thought was interesting. I needed a justifiable way to narrow 3,000 comments to something more manageable yet in some way representative.

At this point I had little hope that reader ratings would provide any real value. I had observed group-think in many reader comment discussions. I knew that early comments and ones that hit the right political buttons attracted votes, while detailed analyses often languished. Unlike *Slashdot* or sites with active moderators, I expected ratings to be essentially useless. Looking at the distribution of ratings seemed to confirm this: the distribution was heavily skewed, with the top comment receiving about three times as many up-votes as the second-highest rated comment, and early comments benefiting most. How could I compare ratings when the numbers were so obviously arbitrary?

I now had software for random sampling, so I started with that. The result was unimpressive. Many of the comments were off-topic. It seemed my initial impression was wrong: I had seen only what I wanted to see, but it was only a tiny minority of what was there.

I started to notice patterns however: certain users showed up multiple times in my sample of 50 (out of 3,000), contributing brief or off-topic comments. Perhaps the

sample was skewed by trolls. More generally, perhaps frequent commenters were less likely to provide interesting remarks. In the back of my mind was Swartz's (2006) own research about Wikipedia, which found that most edits were made by regular contributors, but most original content was written by people who only made a handful of contributions. Perhaps the same would apply here.

I looked at the distribution of comments by user and found a skewed pattern very much like the one for ratings. The vast majority of users only made one or two comments. A handful made large numbers of them: up to about sixty. The most frequent commenters wrote very little; often their comments were off-topic (one quoted arbitrary passages of Shakespeare).

I took a new random sample excluding anyone who had written more than five comments. A quick read through showed better comments, but still short of my initial impression. I tried several different samples just to be sure. The same thing held.

Still, perhaps there were other criteria I might use. The number of votes a comment received might not indicate much, but a single vote show at least someone found a comment valuable. This would exclude many late comments, but perhaps that would not be too bad a compromise. Gómez et al. (2008) found that reply counts on *Slashdot* correlated with reader perceptions of comment quality. I tried again, this time excluding prolific commenters, comments with no up-votes, and comments with no replies. The quality of comments went up. For each exclusion individually or in combination, the mean rating of comments matching the criteria increased markedly. I had found a way to take a random sample of what I subjectively found to be better-than-average comments.

Comments in the sample were mostly missing what I hoped to find. There were a few exceptions. Then I realized that the comments that engaged with the themes from the memorial site tended to be among the top-rated ones. I tried simply taking the top 50 comments to see what was there. Strong political positions were among them—near the top in fact. Maybe they were not the "best" comments that had jumped out at me on my initial read through, but they were substantial and interesting. Even though I had payed attention to ratings while reading comments, it was not until I extracted top comments systematically that they seemed useful.

I decided to analyze top-rated comments. The results were pretty good while the criterion was easier to justify than was my complex selection technique. Better still, ratings were a legible part of the discussion, comprehensible to and indeed produced by participants. Based on past experience, I considered only up-votes when selecting top comments. A comment with many up-votes can be buried by down-votes simply because people disagree with the politics. I did not want my selection criteria to exclude diversity.

After all that analysis, I ended up excluding much of my data and leaving unused many of the software features I had developed. I even used an algorithm to test whether the distribution of ratings was a power curve (it was not) and calculated Gini coefficients for rating distributions (large numbers of zero ratings skewed the graph; even without them, the results did not seem interesting). But I had made several important discoveries:

- Even on *CNN*, where the comment system and moderation were, in my subjective opinion, among the worst I have seen, ratings were as good a judge of quality as anything I could think of. This was a criterion I could use elsewhere.
- The content of a sample selected based on reader votes differed substantially from a random sample. This suggested that a random sample presented a misleading view of reader attitudes. Comments could not be compared, one with another, as though they were of equivalent weight. This would pose problems for any quantitative analysis of comments without regard to rating. My idea of counting themes and co-occurrences was an enterprise of doubtful worth.
- Pure qualitative analysis would suffer from a similar problem. The volume of poorly-regarded or trolling comments would produce a misleading subjective impression for the researcher, as could the length and substance of sparse thoughtful comments.

I confirmed this last point when I investigated expressions of empathy in the *CNN* discussion. The most important lesson I learned was that simply reading comments as one might read a conventional product of authorship like a news article or a book gives a very different impression than might be obtained by considering reader ratings. Attempting to take ratings into account subjectively does not seem to work; this is

unsurprising given their long tail distribution.

Ratings seemed to be meaningful, yet they posed challenges for qualitative analysis and content analysis alike. If posts are often dramatically unequal in their significance, how can counting occurrences across posts provide a meaningful measure of anything? Where reader ratings are present, then, the researcher must consider both qualitative and quantitative criteria. But how? Traditional statistics based on normal distributions do not apply. This realization led me to consider analyzing comment-reply pairs (exchanges), something I had been doing unconsciously for years in casual reading.

These early efforts laid the basis for the method I ultimately adopted and provided some of my data about Swartz.

Chapter 5: Method

There has been little qualitative research of reader comments as a distinct format. My study is therefore exploratory with respect both to data and method. One of my goals is to develop practical techniques for performing qualitative analysis of large-scale reader comment discussions. Briefly, I conduct a comparative study of several discussions (cases) about each of two news stories (the death of Aaron Swartz and the revelation of Edward Snowden) across several relevant news and blog articles. Within those discussions I have performed analyses of purposive samples of comments.

To explain my method, I first outline existing approaches to analyzing reader comments and similar online communication, compare my approach to case study method, describe frame analysis, make some remarks about ethics, and describe the discussions I chose to analyze and my rationale for selecting them.

Methodological Challenges in Analyzing Reader Comments

Three existing approaches have been widely and successfully applied to a range of discourse; I refer to these as symbolic, textual, and ethnographic. Their limitations, however, are particularly acute for analyzing public meaning-making in large-scale reader comment discussions.

Symbolic approaches, to give them a name, rely on the fact that online communications, like human language itself, is composed of discrete symbols (hyperlinks, words, etc.). Symbols are highly amenable to digital representation and machine-assisted analysis, allowing researchers to process very large quantities of data using deterministic criteria for selection and measurement. Consequently, there have been numerous quantitative studies using methods like content analysis (e.g. of words or tags), network analysis (e.g. following hyperlinks), and latent semantic analysis (e.g. assessments of the quality or sentiment of text). These studies can map out relationships, measure popularity, or track changes in subject matter (operationalized as Twitter tags, for example) over time, producing statistical measures with good claims to generality. Like other quantitative methods, however, they are strongly bounded by the initial categories chosen by the researcher; in practice, quantitative analysis ultimately relies on (and is often explicitly combined with) qualitative investigation for interpretation,

and in order to determine categories or other assumptions such as starting points for web crawls.

Textual approaches, such as discourse analysis or conversation analysis, entail close qualitative readings of text and meaning. Typical of qualitative methods, these tend to be much more open ended and provide richer results, but in comparison with quantitative approaches can only cope with limited data and are harder to generalize. They have been developed mainly for the analysis of strongly unified texts characterized by coherent authorship; in that context they excel at developing novel categories, teasing out subtle distinctions in meaning and finding commonalities across divergent media. On their own, however, they may be weak for accounting for quantities in data, especially extreme non-linear non-normal distributions.

Ethnographic approaches analyze online discussion through the experiences of participants, e.g. through interviews. This has produced a substantial body of research featuring rich descriptions of individual experiences, of how participants in online communication relate to others, develop their identities, construct meanings, and integrate the Internet into their everyday lives. As with other qualitative research data set sizes tend to be small.

In practice, researchers often combine approaches to take advantage of a mix of strengths and to address the specificity of the subject of study. My investigation of reader comments poses several challenges: scale (large numbers of discussions, comments, and users), diverse highly structured discussion systems combining textual and numeric data (featuring threading, reader ratings, user sorting, and so forth), wild heterogeneity of data within a discussion (some comments are highly rated and widely seen, others are virtually invisible), a lack of consistent user groups or relationships, and my focus on discourse rather than individual interpretation or experience.

While reader comments can provide valuable insight into the range and intensity of discourse, they pose problems for some techniques of analysis. Because there is no rough equivalence of comments, content analysis risks granting undue weight to prolific but unpopular viewpoints. Holistic qualitative analysis may be better able to cope with diversity, but many also read into a discussion unwarranted unity. Attempts to weight individual comments by ratings or other criteria may lose the important relational and

dialogic construction of meanings. With such large volumes of material, some technique of selection or rationalization is necessary to provide context for the researcher's subjective impressions of scale—yet integrating numeric data like reader ratings can be challenging.

Interviews and participant observation offer to reach beyond the differences between sites to understand experiences and motivations. Yet this a time-consuming approach cannot hope to address the volume and diversity of online discourse. Even were it practical in this instance, one of my aims with this research is to find ways to take advantage of that volume of data. Interviews could supplement analysis of discussions themselves, but they cannot substitute for it.

Another approach is computational, crafting algorithms to select and aggregate comment data. Analyses can consider ratings, reply counts, and the occurrences of words and phrases. Yet this cannot hope to tease out the construction of meaning in interdependent comments; moreover, it first depends on some form of qualitative analysis to define what to look for.

Studying Cases

I am interested in the dynamics and content of large-scale reader comment discussions. Studying multiple cases allows me to take account of the detailed context of individual comments and discussions. I examine a variety of cases across several sites with different characteristics that I must take into account, not in order to contrast sites so much as to compare discussions: for is discussions, not sites, that are the cases in my study.

I attempted to chose exemplary cases that clearly demonstrate the characteristics of large-scale reader comment discussions and differentiate them from other online formats. From some perspectives, these cases may be untypical or unrepresentative. Unlike many reader comment discussions, they are neither short nor dominated by trolling, spam or hate.

Though my method is similar to a case study, that moniker is perhaps misleading. Most case studies feature one or a small number of cases examined at a very high level of detail. My study examines twelve, considered at what I suggest is a moderate level of

detail. While I am not using the term, my work is informed by the case study approach.

Some critics have claimed that case studies are not generalizable. On the contrary, as Flyvbjerg (2006) argues, it is largely through cases that we learn about the world. Such learning entails the transformation of experience into cases in an active process of construction. Flyvbjerg (2006) explains that good cases are frequently the basis for generalization. Everett Hughes (1984, pp. xix) writes, “if one quite clearly sees something happen once, it is almost certain to have happened again and again. The burden of proof is on those who claim a thing seen once is an exception.” Flyvbjerg argues moreover that good case studies are the foundation of deep understanding in a field. Like Latour (2005), he argues that the important features of a phenomenon are often to be found in concrete details rather than abstractions. He (Flyvbjerg, 2006, p. 242) refers to Kuhn: “a discipline without a large number of thoroughly executed case studies is a discipline without systematic production of exemplars, and . . . a discipline without exemplars is an ineffective one.” The goal is not necessarily to boil cases down to a single finding, but rather to present a phenomenon in its ambiguity and complexity.

I argue that selecting reader comment discussions by random sampling would produce incoherent and misleading results. It would discover what is ordinary, not what is characteristic. It would likely miss out both on the heterogeneity of discussion (e.g. in terms of technical features and volume), while also capturing discussions that would be difficult to compare due to their diversity (e.g. in terms of relevance and genre). Examining a number of cases allows me to use my experience to select discussions that highlight relevant aspects of reader comments, and to develop the expertise necessary to analyze them, enabling me to examine a diversity of instances while also delving into each in some depth.

In order to broaden my study, I have chosen to examine a large number of cases (discussions); the trade-off is that I examine each in less depth than I might do otherwise. As I explain later, I do not look at all of the comments in each discussion, only a purposive sample.

I am constructing cases as a scholarly exercise. Participants in discussion are also constructing cases as part of ordinary interaction. The stories I have chosen are cases through which they construct their understanding of the world. World-building is

not an undifferentiated activity: it entails distilling a stream of experience into cases, not as academic exercises but as part of everyday meaning-making activity.

Frame Analysis

Once I have selected my cases, what am I to do with them? Frame analysis offers a way to approach the terrain of comment discussions and their world-building activity. At the level of individual comments and threads, it provides a means to categorize what is said. On a larger scale, the practices of participants in discussion are conditioned by the frames in which they are embedded. Frames can help analyze what people say: and what they do.

Participants in public discourse operate under expectations about their audience and their world. These expectations are molded by what Erving Goffman (1974), drawing on Gregory Bateson, calls “frames.” A frame is a mental model for organizing everyday experience through which individuals form conjectures about past events and expectations of future ones (Goffman, 1974, p. 38). A frame shared by multiple participants in an activity thus establishes a kind of unity for that activity.

For example, the audience frame in the theater creates shared expectations of decorum, respect and attention. Frames can enclose other frames. Goffman (1974) provides several examples. Within the frame of the theater is the frame of the play itself. The actors in a detective mystery are expected to follow the conventions of their performance. Further framings are possible, for example if there is a play within the play. Sometimes frames can exist in tension. For example, when a con artist runs a shell game there are two frames. The mark or victim of the con believes he is acting in a frame in which the game can be won, unaware of the enclosing frame of the con. By framing the situation for the mark, the con artist exerts control.

While Goffman was concerned with “scripts” for everyday life, frame analysis has become widely applied to the study of political movements: their formation and mobilization, how they interpret the issues they are concerned with, how they contextualize these relative to other matters and their interaction with the world at large. Environmental issues, for example, have been analyzed to discover how best to communicate in ways that resonate with particular audiences or motivate them to action

(Crompton, 2010; Lakoff, 2010). As in many of Goffman's examples, framing grants control. In politics rather than a single frame shared by all participants, there are multiple frames competing for dominance.

Consider the clash between pro-life and pro-choice factions debating abortion. Each attempts to impose its frame on the issue, not least in the name chosen to identify its cause. Pro-life advocates argue that the matter is one of human life, while pro-choice advocates argue that it is one of freedom. These are intended to be incommensurable positions. The conflict between them takes place in the context of a larger surrounding frame of public debate; typically this is a frame of rational discourse: each side tries to make a reasoned argument supporting one frame or the other. One tactic is to control this outer frame in order to control the inner one, e.g. by shifting from science to faith as the basis for deliberation.

Framing has become a major concern for mainstream politicians and political parties, often supported by empirical focus group and audience research to discover what frames and language are most effective for attracting and motivating supporters. Within the tradition of studying activist movements, Noakes and Johnston (2005) identify two main levels of frame analysis corresponding to the work of two different scholars: William Gamson and David Snow. The emphasis of Snow's research is on the strategic action of movements and activist organizations (Snow, Rochford, Worden, & Benford, 1986). The object of Snow's interest is not the content of frames *per se*, but the process of framing by which movement participants select and define frames.

Snow and his collaborators provide a number of categories for use in frame analysis. "Diagnostic framing" provides an explanation of a problem, "prognostic framing" proposes solutions, and "motivational framing" motivates collective action based on the diagnosis and prognosis. "Master frames" are particularly powerful frames that span multiple social movements. The *rights* master frame that emerged from the civil rights movement in the United States, for example, has been adopted by numerous social movements, even ones in conflict with each other such as pro-choice and pro-life movements around abortion (Oliver & Johnston, 2005, p. 187). Snow and his collaborators detail a number of "frame alignment" processes through which activists link frames to each other and adapt them to achieve movement goals.

Where Snow examines how organizations choose consciously to frame their communications in pursuit of their political objectives, Gamson's work is focused on how people interpret frames. He is less focused on the strategic action of movement leaders than the interpretation and communication of movement members and ordinary people. He places emphasis on the adversarial character of collective action (Gamson, 1992, p. 85). This is captured by the three components he argues are present in collective action frames: identity—who "we" and "they" are in a political struggle; agency—the idea that "we" can be agents of change; and injustice, which identifies how "they" are responsible for grievances. However, injustice frames can be discouraged by deflecting blame to "actorless entities such as 'the system', 'society', 'life', or 'human nature'" (Gamson, 1992, p. 32). Noakes and Johnston (2005, p. 5) summarize: "at its most basic, a frame identifies a problem that is social or political in nature, the parties responsible, for causing the problem, and a solution".

There are a number of debates and unresolved issues in frame research. Noakes and Johnston (2005) explain the difficulty in measuring frame resonance, defined as the effectiveness of a frame for its audience. The response to a frame is difficult to capture as the movement and the meanings of frames are in constant flux. Furthermore, studies often suffer from a lack of symmetry, focusing on resonant frames that are easy to identify and analyze, while non-resonant frames tend to drop out of sight.

The relationship of frames to ideology is an important point of investigation, one that Oliver and Johnston argue has been neglected by frame researchers. Ideology, as Oliver and Johnston understand it, is concerned with a coherent world view embodying values and norms. Frames, in contrast, are about content and presentation. The same frame can be used by people with vastly different values (just as the rights frame appeals to both pro-choice and pro-life activists), while a given ideology can be expressed through a range of different frames. Oliver and Johnston (2005) argue that the popularity of frame analysis coincides with an emphasis on market research and communication, and a focus on individual psychology. The weakness of this turn to frame analysis, they argue, is that the strategic struggle between frames loses much of the political substance and irreconcilable commitments of ideology.

Yet ideology and frames are intimately related. The relationship between frame and ideology is important even if it can be structured in different ways. The linking of frames to each other and to ideologies is very similar to the articulation of signifiers described by Laclau and Mouffe (1985). The main difference in the discussion is the emphasis in frame research on the strategic manipulation of meanings rather than their ideological significance. Articulation theory emphasizes the paradoxical fluidity of meanings. One example from struggles around copyright is the use of the word "pirate" by the content industries. Long emphasized as a pejorative in accusations of "theft," researchers discovered that the word might not be having the desired effect: described as such, "piracy" was often perceived as cool (perhaps this is not surprising since those industries also engaged mass advertising of pirate-themed media and merchandise)[9]

In elite discourse, strategic action is often primary and is therefore at the center of most movement frame analysis. Pursuing interests is usually more important than expressing values. Westby (2005) argues that movements negotiate a tension between ideological commitments and strategic considerations. They may choose frames for strategic reasons even when these frames do not precisely match movement ideology, creating a gap between frame and ideology (and perhaps shifting the ideological position also). For example, environmentalists convinced that economic growth cannot be sustained in a finite ecology may nevertheless emphasize recycling and energy efficiency rather than the reduction of consumption. The frame of efficiency is compatible with consumption: it can be co-opted by marketers, and ultimately steer an organization away from genuine sustainability.

Popular discourse online appears to be shaped by a wide range of motivations, many of them non-strategic. People write for a variety of personal reasons: because they are inspired; they are angry; they want to express their identity, solidarity, or difference; they want to get something off their chests; they are caught up in the call-and-response of argument and debate; and so forth. Unlike an organization structured around strategic goals (whether they be policy outcomes or organizational survival and growth), participants are not typically acting strategically to advance an end goal.

All else being equal, the relation of frame to ideology will be closer when

9 Regrettably, while I read news stories describing this discovery I have been unable to locate them.

discourse is less oriented around strategic considerations. This may explain the great range of popular discourse that I have witnessed online. Participants are not so constrained by interests and are thus more likely to express values. A commenter writing angrily about corruption or the illegitimacy of the law, or arguing that something is yet another example of the 1% oppressing the 99%, is probably expressing a personal conviction rather than taking a strategic stand.

That said, effective commenters are likely to appeal to their audience; the more they feel rewarded by replies and ratings, the more they are likely to do so. Commenters engaged in political action will also act strategically. For example, I have examined several instances of commenters mobilizing in opposition to copyright laws who post letters they have written or make recommendations to others for how to present their arguments in order to best influence elected representatives. The strategic tone and content of these letters is very different than other comments about the topic. Trolls, writing insincere comments intended to provoke social drama, also divorce what they say from any personal convictions, though some of these are obvious precisely because they violate the norms of a particular community or public.

Though more specific than ideologies, frames are often high-level concepts encompassing actors, problems, and solutions. But a frame is not an essential category: it is composed of ideas and its understanding and expression varies across people who subscribe to it. Most frame research focuses on small numbers of high-level frames. Johnston (2005, p. 242) writes, “few studies of collective action frames have produced in any detail the relationships among the different ideas that make up frames”.

Gamson’s work provides some starting points for attempting to understand framing discourse. He proposes identity, agency, and injustice as essential components of collective action frames—that is, frames that motivate people to act politically in groups. Identity differentiates between a “we” who have certain interests versus a “them” who oppose us. Agency is the belief that participants can have an impact on the issue at hand. Gamson particularly emphasizes the role of a sense of injustice which often motivates the other two.

Actors engaged in the world-building activities described by Warner (2002) are thus creating frames which may or may not be well-received by others. At the very least,

they project the frame of an imagined public with particular characteristics which may then be realized as a consequence of their action.

Gamson codes the occurrence of frames, but unit of analysis is the conversation. Its coherence and unity is fostered by the social dynamics of participants engaged in face-to-face talk; he therefore assesses the presence or absence of frames in the conversation as a whole. An injustice frame, for example, is present only if it is unchallenged by group members (Gamson, 1992, p. 37). This presents a difficulty in online discussion. Whereas in face-to-face discussion conversation may converge or at least be constrained by social pressures, reader comments are likely to instead diverge. As the widely-misapplied Godwin's law implies (Godwin, 1994), in any sufficiently long online discussion thread any point of view is likely to be contradicted. Moreover, reader comments are not of equal significance. Therefore, for the purposes of assessing the occurrence of frames in a discussion I consider only high-scoring comments.

Goffman's analysis indicates that all social occasions and spaces are subject to frames that lend them coherence, like the frame of a theater performance. In addition to the various frames occurring in comments within a discussion, my argument that reader comments exist as a distinct medium of communication suggests that they there is a frame or frames that encompass many or all reader comment discussions. The search for such a unifying frame guides my analysis.

My approach contrasts with that taken by Toepfl and Piwoni (2015) in their use of frames to analyze reader comment discussions. Their frames are based on an analysis of news stories rather than comments themselves; oppositional (consistent with their use of counterpublic theory, each frame counters a claim made by journalists); and narrow (frames are all mutually compatible, so that a comment could reasonably employ all simultaneously). The result is support for their claim that reader comment discussion "can foster the formation of counterpublics" (2015, p. 18).

My frames, in contrast, are primarily drawn from comments, not from news articles; are broader, corresponding more closely to stances, world views and attitudes than to arguments; some are mutually incompatible. I revert to Goffman's (1974) idea that frames shape participants expectations of what kind of action is appropriate. In this sense, frames should explain what commenters are *doing*, not just what they are *saying*.

They should encompass assumptions about meaning, not only conscious rational claims. My preliminary study led me to apply frames that are big and chunky rather than narrow and precise, which I then place within broader frames encompassing discussion as a whole.

Site Selection

Each of the cases in my study consists of a primary article (from a news site or blog) and the associated reader comment discussion. The universe of such discussions is huge and fragmented: there is no master list from which to choose (in fact, many discussions are invisible to search engines). I cannot choose discussions randomly without first defining a population, which raises the problem in a different form. In any case, I must choose discussions systematically so that they are similar enough to be compared, yet diverse enough that the quirks of a particular technical design or group of participants does not dominate.

Selecting a discussion means selecting the news or blog article to which it is attached. Unlike most discussions, such articles are first-class entities on the Internet: they possess unique URLs, they are present in search engines, and they are included in article lists on the sites or blogs to which they belong. While discussions are the objects of my study, I must find them by first choosing articles.

An article is defined by three related factors: the website on which it is published, the story that it is about, and a moment in time. Together, these determine a particular article from the universe of articles. Choosing discussions entails considering these factors.

As with discussions, there is no obvious “site of record” or master list for articles. In their study of general-interest news sites, academic blogs and think tanks, Frick, Guertler and Gloor (2013, p. 2) conclude that “There are no thinkers who really dominate the landscape . . . The era of the great authorities seems to be over”: attention, online and off, is divided among various specialized and general interest sites. The choices of stories and sites are thus interdependent. For brief, large-scale discussions attracting diverse participants, major news sites and stories they cover are obvious choices. These prominent sites also lend their articles and discussion greater significance in public

discourse.

Big news stories spread across days or weeks are covered in articles at critical moments when events drive coverage (and often comments), and by reflection and opinion pieces that appear at idiosyncratic intervals. I have chosen to focus on critical moments when a great deal of attention is focused on multiple sites telling their own versions of the same story, and attracting many comments. My choices are also informed by technical features (I chose only comment systems with ratings), particular relevance to the stories I have chosen, and personal experience.

Stories pose other difficulties. Certain highly politicized topics tend to provoke discussions that degenerate into name-calling; climate change, partisan politics and feminism, for example. Previous studies have examined dysfunction in reader comments; it is not difficult to find. I instead sought exemplary cases in which politically uncommitted commenters express radical views about topics that have not (yet) fallen into existing partisan or ideological frameworks.

I selected two topics about which I had already read extensively, and therefore possessed prior knowledge, expertise and personal interest: the death of Aaron Swartz and the revelation of Edward Snowden's identity. These stories do not correspond neatly to existing political or partisan positions (or at least they did not when the stories broke). They entail tactics of disobedience outside the rules of conventional politics and are of interest to technical communities whose sites offer distinctive commenting systems. The stories are similar enough to compare, but different enough to offer some diversity in discussion. They are not average: what these two stories cannot do—what no two stories could do—is represent the diversity of reader comment discussion.

These cases are therefore likely to under-represent ways in which reader comment discussions fail. In particular, these stories are about two young, middle-class white Americans whose privilege makes them less likely to be subject to exclusion and prejudice. Again, this significant limitation of the study is a deliberate one. That said, this choice makes any evidence of prejudice and hate all the more significant.

What follows are overviews of the sites I have chosen for the study. I will describe the Swartz and Snowden stories in detail when I analyze my cases, along with more

detailed descriptions of the sites.

The New York Times. The *New York Times* is generally considered to be one of the United States's papers of record. As of November 2013, they claimed an audience of nearly 30 million readers with a median age of 47 and a median household income of US$74,843, of whom 49% are male, 52% have college degrees, and 23% are professionals or managers (Times, n.d.).

CNN. I chose *CNN* as a major general-interest news site with vigorous reader comment sections, and because I expected the quality of discussion there to be relatively poor compared to the other sites in my study. As of December 2013, *CNN* (n.d) reports 100 million unique users per month with a median age of 39 and median house hold income of US$77,686, each of whom spends an average of 40 minutes per month on the site. The site claims that 46% of these visitors "did not visit a major news competitor in the last 30 days."

Slashdot. I chose *Slashdot* because of its unique discussion system, past studies of the site, and my depth of personal experience with its comment system.

Slashdot pioneered user moderation; one of the Indymedia centers adopted the software for some time (Milberry, 2003). Several scholars have studied it and its unique discussion and moderation system (Chan, 2002; Gómez et al., 2008; Halavais, 2001; Kaltenbrunner, Gomez, & Lopez, 2007; C. A. Lampe, 2006; C. A. C. Lampe et al., 2007; Saunders, 2006). Among computer professionals it was once one of the most influential news and discussion sites. In Gabreilla Coleman's interviews with free software developers, *Slashdot* comes up again and again as a place where they met with others, became enculturated into hacker culture, and learned of the connections between technology and politics (particularly around intellectual property issues).

Since about 2006, the site's popularity has declined. It competes with a plethora of other technical news sites, many of which provide more current and better researched news. There is a sense among *Slashdot* participants that they have aged with the site; that young hackers go elsewhere now. *Slashdot* effectively sat out of the SOPA protests in 2012; it is conspicuous by its absence in the study by Benkler and Roberts (Benkler et al., 2013) of sites that influenced participants in those actions. Still, *Slashdot* is

sometimes referenced on other prominent blogs, such as *Naked Capitalism*, *Boing Boing*, *Tech Dirt*, or *Ars Technica*. One of Edward Snowden's leaks about government spying revealed a program by the British Government Communications Headquarters to spy on *Slashdot* users (Cyrus, 2013).

Ars Technica. Founded in 1998, *Ars Technica* is a popular news site for technologists owned by Condé Nast (Technica, n.d.). I chose the site as a companion to *Slashdot*. The two sites appear to attract a similar readership, but differ in important ways. *Slashdot* seldom publishes substantial original content; *Ars* in contrast produces a significant volume of original reporting and analysis. The *Ars* comment system is unlike the others in my study. It resembles a forum, with a non-hierarchical list of comments from oldest to newest. Whereas *Slashdot*'s popularity has declined in recent years, that of *Ars* has risen, as has its influence. A study by Benkler, Roberts, Faris, Solow-Niederman and Etling (Benkler, Roberts, Faris, Solow-Niederman, & Etling, 2013) of the successful 2012 mobilization against the Stop Online Piracy Act (SOPA) found that *Ars* was one of the key news sites involved. Edward Snowden registered as an *Ars* forum user in 2001, and continued participating in discussion there until May 2012 (Mullin, 2013).

"When you advertise on Ars Technica," the site tells advertisers, "you are reaching technology experts and influencers, from CIOs, to systems administrators, to programmers, to devoted technology enthusiasts who live and work on the leading edge of computing." According to a voluntary survey of their readership, the vast majority of its readers (89%) do computer-related work (Fisher, 2013). Most are professionals; most are almost certainly male. The *Ars* "About" page features a photograph with sixteen of the site's writers, of which only two are women.

The Guardian. I chose the *Guardian* because it was the paper that initially broke the story of Snowden's NSA leaks, and was widely cited in other coverage. I did not include it for the Swartz story, however, as that *Guardian* article did not allow comments.

The *Guardian* and Glenn Greenwald, the journalist responsible and a long-time critic of the U.S. government, soon became closely associated with the Snowden story. Greenwald came to be widely perceived not simply as a neutral conduit for information, but as an interested participant with an agenda. Critics argued that *The Guardian*

violated a duty of journalistic neutrality. At one point agents of the British government destroyed a computer in the *Guardian* offices containing leaked data, claiming that the leaks were not in the national interest.

As of January 2014, the *Guardian* (n.d.) website boasts 90 million "unique browsers" per month, 29 million of them in the U.K. and 26 million in the U.S. Their audience is 54% male with an average age of 37 and a household income of over £50,000. The site specifically touts its appeal to "progressives."

Lessing v.2. While Lawrence Lessig's site is a personal blog, not a news site, it was widely cited about the Swartz story, attracting a surge of attention and comments. Lessig is a law professor well known in technology circles for copyright activism. Gabriella Coleman cites him as a major influence on free software developers and their understanding of copyright (Coleman, 2013). Lessig and his family were personal friends of Swartz; he was initially involved in the case until it came into conflict with his professional obligations. On January 12 he wrote an impassioned blog post about Swartz's death; this was picked up by the *Ars Technica* story and the *Volokh Conspiracy* posts in my study, and was linked to by comments on the *Times* and *Slashdot*.

The Volokh Conspiracy. The *Volokh Conspiracy* is an American legal blog, rather than a news site. Like Lessig's blog, the analysis there was cited by other news sites (I discovered it via a CBC story). I also found the discussion about civil disobedience and the legitimacy of Swartz's actions in the second article particularly detailed and interesting.

Most of the bloggers who contribute to the *Volokh Conspiracy* are law professors who describe themselves as "generally libertarian, conservative, centrist, or some mixture of these" (*Volokh Conspiracy*, n.d). Since 2014 the blog has been hosted by the *Washington Post*, but when Swartz died it was independent. The blog is extremely influential; it has even been cited in court rulings (Guess, 2008). Orin Kerr, one of the *Volokh* bloggers, wrote two articles about the Swartz case the week after his death.

The two *Volokh Conspiracy* articles I look at were posted a few days later than my other cases. By this time, the discussion had begun to change somewhat, with more focus on civil disobedience elsewhere. Comments are still discussing the same story, but

frames may have developed and shifted a little. This is something to keep in mind (similar frames may have shown up later elsewhere), but at my level of analysis I do not think it is a significant problem. For my purposes the most important criterion was articles that attracted surges of comments: on other sites this happened when Swartz's death become known, on the *Volokh Conspiracy* the surge was for this pair of articles.

The Discussions

The twelve cases in my study are discussions on these sites about the Swartz and Snowden stories. I did not choose every possible combination as a case. For example, I excluded the *Guardian* discussion about Swartz's death. Table 1 summarizes the site/story combinations in my study.

Table 1: *Cases by Story and Site*

Story	***Ars***	***CNN***	***Guardian***	***Lessig***	***Slashdot***	***Times***	***Volokh 1***	***Volokh 2***
Swartz	x	x	–	x	x	x	x	x
Snowden	x	x	x	–	x	x	–	–

Note: Each case is indicated by an "x."

The following table lists the cases about Swartz:

Table 2: *Article selection for Swartz story*

Site	**Date**	**Article**
Ars Technica	Jan 12	Internet pioneer and activist takes his own life
CNN	Jan 12	Internet prodigy, activist Aaron Swartz commits suicide
Lessing v.2	Jan 12	Prosecutor as bully
New York Times	Jan 12	Aaron Swartz, Internet Activist, Dies at 26
Slashdot	Jan 12	Aaron Swartz Commits Suicide
Volokh Conspiracy	Jan 14	The Criminal Charges Against Aaron Swartz (Part 1: The Law)
Volokh Conspiracy	Jan 16	The Criminal Charges Against Aaron Swartz (Part 2: Prosecutorial Discretion)

The Snowden cases are all from 9 June 2013:

Table 3: *Article Selection for the Snowden Story*

Site	Article
Ars Technica	Whistleblower who exposed NSA mass-surveillance revealed by The Guardian
CNN	NSA leaker comes forward, warns of agency's 'existential threat'
New York Times	Edward Snowden, Ex-C.I.A. Worker, Says He Disclosed U.S. Surveillance
The Guardian	Edward Snowden: the whistleblower behind the NSA surveillance revelations
Slashdot	WhistleBlower Outs Himself

I collected my data by viewing all of the comments in Firefox, using a browser debugging feature to save the complete discussion HTML, and parsing the result with scripts I wrote myself. See Appendix B for details.

Exchange Analysis

I contend that reader comments should not be treated equally. Some contribute more than others to discussion, some are more prominent on the page and some reflect the views of discussion participants better than others. As a practical matter, they are too numerous for a qualitative researcher to give them equal attention. An adequate analysis must discriminate.

Based on personal experience, preliminary study and existing research, I argue that ratings are a key criterion for choosing comments to analyze. Random sampling provides an unrepresentative view: it fails to account for comment inequality (a problem worsened by the long-tail distribution: comments with high ratings would be likely to escape random samples), and it treats comments as isolated fragments of discourse when in fact they are embedded in threads of dialog.

Top-rated comments often cluster together in threads, in part because ratings attract attention, in part because comments in such threads share subject matter that is of interest to raters, and in part because other factors can increase the chance a comment will attract ratings (for example, early comments tend to be rated higher).

While I conduct a few random samples, my main method takes into account both ratings and reply patterns in order to select and compare comments. I select the

comments with the most up-votes in each discussion, and analyze them in context: I compare them with other comments to which they respond (and sometimes to the parents of those comments and so forth) in order to see the role that they play in discussion. A top-rated comment thus draws my attention to the thread that contains it, and to other comments in that thread.

I focus on up-votes. I cannot rely on down-votes. Only some sites feature them; more importantly, up- and down-votes are incommensurable. Readers chose to vote up more often than down: there are more up-votes than down in every discussion I examined. This may be due to site norms (e.g. *Slashdot* encourages moderators to vote up rather than down), a log-in requirement (some sites allow anyone to vote up, but only logged-in readers to vote down), a desire to promote favored comments, or a preference by raters to praise rather than criticize.

Where possible, I compare comments in *exchanges*: pairs of comments, one responding to the other, that express contrasting views about a topic. If these are close together in time and space (and raters have a roughly equal opportunity to see both) their ratings can provide an indication of relative support among raters. Since the earlier post in an exchange has a greater opportunity to attract ratings, this comparison is most reliable if it is the later comment (the reply) that attracts more votes. For example, if comment A says “X” and attracts 20 up-votes, to which B replies “not X” and attracts 50 up-votes, it is reasonable to infer that “not X” has more adherents. Obviously exchanges with higher aggregate votes are more reliable.

An exchange need not be reflected in the structure of a thread; in some discussions, for example, comments reply to others at the same level. The important features for comparison are divergent approaches to a topic that are close together in space (by which I mean few other comments visibly intervene between them) and time: for example, two contrasting replies to the same parent comment, or two top-level comments addressing the same aspect of a story. In this case, a higher vote total for the later of two comments is a more reliable indicator than the converse. This approach can be expected to provide weaker evidence than exchanges, but may be necessary for comment systems where replies attract fewer votes (e.g. those used by the *New York Times* or *CBC*, where most or all replies are hidden unless readers choose to view

them).

Compared to many other forms of expression, comments tend to be brief and narrowly focused. This makes them ideal for comparisons of this type. Nevertheless, there are nearly always other factors that might contribute to a high or a low rating, such as tone, adherence to norms, or other arguments or claims in a comment. For the comparison to be valid, the topic being compared should be a major focus of both comments. Making this determination often requires attention to the discussion as a whole and to the back-and-forth within the particular thread.

Although exchange analysis relies on quantitative data, it can seldom produce a reliable quantitative result. Given all the other factors involved, it is important not to read too much into particular numbers and to be wary of counting exchanges or otherwise treating them as numeric. An analysis of exchanges does not stand on its own. Not every important topic or discussion will contain convincing exchanges. A debated topic in an exchange is like a theory about popularity which can be bolstered or undermined by an analysis of ratings.

My primary goal in analyzing ratings is not to reduce the complexity of discussion to manageable numbers. I believe that it would be irresponsible to conduct a qualitative analysis of reader comments without taking into account the important and prolific activity of raters. This is unlike content analysis: in most instances I am not assigning phenomena to categories so that statistical techniques can apply. On the contrary, ratings *already exist* as elements of discussion legible to and produced by participants themselves. Nor are ratings simply indications of popularity: rating is an active practice of participation. Like comments, ratings judge. Unlike many popularity metrics, ratings can be representative: by voting a comment up, a rater may be saying “this comment speaks for me” (though *Slashdot* in particular discourages this); in this sense, a rating is like a comment and the act of rating may substitute for writing one.

Taking ratings into account, I do not simply assess the presence or absence of frames in the discussion as a whole. In any sufficiently long discussion nearly every frame is likely to be present. I look at the presence or absence of ratings among top comments only, and use rating comparisons to weigh the strength of competing frames in exchanges.

As with any criterion for data selection, the use of ratings entails bias, systematically excluding certain comments and practices (most obviously trolling, but also presumably certain political views, styles of communication, and so forth). In order to provide some insight into what is missed, and to provide evidence supporting my claim that ratings are significant and have a material impact on discussion, in several instances I provide data from random samples. I believe my analysis backs up my use of ratings.

Rating systems vary from site to site. Except on *Slashdot* (with its fixed ratings scale of -1 to +5), popular comments attract exponentially more ratings than those that are less popular; early comments benefit from this disproportionately. Selection by site moderators can also influence comment ratings, as can other site features like thread hiding that make some comments more visible than others.

In some instances, down-votes can be useful. While it is risky to infer representativeness or popularity when up-votes exceed down-votes, a larger number of down- than up-votes can be a meaningful indicator. The one partial exception among the sites in my study is *Ars Technica*, where both voting up and down requires logging-in, making the two more comparable than elsewhere. It might be meaningful to compare up:down ratios, but in most cases the down-vote numbers are so small that the reliability of the ratio is questionable.

Using ratings to select data anchors my analysis, while analyzing exchanges allows me to take ratings into consideration and give some account of the relative strength of competing frames. But the fragmented diversity of reader comment discussions does not reduce neatly into a single interpretation or narrative. The frames I develop for each story are present in discussions but not exhaustive of them. I intend my analysis to both justify the frames I identify and to place them in the context of the larger diversity of discussion. Moreover, these are only a handful of discussions selected from the universe of the Internet. My choices and my analysis are ultimately influenced my subjective experience. All I can do is tell a story (not *the* story) of these discussions, and propose that it is may help illuminate other discussions elsewhere. To this end, I believe this attempt to illustrate my data should must be dense and detailed. Flyvbjerg (2006) and Peattie (2001) caution that the details of case studies can be more useful and

interesting than the findings. In my view, my analysis is less valuable for its results than as a practice and narrative to help the researcher and reader navigate a complex phenomenon exceeding the scope of any particular study.

Ethics

The material I am studying is all public and was all written with the intent that it be public. While this exempts it from ethics review, I believe there is more to be said.

Comments written for public consumption can still reflect poorly on their authors. Time passes and people change. There is merit to the ideal of a freedom to be forgotten, recently established in the European Union as a (problematic) *right* to be forgotten. I do not wish my research to cause harm or distress even to those who say hurtful things in public. Study of online discourse would be essentially impossible without taking at least some risk; still, there is a certain selfishness to proceeding. It helps that most commenters have used pseudonyms. The least I can do is to be mindful that these are real people, flawed as we all are, and strive to treat them with respect.

I initially wrote several paragraphs to bear witness to the great sadness of those who knew Aaron Swartz and many who did not (including me). Would it be honest to apologize, yet proceed? My analysis is what it is; I hope it does not add to anyone's sorrow.

Techniques for evaluating reader comments could be useful to intelligence agencies, political operatives and public relations managers. Many of these are people I would not wish to help: yet they have the resources and are in the best position to apply anything I might have learned. This is a risk run by much academic research. Scholars are like the dissidents on the island in Huxley's *Brave New World*: serving domination even in their dissent. To abandon the search for useful knowledge seems an even greater surrender. This does not absolve the researcher, however. I would rather focus on what might help than hurt; I might rather remain silent on what has little humanistic use. I would not stop anyone from reading this; to those who do, I can only say that covert efforts to surveil reader comments in order to control people, or to "manage" them in pursuit of ulterior goals do not have my blessing.

I am not suggesting that it is wrong for journalists, bloggers, experts or corporate

parties to participate in reader comment discussions. I believe that strategic participation in reader comments can be legitimate: public discussion matters because of its relationship to the institutions that shape our polity and our lives. But such participation should be undertaken with honesty, sincerity and respect.

Chapter 6: Comment System Designs

Discussion systems vary dramatically from site to site, with different technical features, reader populations, practices and norms. Such differences influence comment and voting patterns, what participants write, and what kinds of analysis are possible. (For example, determining reply patterns is difficult on a site like *Ars Technica* that lacks hierarchical threading.)

Some sites display comments on the same page as the story itself, while on others the comment page is linked, while others display comments on the same page but do not load them until the user scrolls down or clicks a button. Placing comments on a separate page reduces their prominence, likely affecting the number of people who read or participate.

Not all sites support hierarchical threads; those that do often only permit one level of replies. Hierarchical threads promote fragmentation of discussion, as each reply can digress on a different topic. This is particularly evident in discussion on *CNN* and *Slashdot*. Hierarchy also limits the capacity to sort comments, as the relationship between a reply and its parent must be maintained regardless of sort criteria. Hierarchy allows replies to piggyback on the popularity of parent posts; often the visibility of a top comment will attract a long thread of replies.

A site with collapsed threads initially hides some or all hierarchical replies to a comment; the user must then click a button to view the hidden replies. This decreased visibility can result in much lower scores for replies than for top-level comments, as is evident in the *New York Times* discussions. A reply piggybacking on a high-scoring comment is unlikely to gain much visibility this way.

A few sites allow anonymous or guest commenters. *Slashdot* features a particularly high proportion of anonymous replies; in part perhaps because commenting anonymously is a way to get around the *Slashdot* restriction that one cannot both moderate and comment on a given discussion (anonymous comments start with a rating of +0 rather than +1, however).

All the sites in my study allow readers to vote comments up; some also allow down-votes. Separate aggregate up- and down-vote totals are displayed for all sites

except *Slashdot*, which shows only an overall score calculated by subtracting down-votes from up (*Ars* also shows a calculated score alongside up and down totals). Among commenters, down-votes seem to be much more contentious than up-votes, as I explain in later analysis.

Not all sites require users to log in to vote. Some permit up-votes without a log in, requiring it only for down-votes. Different log-in requirements for up- and down-votes underline the difficulty comparing these numbers. Potentially, down-votes on such sites can help to distinguish the views of logged-in users from others. In practice, however, down-vote totals are usually much lower than up-vote totals. Most sites allow participants to vote on an unlimited number of comments. *Slashdot* is an exception: logged-in users are allocated a small number of time-limited moderation points by an algorithm. Once those points have been expended the user can no longer vote until again awarded with points.

Moderator participation is a social practice rather than a design feature. On some sites I observed no moderator participation; on others, participation is rare; on others the story author takes part in discussion. In my experience, moderators and authors on amateur blogs are much more likely to participate in discussion than are their counterparts on commercial news sites.

There are other important differences, such as accessibility from mobile devices (some sites hide comments or limit features), the presence of comments in syndication feeds and so forth. The descriptions above are all based on the desktop Web interface.

The initial ordering and selection of comments varies. Most of these sites show oldest comments first, but the *Times* shows editor picks first. For most sites it is possible for users to change the sort order; typically, to show newest first or highest-rated first. Where the volume of comments is very high—in the hundreds or thousands—this makes a difference as to whether later comments are seen at all.

Early comments have more time in which to amass up-votes and benefit from compounding popularity. Later comments can remain unnoticed. Showing newest comments first by default addresses this, but limits the opportunity for any comments to

accumulate high ratings.[10]

The following table summarizes some key features of the systems I studied as they were when I collected my data. Though some sites have subsequently changed, I am not aware of any changes to the design of these comment systems between the time when the discussions took place and when I examined them.

Table 4: *Key Comment System Differences*

Site	Same page	Threads	Collapsed replies	Anon.	Voting	Log in to vote	Moderator participation
Ars Technica	no	flat	n/a	no	up-down	yes	story author
CNN	yes	full	no	yes	up-down	down	no
Guardian	yes	2-level	no	no	up	no	rare
Times	yes	2-level	yes	no	up	yes	rare
Lessig	yes	full	yes	yes	up-down	down	yes
Slashdot	yes	full	yes	yes	up-down	special	no
Volokh	yes	full	no	no	up-down	down	story author

Ars Technica

Of all of the systems in my study, that of *Ars* most closely resembles a forum, probably because the site featured forums long before it supported comments. Comments form a long list from oldest to newest, broken into pages of forty each. There is no hierarchical threading and no means for readers to filter or reorder comments. Just as in a forum, the reader is guided through a discussion that flows and changes over time. Discussion is less fragmented than it is with hierarchical systems. A voting system was added later as an experiment in 2012, following which 16% of readers reported using the comments more than before while 2.5% reported using them less (Fisher, 2013).

Only logged in users registered with *Ars* can comment or click up or down arrows to vote. Totals of up and down-votes are displayed. Also shown is an overall rating (up-minus down-votes): if it is negative enough, the comment is hidden. This only really happens for earlier comments: later comments farther down the thread seldom

10 For some time the *CBC* comment system did this; the result is smaller vote totals compared to their previous oldest-first default. Users can re-sort comments, though this must be done each time a discussion is viewed. Some users clearly do so: early comments still rise to the top, but maximum ratings are lower.

accumulate enough votes. Logged in users can click to view hidden comments, but for other users there is no way to see them, nor is there any indication that they even exist. Site moderators can also hide comments; a flag button is used to attract their attention.

Readers can filter comments visually, picking out and reading those with high ratings. The site provides other cues that can be used for visual filtering. Colorful indicators at the top of a post indicate whether it is “Controversial” (high up- *and* down-votes), by the “Story Author”, a “Reader Fav” (high rating), or an “Editor’s Pick.” While the discussion is on pages separate from the article itself, sometimes Reader Favs and Editor’s Picks are also shown on the same page as and underneath the article. Here, they can only be read: no vote totals are shown, and there is no hyperlink to the comment in context in the discussion where voting is possible. Presumably this is intended to prevent compounding votes for promoted posts.

Compared to other comment systems, but in common with many discussion forums, the site displays a lot of information about the author of each comment. Every comment displays the user name of its author, when its author registered as a user, the total number of posts he or she has written and a title reflecting past participation (from “Smack-Fu Master, in training” for users who have written fewer than 100 comments to “Ars Legatus Legionis” for ten-year veterans with at least 10,000 comments to their name). User profile pages display comment history and a private message button, although I have never seen a mention of its use. Optional additional information includes location, occupation, gender and email contact. Despite this emphasis on comment authors, in years of reading I have seen little evidence that commenters recognize one another by name or reputation.

Like many forums but unlike many comment systems, *Ars* supports a fairly rich range of comment markup. Comments can include hyperlinks, images and block quotes. (Of the other systems I looked at, only *Slashdot* supports hyperlinks and block quotes; the others support none of these features.) Images are sometimes used with pictures to make a point, as in many discussion forums. When writing a reply, the content of the previous comment is automatically entered in a block quote at the start of the new comment. The author of the reply can then edit or shorten the quote. These quotes can nest, in effect providing a record of the thread of discussion.

After a comment is posted, there is a period during which it can be edited. Comments display their original posting time and the time of the most recent edit. Many users write their edit as an addendum and explain the reason for it.

Commenters are advised to read the posting guidelines for participating in both comment and form discussions: “comments aren’t run by rules, *per se*. Rather, they are moderated by a tight-knit group of moderators who have volunteered to put themselves in the position of making difficult judgment calls” (Fisher, 2000). The guidelines detail nineteen rules, ranging from respect for moderators and other participants to descriptive subject lines. With the lack of filtering features, it is the task of these moderators to maintain the quality of discussion.

CNN

The CNN comment system is very lightly moderated. Readers must rely on filtering and ratings if they wish to focus on better comments. The site uses the Disqus comment service, featuring multi-level threads, up- and down-voting (but no overall scores), and sorting by “Best” (the default), newest, and oldest. The result is often long, fragmented hierarchical threads. Sorting by “best” is minimally effective because top comments are displayed with intervening reply threads.

In long discussions, readers must click a button to load comments in batches; long comments are also displayed truncated unless a button is clicked to load the remainder. It is very difficult to see all of a long discussion.

The site provides is no indication of comment practices or norms. Comments do not need to be approved by a human moderator before being displayed. There is a flag button with which other readers can notify site maintainers of inappropriate comments.

It is difficult to see all high-rated posts with “Best” ordering, as the sorted top-level comments are interspersed with reply threads including low-rated comments. This exposes those replies to additional views. Highly-rated posts tend to have highly-rated replies as well.

In addition to comment text, comments also display the user name of the person who wrote the comment. Disqus allows users to sign in with Disqus, Facebook, Twitter,

or Google. There is an avatar display, but many users choose not to customize it.

The Guardian

The *Guardian* uses JavaScript to display comments below the associated article in a two-level thread hierarchy. At the head of each thread is an indication of how many people commented in that thread (e.g. “7 people, 9 comments”). Each comment includes the user name of the commenter, the time of posting, an avatar image or a placeholder, and the date and time of posting. Users can share a comment on Twitter or Facebook or flag it for moderators. Commenters must be logged in with a *Guardian*, Facebook or Google account. Optional profile information includes location, interests, a web site URL and description. Comments themselves can contain block quotes, italics, boldface, and hyperlinks. Even non-logged in users can “recommend” a comment to increment its rating.

A “community standards” link at the top of the discussion leads to a guidelines document (Guardian, 2009a) explaining rules and expectations for participants. Threats, personal attacks and so forth are prohibited. A more detailed Frequently Asked Questions (FAQ) document explains standards and moderation practices in detail (Guardian, 2009b). Moderators can remove comments, suspend commenting privileges, or pre-moderate comments by problem users. There is no indication that moderators take part in discussion itself, though the paper encourages article authors to do so.

Readers can filter to view staff replies or “Guardian picks” (the FAQ explains that reader recommendations are a basis for selection), if any. By default, all comments are shown. Comments are paged; initially 50 conversations (threads) are displayed per page, but users can choose to view more. Readers can view oldest first (the default) or newest. By default, all replies in a thread are expanded; an option collapses them, in which case a button loads more.

Lessig v.2

Lessigs’ blog, hosted on Tumblr, uses JavaScript to load Disqus comments below each blog post. These are preceded by a sometimes lengthy list of Tumblr comments, though very few of these are substantial (most consist almost entirely of the names of

users who “liked,” reblogged or retweeted the post). These do not constitute a discussion and I am not considering them in my analysis.

Users must sign in to Disqus, Facebook, Twitter or Google to comment or to down-vote a comment, but anyone can vote a comment up. Comments are threaded in a multi-level hierarchy. Comments can be shown with newest, oldest, or “best” (highest rated) first. The system only shows a limited number of comments initially; readers must click a button to load more. Long comments have a similar button for reading their complete text.

Since I collected my data, Lessig has moved his blog from Tumblr to a new site. While the new page links back to Tumblr, the comments I collected have been lost. (Some new comments have appeared on the new site and on the original Tumblr post.) My analysis is based on the 686 original comments.

The Volokh Conspiracy

Like Lessig’s blog, the *Volokh Conspiracy* uses JavaScript to load Disqus comments below a blog post. Readers can log in using Disqus, Facebook, Twitter or Google to write comments or to vote comments down. No log-in is required for voting comments up. Comments are threaded in a multi-level hierarchy. They can be sorted by newest, oldest, or “best” (highest rated). The system only shows a limited number of comments initially; readers must click a button to load more. The text of long comments is similarly truncated with a button for loading the remainder.

The New York Times

Only a minority of *Times* articles permit comments. When comments are enabled, the paper’s home page shows a comment count and a link to discussion.

Comments can be filtered to show only “NYT PICKS” selected by site editors (the default), top-rated “READER PICKS,” or “ALL.” When all are shown they can be sorted oldest first (the default) or newest first. The site allows for one level of replies. “Picks” are shown without context; clicking through to the originating thread requires a slow page load, discouraging the use of this feature. Articles often attract large numbers of comments, which must be loaded by the reader in batches of 50.

Readers are prompted to “Share your thoughts.” Logged in users can choose to reply, recommend, or flag abuses. There is a thumbs-up symbol next to the recommend button, followed by the number of recommendations. A linked FAQ explains that comments are pre-moderated, and that the paper is “interested in articulate, well-informed remarks that are relevant to the article.” There is no guidance for rating comments.

Associated with each comment is a user name, avatar, location (city and state if the user provides them), and an “NYT PICKS” icon if the comment has been chosen by staff. The commenting FAQ requests users to user their real names; most appear to do so. Custom avatars are rare.

NYT picks typically have high ratings. Other comments tend to have much lower ratings. On other sites replies to high-rated comment tend to attract votes, sometimes more than in the original comment. On the *Times* site, it is not uncommon to see a comment with ratings in the hundreds that has attracted a dozen or so replies with single-digit ratings. The default display of “picks” gives no indication of replies (even when a reader choses to write a new one) and makes it tedious to retrieve context. Most readers likely never see replies to high-rated comments they view or rate. The result is an exceptionally strong winner-take-all dynamic for picks. Back-and-forth dialog among replies happens, but is relatively invisible.

Slashdot

A typical news story on *Slashdot* consists of a short summary of a linked article on another site, followed by comments. Stories typically attract a few hundred comments or more displayed as hierarchical threads. Users can customize the view, hiding or minimizing low-scoring comments. Back-and-forth replies predominate. Most summaries are written by editors or selected from suggestions submitted by readers, often including a quote from the linked article. Some are book or movie reviews or requests for advice written specifically for the site, but most are references to stories elsewhere. The latter are typically short, and often poorly-written, biased or inaccurate. The low quality of the blurbs and headlines is a common complaint of commenters. The summary frames discussion but lacks authority: discussion often takes the form of critique or correction.

Many readers do not follow links to or read original sources, giving rise to the expression "RTFA" (read the fucking article).

The real meat of a *Slashdot* story is not the summary or the linked article; it is the discussion. In this respect, *Slashdot* is more like a traditional discussion forum, where discussion, not an article, is primary. But while the primary articles for *Slashdot* stories are seldom original, it is the articles and the site editors who select them and who therefore set the agenda for discussion. Most discussion is about news events or stories. The lack of original content makes *Slashdot* less of a go-to site for news stories, with less likelihood of popularity surges. The site appears to have a relatively strong sense of community compared to the other news sites I studied. Despite this, *Slashdot* features a distinctive and complex commenting system, which has been the subject of existing research.

While anyone can post comments on *Slashdot*, even anonymous visitors, by default not all comments are shown. There is no abuse reporting system, nor is there a mechanism for removing offensive posts. Site moderation is performed collectively by the members using a unique and somewhat complex system.

Each *Slashdot* post has a score from -1 to +5. Readers set a threshold (usually 1 or 2) below which only the title of each post is displayed. Clicking on the post's title will expand the text of that post and others at the same level, and perhaps below it (the software tries to be somewhat smart, and is configurable, so the exact behavior may vary from user to user). Posts with high scores are prominent and easy to find, even among hundreds of others. There is a second user-configurable threshold below which not even post titles are displayed: instead, *Slashdot* provides a single entry, e.g. "3 hidden comments," which can be clicked on to display the low-scoring comments. Both thresholds can be set to -1, in which case all posts are displayed in full.

The site maintains, but never reveals, a karma scare for each member. By participating and writing highly-regarded posts and by moderating fairly (as determined by meta-moderation, described subsequently) a user can increase his karma (certain too-frequent behaviors may also lower it). Those with good karma gain a bonus to their comment scores.

When a member with good karma visits the site he may be randomly assigned "mod points" (the number has varied as the software is tweaked and is influenced by karma and other factors about the specific user; at different times over the years I have received 5, 12 and 15 points). These points expire after three days. Until they expire or are used up, the member can choose to expend one to increment or decrement a post's score by one point by selecting an adjective (e.g. "informative", "off-topic", "troll"); the scores of comments that have been moderated are often accompanied by one of these adjectives. Moderators are encouraged to base their moderation on the quality of comments rather than their personal views, and to prefer moderating up to moderating down. The site administrators also have unlimited mod points, which they claim they use mainly to mod down floods of posts by bad actors (Slashdot, n.d.).

Members with good karma can also choose to meta-moderate. The mechanism was not always present, and the specifics have changed over time. The generally idea is to have moderators second-guess one another so that the system can detect rogue moderators. The system is remarkably successful: according to meta-moderation statistics, 92-93% of moderations are judged as fair (Poor, 2005). (Note, however, that the system no longer behaves as Poor describes.) In general, *Slashdot* discussion participants have reported that they find comment scores are useful (C. A. Lampe, 2006, p. 66).

Slashdot developers have commented on some of the design choices. Given its audience, *Slashdot* is a prime target for technical mischief. In an intriguing contrast to the "gamification" trend, the system is deliberately designed to avoid score-like reward systems so that it will be unattractive to "game" (O'Brien, 2004). Karma, for example, is an adjective ("good" or "excellent"), not a number. Top scores are low. There is little reward for disruption, as bad posts tend to rapidly be voted down and become invisible.

I have been reading and participating in discussion on *Slashdot* almost since its inception in 1997 and am very familiar with it. My experience is that comments are directed at what others have said, not at individuals themselves. Discussion is aggressively rational: comments are often blunt to the point of rudeness, but are judged primarily on the perceived merits of their argument. For example, *ad hominem* attacks are criticized for being fallacious, not because they are unpleasant. I have found that it is

important to comment while a topic is still on *Slashdot*'s front page. Replying to a highly-rated comment gives the reply good visibility. Argument should be aggressive and direct, but never personal. The comment should say something that has not already been said. It should demonstrate expertise with sufficiently detailed explanation, and if possible support this with the use of quotations or links to external sources of evidence. It should avoid diverting from the argument at hand to more general ideological stances which might deter up-votes. Such an approach can regularly achieve high ratings even when making an unpopular argument (e.g. references to Marx). The greatest risks are a) posting late or far down in the thread where the comment is unlikely to be seen, and b) an early down-vote that hides the comment from subsequent raters.

Reading vs Rating

Several of these comment systems illustrate a tension between reading and rating. Ratings are intended to guide readers: but they also influence subsequent raters. How can "good" comments be promoted while allowing other comments a reasonable chance to be rated? In order to rate comments, one must read them (or at least look at them). If readers and raters were different groups, or if reading and rating were distinct practices, there would be no difficulty. But they are not: anything that guides readers also guides raters. A winner-takes-all system inherently privileges early comments over later ones and downplays the dialog of replies. On the other hand, showing newer comments first reduces ratings and limits their use as a means to differentiate comments.

The problem is perhaps most pronounced for the *Times*, where editor and reader picks are most prominent. They gain the lion's share of the votes, while the vast majority of replies languish in relative obscurity. The *Times*'s emphasis on staff picks potentially compensates for this weakness by using a separate group to influence visibility and voting. Yet as my analysis will show, staff picks appear to be strongly influenced by reader ratings.

CNN takes a contrasting approach: even when top comments are shown first, their full threads are included. Thus, the comment with the highest rating is easy to find, but it tends to attract a large volume of replies; subsequent top-rated comments are pushed far down the page. The same applies to Lessig's blog and the *Volokh*

Conspiracy.

Ars Technica separates reading and rating activities where picks are involved: they are shown on the article page with no means to rate them; the only way to vote for or against them is to seek them out in the discussion itself. This prevents their ratings from snowballing, but for readers it also cuts them off from their context in the discussion.

Finally, *Slashdot* eliminates run-away ratings by restricting them to the range of -1 to +5. Rating is treated as a distinct activity: at any given time, few readers possess mod points, those who do are advised to turn off filtering in order to detect abuses by other raters. Even when filtering for top-rated comments, context and dialog are maintained by collapsing low-rated comments: raters can expand them individually by clicking on them.

Heavily skewed ratings, like those in *Times* discussions, pose difficulties for analysis. Analyzing exchanges will not work because the ratings for replies are consistently lower. The best that can be done is to compare two comments, A and B, posted in that order; if B has a high rating which is also higher than A then it is probably the case that B is more popular than A—provided that the difference is not due to B being an NYT PICK while A is not. Absent such a clear case, rating comparisons are no more than suggestive.

The *Times*'s bias in favor of staff picks presented me with a dilemma. I could minimize the effect of staff picks on my analysis by excluding them from the sample, e.g. by taking a sample of the top 50 comments that were not picked. This might well provide a more representative view of what *Times* readers really think about the story. I have not done so. My study is of discourse produced collectively through interaction. For better or for worse, the editorial practices of the *Times*, like the technical features of other comment systems, are important contributors to how that discourse is produced. It is necessary to consider them when assessing reader comments as public discourse, not bracket out their effects.

Chapter 7: The Death of Aaron Swartz

> Our governments are corrupt to the core, and it's going to be either us or them in the long run, and the battle is being fought over the Internet. The Internet which allows for true free speech goes directly against every control method the governments of the world and their business partners have. The fight for the Internet is literally the fight for humanity.
>
> — commenter reacting to the death of Aaron Swartz (Lessig, 2013b)

Aaron Swartz, 26, took his own life on 11 January 2013 while under indictment by U.S. federal prosecutors who accused him of computer crimes. There was an outpouring of mourning, anger and debate across online news and discussion sites, including criticisms of the U.S. legal system, the legitimacy of the law, and the proper boundaries of civil disobedience.

Swartz was well known for a number of projects, including his work on one of the RSS standards[11], as co-founder of the popular Reddit discussion site, and for his role organizing the successful anti-SOPA/PIPA protest campaign in 2011 and 2012. I had previously been aware of the Swartz case because of widespread coverage online. I also realized that I had cited him in my M.A. thesis: at the age of 19, he had written a blog post in which he analyzed contributions to Wikipedia. He overturned the assumption that a few core users made most original contributions to Wikipedia, revealing that most new material was in fact written by one-time or infrequent contributors.

In 2008 Swartz was responsible for the release of court records from the PACER database. Although under U.S. law federal government documents are automatically in the public domain, access to court records requires payment for access to the government-run PACER database. Swartz legally accessed the database then began downloading it in its entirety. He accumulated nearly twenty million pages before PACER noticed and shut him out. Swartz then released the public domain records freely on the Internet. The FBI investigated but found no basis for laying charges.

In 2010 and 2011 Swartz downloaded thousands of scholarly articles from the

11 Contrary to some reports, Swartz did not invent RSS. His work was on the use of RDF metadata in RSS 1.0. Despite its name, RSS 1.0 was not the precursor to the popular RSS 2.0 and failed to gain much traction.

JSTOR database of academic articles. When JSTOR and MIT administrators noticed his high-volume access, they blocked his IP and MAC addresses. Swartz circumvented this by changing his addresses.[12] When found out, he was charged by Massachusetts state prosecutors. They later dropped the case, which was taken up federally. The Department of Justice indicted him on thirteen felony counts including wire fraud and computer fraud. The prosecutor threatened him with penalties exceeding four million dollars and 35 years in prison.

By 2013, the cost of Swartz's legal defense had depleted his windfall from Reddit. His lawyer tried to arrange for a plea with no prison time, but was rebuffed. On 9 January 2013 the prosecutor proposed a six month prison sentence if Swartz pled guilty to all thirteen counts, accepting a felony conviction. Swartz refused the deal. Two days later, Swartz, who had previously described bouts of intense depression on his blog, hanged himself in his New York apartment.

His death was widely mourned across diverse Web sites. Legal scholar Lawrence Lessig, writer and copyright activist Cory Doctorow, social media researcher danah boyd and economics blogger and political activist Matt Stoller recounted their relationships with him and expressed their sorrow at his death. Within days the story was being covered by mainstream news, including the *New York Times*, the *Economist*, CNN, the BBC and the CBC.

The proliferation of discussion about Swartz was so great that on 17 January Nate Anderson (2013) wrote a story on *Ars Technica* titled "Opening arguments in the court of public opinion after Aaron Swartz's death." Quoting from several well known blog and media sites, Anderson writes, "It's remarkable just how quickly one young geek's death has mobilized even national political columnists—who by this point must have seen just about everything—into an outrage that grew beyond Swartz and has quickly opened up a national conversation about justice and about how we seek it." More than six months afterward, Swartz's name continued to appear in stories and discussions about Internet politics, such as Edward Snowden's NSA leaks. Over a year later his

12 IP (Internet Protocol) addresses are usually allocated dynamically to users on a network, and can change over time. Network hardware includes a unique MAC (Media Access Control) address. These addresses are used by hardware and software to manage network connections. These addresses do not correspond to any legal identity; there are a number of technical reasons for using standard network software to change them.

name still pops up in reader comment discussions.

Anderson's article focuses on debates about the U.S. justice system, but online commentary went far beyond this. Suicide and depression, the legitimacy of the law, the nature of civil disobedience and the character of contemporary American capitalism all came up repeatedly in discussions, often in response to articles or blog posts that said little about such matters. It this popular response that interests me.

Before delving into individual discussions (cases) I provide some general statistics about the discussion cases and define the frames I developed through my analysis.

Swartz Case Statistics

Statistics about comments and ratings provide important context for the qualitative analysis that follows, and for the vote totals I will cite. The comparison across sites also provides useful insights about how readers participate differently on each. Table 5 provides general statistics for the Swartz discussions, while Table 6 gives indications of vote distribution.

Table 5: *Swartz Discussion Comments and Commenters*

Site	N	Named Users	Anonymous Comments	Mean Posts	Median Posts	Reply %
Ars Technica	265	133	–	2.0	1	–
CNN	2,977	942	321	2.8	1	69
Lessig	686	384	8	1.8	1	58
New York Times	444	314	–	1.4	1	48
Volokh 1	649	176	–	3.7	1	86
Volokh 2	793	161	–	4.9	1	88
Slashdot	589	170	288	1.7	1	86

Note: N is the number of comments for each article when data was captured. *CNN* reported eight comments that were subsequently deleted. The *Times* reported 2,182 comments, but due to a bug on the site I could not fetch the final 383. The mean and median numbers are per named user (anonymous comments are excluded from the calculation). The Reply % figure is based on the hierarchy of comments on the site; these figures may not be comparable across sites due to different threading systems

(*Ars* discussions are flat, while the *Times* and the *Guardian* allow only one level of replies.)

Table 6: *Swartz Discussion Votes*

Site	N	Top Rating	Most Votes	Most Up	Most Down	Mean Up	Median Up	Votes /N	Votes /User	Most Up/N
Ars	265	181	229	199	152	17.0	7	26.8	53.3	0.75
CNN	2,977	677	721	699	26	3.1	1	4.5	14.2	0.23
Lessig	686	797	817	807	62	8.9	2	10.0	1.8	1.18
Times	444	452	452	452	–	29.1	12	29.1	41.1	1.02
Volokh 1	649	34	41	35	15	3.1	2	3.9	14.4	0.06
Volokh 2	793	46	48	47	7	3.1	2	3.7	18.1	0.06
Slashdot	589	5	–	–	–	0.8	0	–	1.9	–

Note: Slashdot ratings are always in the range of -1 to +5. The table only provides the final score for each *Slashdot* post; mean and median up-votes are thus actually mean and median ratings. The total votes per user calculation is based on total votes (up plus down counts) per named user; for *Slashdot* this means something different as it based on overall ratings.

Swartz Frame Definitions

Frames exist at multiple levels. Despite the fragmentation and narrative incoherence of reader comment discussion, there must exist a frame within which the vast majority of participants operate. What do readers expect when they read or participate in such discussions? There may be a frame or frames embracing all reader comment discussions, and one particular to each story (a “story frame”). Frames are not arbitrary, but the choice of which ones to construct and apply is partly subjective, reducing a panoply of replies and discussion to a single dimension. The practice of seeking such a story frame, a minimum spanning frame for discussions about a story, is a method I use for navigating and attempting to make sense of discussion.

Possible frames for the Swartz story might be Witnessing (“too soon bro,” reads a comment on *CNN*), Understanding Depression, or Trolling. One of the most prevalent frames, however, is that of rational debate, governing expectations for what disagreements are about, what kinds of questions are relevant, and what kinds of answers are expected. This is overly broad, however, as a device for analyzing the

specific claims made in the Swartz discussions. Blame is a driving force in most disagreements, and is assigned in a large proportion of comments. The story frame I propose for discussions about Swartz's death is a question that many or most comments address: Who is to blame?

The implied *for what* (his death? law-breaking? bad law?) is entailed in the answer to the question. The answers (the prosecutor, the government, society, capitalism, himself, his friends) point to different political frameworks for organizing an interpretation of the story.

Below are categories of answers to the question, Who is to blame. These are frameworks of understanding used by article authors and discussion participants, who often argue or assert that a particular frame is the correct or true one for understanding the story. Each is broad enough that it could easily be generalized and applied to other stories. Following Gamson, who codes fine-grained frames like these, I refer to these as frames, even though none of them successfully sets expectations for discussion as a whole.

Troubled Genius. Swartz was young, brilliant and accomplished, yet that genius had a dark side. Something in his character drove both his success and his tragedy, leading him to make questionable choices that resulted in run-ins with the law. His temperament (evidenced by a history of depression) appears to have contributed to his untimely death. This frame particularizes Swartz's story. Its essence is that he was *other*, and both his accomplishments and death are correlates of that otherness. A weak expression of this frame is present in *Slashdot* comments that quote *Bladerunner*: "The light that burns twice as bright, burns half as long." The frame does not qualify if blame is placed with others instead of with him (the idea that society punishes genius, for example, is an instance of Institutional Failure).

Law Breaker. Swartz broke the law and is therefore responsible for the consequences. The prosecution and the legal system responded appropriately. Swartz's suicide is tragic, but he was entirely responsible for making that choice. Any flaw in the law is a secondary issue that should not impede its impartial application. Comments espousing this frame typically characterize Swartz's actions as selfish, unethical or irresponsible: certainly not instances of legitimate civil disobedience. His attempt to

avoid punishment only confirms that his action was not legitimately political. The argument is often buttressed by citing actions of his that are not in themselves illegal, but that reveal what kind of character he has: a law breaker is someone who does not respect or who has contempt the law.

Bad Actor. One or more individuals within the government or other organization with authority acted irresponsibly. Accusations are almost always levied at prosecutors Ortiz and Heyman. Though bad actors may be present throughout the system, this frame does not extend the argument to a critique of the system as a whole. Only individual responsibility is cited. The presence of a bad actor does not exonerate the accused; this frame is therefore compatible with the Law Breaker frame.

Institutional Failure. Institutions suffering from systemic problems contributed to the failings in this case. The purpose of these institutions is sound, but in this instance they failed to achieve it: the prosecution was unwarranted or over-zealous, MIT failed to act with appropriate restraint, the laws on copyright or computer crime are flawed, etc. The failure likely resulted in an injustice, but a technical one: while criticism of the "system," "machine," etc. can be very harsh, comments applying this frame do not identify a structural antagonism and power inequality between different groups. The solution is to fix the institutions so that they work properly. Unlike with the Bad Actor frame, there is an implication that this case reflects a pattern larger than this particular case.

Unequal Justice. The law is applied unequally to different groups. Swartz, like others, was pursued while more elites guilty of worse crimes go free. His case is often compared to the treatment of those responsible for the financial crisis. Swartz may still be guilty, but that is less important than the demonstration of systemic injustice. The legal system is not simply flawed or arbitrary: it is systematically biased to the benefit of some and the detriment of others. This is above all an antagonistic frame that identifies injustice, an us, and a them. If there is no privileged "them" (however ambiguously identified) then this frame does not apply.

Democratic Struggle. This frame rejects a legal or individualized interpretation of Swartz's actions for a public and political one. Swartz was more guilty of a right than a wrong. He was engaged in a legitimate political struggle with aims as narrow as open

access to scholarly research or as broad as democratic emancipation and social justice. Swartz's claim to civil disobedience is legitimate and consistent with other struggles for justice. The civil rights movement is often invoked; sometimes also Gandhi, WikiLeaks, or Bradley Manning (who leaked diplomatic cables to WikiLeaks). This frame is close to agency; exponents may call for carrying on Swartz's work. This frame need not set up an antagonism, as the struggle may be against broken systems (Institutional Failure) rather than privileged elites (Unequal Justice).

Other frames. Other frames that occur in the sample include Tragedy, which rather than engaging in debate expresses sadness and loss; Conspiracy, which claims the official story of Swartz's death is a cover story for some kind of skulduggery; and Pawn, according to which supposed friends or allies of Swartz (usually Lessig) used and sacrificed him as a cat's paw in their ideological struggles. While this last frame is rare and invariably low-rated, its inversion of Democratic Struggle and Bad Actor is intriguing. I imagine there might be a reservoir of this frame somewhere online.

Ratings Notation

Reader ratings are an important part of my analysis. I generally include vote totals in parentheses when quoting discussion comments, e.g. (+5) or (+10/-2) for a site with both up- and down-vote totals. I always include the sign when referring to totals: plus for up-votes, minus for down (-5 down-votes means five votes down).

As explained previously, I focus on up-votes rather than down-votes. When I refer to high-rated or top-rated comments, I mean those with the most up-votes, regardless of the number of down-votes on those comments. If I am discussing down-votes or overall ratings (up minus down) I say so explicitly.

If I write that post A says "X" (+50/-4), while post B replies with "not X" (+90/-10), the reader should understand an implicit comparison between the up-vote totals and their associated claims: in this case, "not X" attracted substantial and greater support from raters and is likely more representative of the opinion of discussion participants. Such comparisons are frequent in the text; to avoid clutter, I seldom elaborate on them.

I looked at a purposive sample of the 50 comments with the most up-votes in each discussion. In some cases the count is greater than 50 due to ties. For *Slashdot* I

simply chose comments with scores of +4 or +5. These are not the only comments I looked at, however; I found other comments following threads, by accident reading a discussion, or as part of a random sample. I do not always mention whether a given comment is one of the top 50 sample; the reader can determine by comparing the number of up-votes with the top-50 cut off given in the introduction to each discussion.

Lessig v.2

Lawrence Lessig's (2013a) January 12 post, "Prosecutor as Bully," is a blog post, not a news article. Lessig is an expert communicator, but he is a scholar and advocate, not a journalist. He is not writing news. The inverted pyramid of the news story is entirely absent. Even though Swartz died only the day before, Lessig takes for granted that readers know the story. He writes about "Aaron" as someone his audience will know and recognize by first name or from the photograph at the head of the story; not until the second last paragraph does he even use Swartz's last name.

Yet though it is different from the other articles I look at, this is also an important and prominent article. The *Volokh Conspiracy* and *Ars Technica* articles link to it. The other news articles in my study do not: but comments in every discussion do, with the exception of *CNN*. Where other articles are *about* the story, to some extent the Lessig article is *part of* the story. This is why I describe it first: the ideas from here and links to here show up in other cases I consider.

Over the course of the article, Lessig constructs a rational and emotional argument for this particular story as a manifestation of a larger Institutional Failure. The essence of the article is a heartfelt cry of pain and an accusation leveled at the prosecution of this case and (as a Lessig makes clear in a post a week later) at American justice more generally. "Please don't pathologize this story," writes Lessig. "If we're going to learn from this, we can't let slide what brought him here." Alluding to Democratic Struggle, Lessig writes, "The causes that Aaron fought for are my causes too." But "If what the government alleged was true . . . then what he did was wrong." (Lessig has affirmed repeatedly that "piracy is wrong" (Lessig, 2008, p. 113)). Regardless, treating Swartz like a terrorist and threatening him with fifty years in jail over academic papers is an outrageous instance of Unequal Justice "in a world where the

architects of the financial crisis regularly dine at the White House."

Despite Lessig's argument, in my opinion the passion with which he recites key details of the case makes it difficult to keep in mind the larger context. The article is a unity that must be read from beginning to end to complete its argument. It concludes by blending personal tragedy and social advocacy in a powerful appeal for solidarity around shared values of justice:

> Fifty years in jail, charges our government. Somehow, we need to get beyond the "I'm right so I'm right to nuke you" ethics that dominates our time. That begins with one word: Shame. One word, and endless tears.

The discussion. Lessig's article attracted a surge of 686 comments: an order of magnitude more than his subsequent blog post on the topic a week later (Lessig, 2013b), which received only 47. I examined the 51 comments with the most votes up (51 rather than 50 due to a tie). The comment with the most up-votes received +807; the one with the most down-votes received -12. The lowest number of up-votes among the top 51 was +26. Aside from those in an extended digression about Julian Assange, all of the top 51 comments on the article are in general agreement with Lessig. None criticize Swartz; none absolve the government. For comparison, only two comments in a random sample of 50 are critical of Swartz.

Lessig's article addresses a public who "pathologize" Swartz as in some sense "crazy" (the Troubled Genius frame is an example of this). He aims to redirect concern from the particularities of this case to more general problems with an unjust system. Top comments embrace Lessig's argument, making Swartz's story an example among others. IRMO says that "this looks to be yet another case where the DOJ prosecutor cast aside the public interest" (+65/-0). "Remember Ida Craddock . . . hounded to her death at the beginning of the last century for supplying birth control information to women," writes impboy (+52/-0). The highest rating is for a comment by andreasma who draws more comparisons:

> Aaron, Manning, Assange, Kyriakoy, Occupy, all persecuted, hounded, some tortured. For what? For speaking truth to power, for revealing corruption, war crimes. For liberating information. . . . Meanwhile, Yoo, Addington, Libby, Cheney, Rumsfeld, Gonzales . . . The murderers, torturers and torture apologists are celebrated. . . . the rule of law has

> become a joke . . . the poorest get relentless, unforgiving, zero tolerance prosecution for the tiniest of misdeamenors. . . . There are only two crimes that are punished in this country now: being poor or challenging the powerful. (+807/-10)

This comment proves to be controversial, however, for the comparison to Julian Assange. It kicks off a long thread containing most (30) of the top 51 comments. The first reply, by Angie Stich, says that Julian Assange "wouldn't be in the position he's in if he weren't a rapist" (+68/-62, a large number of down-votes). Her claim attracts substantial support and opposition, launching a prolific subthread with many top comments. The first reply to her disagrees, gaining a higher rating. Ophelia Millais writes that she "wasn't aware that he had already stood trial for that and was found guilty" (+208/-2).

In the discussion that follows, some comments imply skepticism about justice in the US, UK (where Assange too refuge in the Ecuadorian embassy), or in Sweden (which ordered his extradition on rape charges). "Wasn't the point of the Assange 'so-called RAPE' prosecution to sully his powerfully good image? And of course to deter 1) him; 2) those who might whistleblow in his direction; and most importantly, 3) all possible imitators," writes Villainess (+64/-3). "Assange . . . knows that they want him delivered to the US via Sweden, and from the ghastly, barbaric manner the US treats its prisoners these days, he is doing what he must to stay out of their hands," writes w00t (+140/-11). The comment by w00t actually ends, "shut the fuck up," leading to a bitter back-and-forth that includes charges of antisemitism against w00t when he criticizes another commenter for using "the Jew card."

Top comments like andreasma's embrace the radicalism in Lessig's argument, linking Unequal Justice to claims that Swartz was engaged in Democratic Struggle. In addition to several in the digression about Assange, other comments in the thread agree with the assessment of American justice. Omnivore1 thanks andreasma: "I shared your entire comment on my FB page. . . . You said what I know to be true, much better than I could have" (+144/-1). "The US criminal justice system is a business in which employees get promoted for high arrest and conviction rates," writes Manuel Laboriante (+61/-1). When Prokofy disagrees, Ophelia Millais rebuts her with a long reply. "There are many who feel, as Swartz did, that some things morally and ethically belong in the public domain," she writes, "To liberate this type of content isn't just a form of protest, it's a

public service" (+32/0).

The Unequal Justice and Democratic Struggle frames appear elsewhere. Paul Shuster, for example, argues that Swartz was engaged in civil disobedience, but "Whether you get prosecuted for white collar crimes is based solely on whether you have political juice or not, not the magnitude of the crimes. The current composition of our two party system ensures that this will be the case for the foreseeable future" (+34/-0).

In summary, commenters embraced the Unequal Justice frame Lessig presents. I approach the comments on the articles in my study as though they were written in isolation at a point on time; in fact, a number of them link to other sources, including to this passionate post by Lessig. As a widely-referenced early reaction to Swartz's death by a public figure who knew him, it is likely Lessig's framing of the story (particularly his use of Unequal Justice) shaped how it was framed elsewhere.

The New York Times

The *Times* article, "Aaron Swartz, Internet Activist, Dies at 26," tells a story of Troubled Genius: a young man who accomplished much but succumbed to depression (Schwartz, 2013). It outlines Swartz's death, some of his accomplishments and reputation, the charges against him, and his depression. The center of the article is devoted to his activism, including his PACER and SOPA efforts.

The JSTOR case is introduced by the federal indictment. "In an effort to provide free public access to JSTOR, he broke into computer networks at M.I.T." A quote from prosecutor Carmen Ortiz invokes the Lawbreaker frame: "stealing is stealing." The scholarly publications belonged to JSTOR, which needs subscription money. Swartz's possible motives are explained by Carl Malamud who, while he "did not approve of Mr. Swartz's actions at M.I.T." hints at Democratic Struggle with his belief that "access to knowledge and access to justice have become all about access to money, and Aaron tried to change that." Describing the PACER documents, the paper reports that "activists like Carl Malamud . . . have long argued that such documents should be free because they are produced at public expense."

The article ends with an extended excerpt from a 2007 blog post by Swartz in which he describes the intensity of his depression: "you feel as if streaks of pain are

running through your head, you thrash your body, you search for some escape but find none."

An addendum states that details about his arrest have been corrected. There is no correction to the inaccurate claim by "federal officials" that Swartz used a "false account" to sign in to the MIT network. The article presents an objective recounting of facts and quotes from sources. What emerges from it is a picture of individual Troubled Genius: a nebulously idealistic, bright and accomplished but troubled young man who was driven over the edge by depression and his run-in with the law.

The discussion. There are 444 comments, the second most comments for any article that day (see Appendix A). Of these, 10 are staff picks. Users needed to be logged-in to vote. The highest number of up-votes was +452; among the top 50, the lowest number of votes is +61. The site does not support down-votes.

While the *Times* reported this as a case of Troubled Genius, the comments are striking in their avoidance of it. Among top 50 comments, only Soleil presents a weak version: "Sounds like a severe prosecution and a sensitive, brilliant young man" (+74). Instead, comments are almost all concerned with the rights and wrongs of the case. All of the 50 top-rated comments support Swartz or criticize the government. A few say that what he did was wrong (focusing on actions rather than character) but emphasize the greater wrong done to him. DG writes, "It is completely ridiculous that this was treated as a criminal offense . . . What he did was wrong, but, at most, he should have been required to have counseling" (+61).

The *Times* hides replies by default; they therefore seldom receive many votes, and are scarce among top comments. Nonetheless, there are 5 replies among the top 50. Of these, four are responses to other commenters who criticize the idea that information or journal articles should be freely available. "I don't have much empathy for his view that subscription information should be 'liberated', basically stolen and distributed for free," writes JS (+18). That is not a top 50 comment: but two replies are. User rachel replies with Democratic Struggle: "the data he 'liberated' was already paid for by the public, and in essence privateered by corporate interests" (+116); eb writes, "He fought to keep people from having to pay for access to information they already owned - it was not entitlement, but a clear-cut sense of right that drove him" (+162, NYT

pick). As an NYT pick this comment is given greater prominence, so its rating is not comparable.

The pattern recurs later: Sergio Georgini writes, “I find it interesting that so many people think creators of intellectual property should have to share their work with the public for free” (+22). The reference to “so many people” implies that this is effectively a reply to other commenters. To this adam replies, “Again, a profound misunderstanding of the difference between intellectual property created for profit and royalties, and academic intellectual property most often paid for by the public” (+94, NYT pick).

Michael Lissack criticizes “this ‘crime without victims’ argument,” arguing that Swartz “trespassed” and “abused two networks massively. Nothing gave ”him the right to make his own rules and to decide what was right/wrong without consequence” (+31). ATCleary emphasizes Unequal Justice over Lawbreaker: “The issue is not whether this was a victimless crime. Rather, the issue is the proportion of the response . . . Especially when it comes hard on the heels of the government’s refusal to pursue criminal prosecution of HSBC’s employees” for laundering drug money (+83).[13]

The final top-50 reply agrees with its antecedent. Lola writes, “you can rape a woman and only get 2 years” (+130), to which e.s. responds, “Apparently you can also loot billions of dollars and not only avoid all prosecution, but get millions in bonuses” (+69). (The mention of rape may be a reference to the recent Steubenville case, in which the sexual assault of a high school girl by multiple assailants was captured on cell phone cameras. Comments in several discussions make comparisons like this.)

The issue of access to scholarly work is repeated in many of the top comments, including the highest rated of all, by Valerie:

> . . . as a researcher, I can vouch for the fact that Mr. Swartz was right. The results of scientific research are in the public interest and are generally taxpayer funded. . . . the system is broken for everyone except the publishers. Online acces to research should be FREE. RIP Mr. Swartz, and thank you for trying. (+452)

13 HSBC had been laundering billions of dollars for drug cartels. A month earlier, the Department of Justice reached a settlement under which the bank avoided prosecution by paying $1.92 billion and agreeing to certain conditions (Protess & Silver-Greenberg, 2012). Many commentators criticized the settlement as a slap on the wrist, contrasting it with harsh American drug enforcement (Taibbi, 2012).

Again and again, the determined prosecution of Swartz is characterized as Unequal Justice in comparison with other cases. In a couple of instances the contrast is with rape convictions, but in most it is with the lenient treatment of those responsible for the financial crisis. The second highest-rated comment makes this argument, as does bikegeezer with the third highest-rated comment,

> The guys who blew up the world economy with credit default swaps and toxic mortgages labeled triple A. Nobody went to jail. The banks that forged tens of thousands of documents that threw people out of their homes. Nobody went to jail. HSBC which laundered money for drug cartels and terrorist organizations. Nobody went to jail. And this guy was indicted? Eric Holder and the "Justice Department" is a corrupt organization. (+360)

None of the three replies disagree. "And the entire corrupt system breeds contempt for the law," replies Frank Knarf (+22). I remember America rebuts Ortiz: "'Stealing is stealing,' except when you're a bank" (+186). TonyR agrees: "Worse still, it's not even stealing. Legally, stealing requires the original article to disappear" (+5).

While the *Times* article mentions Democratic Struggle, it focuses on personal tragedy. In contrast, comments focus on Unequal Justice and the Democratic Struggle for open access. They reframe the story: but not explicitly. There is no rejection of the *Times* article as inadequate or poorly framed. The discussion almost seems to be talking about a different article; it could well have been a response to an article about Swartz's ideals and the failure to prosecute of bankers.

Presumably commenters are reading other sources. The article does not mention Lessig (though some lower-rated comments do), yet DG writes without explanation, "Lessig is correct. There should be judicial review" (+61). User lwlegel quotes further information about the charges against Swartz from *techdirt*, a blog that focuses on critiquing intellectual property law (+109).

One comment, by frank, is identical to a comment posted by andreasma on Lessig's blog, where it is the top comment (here it rates 10th with a rating of +205). The Lessig posting has an earlier timestamp; this suggests that arguments from Lessig's blog are migrating to the *Times*.[14] It is possible that this could in part account for the

14 Timezone confusion could cause difficulty. I checked the timezone behavior of the sites I looked at: both the *Times* and Lessig's blog report the reader's timezone, which is to say mine

spread of the Unequal Justice frame, especially comparisons with the treatment of financial wrongdoers.

Staff picks present a further complication. The ten picks in this discussion include condolences, a discussion of the economics of academic publishing, the open access movement, a contrast with the treatment of bankers, and an argument that this does not constitute genuine civil disobedience because Swartz did not accept the consequences of his actions. While the *Times* article itself follows journalistic norms of objectivity, staff picks stake out positions. I will deal with this issue in more depth later.

CNN

Like the *Times*, *CNN* tells a story of Troubled Genius. Swartz was "an Internet savant who at a young age shaped the online era" (Martinez, 2013). The article catalogs his accomplishments. "A prodigy, Swartz was behind some of the Internet's defining moments . . . At the same time, he was plagued by legal problems arising from his aggressive activism, and he was also known to suffer depression."

The article's description of the PACER incident is ambiguous: according to the FBI, Swartz downloaded "18 million pages with a value of about $1.5 million," but "no charges were filed." After reading his FBI file, he remarked that "it's truly delightful." The article provides no justification for the "alleged hack," leaving the impression that Swartz was a Lawbreaker who took the law lightly but was fortunate to be let off the hook.

Similarly, Swartz's possible motivations in the MIT case are not mentioned:

> In 2011, he was arrested in Boston for alleged computer fraud and illegally obtaining documents from protected computers. He was later indicted in an incident in which he allegedly stole millions of online documents from the Massachusetts Institute of Technology. (Martinez, 2013)

Swartz's family defends him, arguing that this was "the product of a criminal justice system rife with intimidation and prosecutorial overreach" for "an alleged crime that had no victims." Their claim about his "commitment to social justice" is left hanging with no justification.

---

when I fetched the comments, so the timing should be correct.

The article wraps up with over 180 words of excerpts from his blog in which he "wrote about death" and described depression, ending with the same passage that concludes the *Times* article: "You feel as if streaks of pain are running through your head, you thrash your body, you search for some escape but find none. And this is one of the more moderate forms."

The article's Troubled Genius frame presents Swartz as a bright but troubled young man with ambiguous political ideals, who did not take the law seriously and got away with one theft, was caught when he tried again, and ended his long-standing struggle with depression with suicide. He is unique, as is his tragedy. This is perhaps the pathologization that Lessig fears.

The discussion. The *CNN* article received 2,977 comments. Six *CNN* articles that day received more (one had 16,759). The comment with the most votes up received +699. While anonymous readers could vote up, only logged-in users could vote down. The comment with the most votes down received -26. Of the 52 top-rated comments, the one with the fewest up-votes received +29.

With nearly 3,000 comments and a very broad audience, I expected *CNN* to have poor quality and uninformed discussion. I ultimately concluded that *CNN* comments had important things to say. Before I explain, I illustrate why the comments on *CNN* made a negative impression on me. Compared to the other discussions I looked at, *CNN* is an extreme case: it demonstrates the merit of comments even when discussion is not at its best.

Of all the discussions I looked at, *CNN* probably featured the highest proportion of off-topic, perfunctory, and unempathetic comments. There are threads of partisan name-calling, religious preaching, and irrelevancies. One of the first high-scoring comments jokes about Swartz's death: "RSSt in peace?" asks ßen Murphy (+142/-23).

A random sample of 50 comments illustrates the lack of empathy or seriousness in many of them. There are two unexplained quotes from Hamlet by GoRemoteKCI (one of the discussion's most prolific posters with 55 comments). "We must enact legislation to ban ropes, NOW!", writes nomoreropes, presumably referring to the gun control debate (+4/-0). Dizork Mage writes several jokes, e.g., "Kid commits suicide with a drone

attack" (+0/-0). An anonymous user writes, "someone wasn't taking their prozac" (+1/-0). "Reading CNN comments might have pushed him over the edge," writes MedianD (+7/-2). "Aaron who?" asks a post by JohnkinsBob (+2/-2). Conspiracy theories pop up. "He died from 'suicide' just like that guy who won the lottery died from 'natural' causes," writes Bull31 without explanation (+0/-4). Swartz's sexuality is noted by nycmcmike: "I heard even though he occasionally hooked up with guys he didn't identify as gay" (a blog post by Swartz does say something along those lines) (+2/-0). LeftyGrove writes in a top-level reply, "nearly every super-intelligent person is liberal" (+4/-3); there are several threads of partisan political back-and-forth buried in low-rated threads.

Some in the random sample are serious. "So far only a subset of the JSTOR articles are free. I tested it out today," reports Kenneth Kleefeld (+0/-0). "Whatever happened to 'if you cannot do the time, don't do the crime'?" asks Jason McKay (+3/-0). "It's sad that he lacked the maturity or guidance of mentors to pull back," writes NoodleDogg (+5/-3). Priyath Ruchike Foneska states, "Taking your own life is cowardice" (+8/-13). User jacalder suggests Swartz's suicide is evidence that "perhaps we really are raising an entire nation of narcisists" (+4/-2). Gerrie Warner appears upset by unfeeling remarks: "Many of the comments conserning this young person's death make me embarrased to be a human being" (+4/-0). John Smith says some of his friends knew Swartz; he describes on the effectiveness of various treatments for depression (+1/-0).

Compared to the other sites I studied, I put extensive effort into trying different ways to filter comments on *CNN*. I thought the discussion as a whole would reflect my first impression; nor did I expect reader ratings to make much difference. In the end, I examined the top 50 comments in detail and came away with a very different impression. Unlike the random sample, top-rated comments nearly all stay on topic, responding to the article's emphasis and failings.

The top-rated comment, by BinaryTruth, tries to fill some of gaps in the *CNN* article, though it erroneously substitutes MIT for JSTOR:

> charged with 13 felonies in September (up to 50 years jail time if convicted) for copying publications from MIT, then a couple days ago MIT voluntarily released over 4 million of those same articles to the public for free. Facing that much punishment for a "crime" that MIT had just rendered moot, it's no wonder he killed himself. (+699/-22)

An anonymous reply emphasizes that this is missing from the article: “funny the article doesn’t mention that he was facing years in prison” (+214/-3). Two other top 50 comments follow up with criticism of *CNN*.

Replying to BinaryTruth, WeRTM criticizes Swartz as “a young man who had a great deal of opportunity in life” but “made many impulsive and questionable decisions” that “caught up with him” (+59/-18), reflecting the Troubled Genius frame presented by *CNN*. A subthread follows in which a number of top 50 comments debate Swartz’s actions. When Bill invokes the Lawbreaker frame (+47/-18), Fatemeh Khatibloo responds, “when your laws cannot keep up with the pace of technology… we need people who are willing to break them” (+210/-9). Institutional Failure appears in several other comments, as when Blake asks “whether or not his actions SHOULD be illegal” (+41/-1). Although Unequal Justice does not occur in top comments, handleym makes a claim for Democratic Struggle:

> So you think Ghandi should never have broken the law? Likewise for Rosa Parks? Likewise for the Stonewall demonstrators? Likewise for Mohamed Bouazizi? In the face of evil laws, resistance is the moral response. . . . We are talking about work that was done by public universities, all paid for by US taxpayers . . . but which is kept locked up by parasites like Elsevier which contribute nothing to the actual creation of this knowledge. (+88/-2)

Other comments blame Swartz; responses that defend him are almost invariably more highly rated.

Several threads center on depression and suicide. Matthew Lawliss replies to WebRTM: “Suicide alone is more complex than any comment here could describe. You do this human being disservice by simply saying that the entirety of his problems were impulsive behaviors” (+159/-4). John E. Vargo disagrees: “Suicide is selfish; there’s no defending it” (+24/-19). Two highly-rated comments rebut. An anonymous commenter writes, “Do you deem yourself able to judge all other human beings?” (+68/-1), while blackroseMD1 writes of working with homeless youth: “I tend to think it’s more selfish of the people around the person, as they want the person to continue suffering so that they can continue to enjoy their presence” (+59/-1). Robert Mccall writes, “Committing suicide is inherently dumb” (+17/-15), to which lte-21 replies, “obviously you haven’t been depressed” (+47/-2). Elsewhere, Idean50 writes of finding “my father 3 months ago with

the gun still in his hand" (+51/-2).

The Troubled Genius frame of the article presents Swartz's situation as though it were unique. But while top comments take up the article's focus on Swartz's depression, they reject the particularization of his case. Personal experiences and expressions of empathy shatter the distance from him as a particular individual, revealing him to be one sufferer among many.

While empathy is one response to suicide, other top comments suggest that Swartz was assassinated. Blogger Blogger catalogs several deaths that he calls "'Suicide' by hanging" (+181/-3). Seven comments in agreement in the following thread are among the top 50. Some are virtiolic: "from the way you talk, I'll be glad when you're not breathing my air," writes Christina Davis (+31/-2) in response to Indigo Wizard's statement of "Good riddance" to Swartz (+14/-11). Several comments defend Swartz against accusations of cowardice or selfishness for committing suicide. User srichey321 argues that "Associating cowardice with mental illness is what causes people not to get help" (+29/-1).

The *CNN* discussion was among the first that I analyzed in detail. I experimented with several random samples of comments and with filtering based on various criteria before settling on an emphasis on top-rated comments. I made a number of important discoveries. First, selecting top comments was as good a way as any of finding comments I thought were of better quality or greater interest. Second, some of these comments filled in important gaps in the story and the debate. Third, despite a perception expressed in many comments that other commenters lacked empathy, ratings told a different story. I will deal with this last point in some detail later.

Slashdot

The blurb for the article "Aaron Swartz Commits Suicide (from the rest-in-peace dept.)" is characteristically brief, resembling an Objective news summary:

> maijc writes "Computer activist Aaron Swartz committed suicide yesterday in New York City. He was 26 years old. Swartz was 'indicted in July 2011 by a federal grand jury for allegedly mass downloading documents from the JSTOR online journal archive with the intent to distribute them.' He is best known for co-authoring the widely-used RSS 1.0 specification when

> he was 14, and as one of the early co-owners of Reddit." (Soulskill, 2013a)

The blurb includes links to a previous *Slashdot* story about the JSTOR case, a *Boing Boing* article about his work on RSS 1.0, and an MIT article about his death.

The discussion. There are 589 comments in the discussion, the most of any *Slashdot* article that day, of which 41 are rated +4 or +5.

An anonymous first post kicks off the first main thread of discussion with what turns into a digression about the nature of freedom: "He wanted data to be free. Now he is free" (-1). Another anonymous comment writes about the attitude of Slashdotters towards open source software and possibly copyright. "We want everything to be free and open . . . Are we not hypocrites to say he cannot be free with his own life?" (+5 Insightful). In my long experience, the rational tone and individualistic logic of this discussion is typical for *Slashdot*. When circletimessquare replies that "death represents zero freedom" (+0 Troll), Xiph1980 replies with an argument that seems to be drawing on mathematical set theory: "Death doesn't represent zero freedom, it represents an empty collection of freedoms." In the thread that follows, circletimessquare is only able to make headway (so far as ratings are concerned) by explaining his own past struggle with pain: "if i had killed myself, i would have permanently destroyed the freedom i have now. suicide is a freedom destroying choice" (+4 Insightful).

This same thread on death and freedom produces an exchange about copyright. Swartz harmed the "freedom to make a living under existing copyright law." His "actions amount to reducing the collection of freedoms available of everyone in the entire scientific journal system" (+3 Funny). It is not the winning argument: gomiam replies with an argument in terms not of individual choice, but overall benefit. "The current copyright system restricts the freedom of the majority for no proven reason to provide monetary gain to a minority . . . That looks like a net loss of freedom to me" (+5 Insightful).

The second major thread in the discussion begins with a post about empathy by benjfowler, who criticizes the tone of discussion. "A young man took his own life. And so far, I'm only reading sick jokes and flamebait" (+5 Insightful). Despite the high rating, the top comments in the discussion that follows have a more rational than empathetic tone. MartinSchou defends the jokes, saying he is a "failed suicider" who finds that "humour is

a stress reliever and coping mechanism" (+4 Informative).

Part of the thread is about the mental state of prodigies. "Many people who are successful that early in life are rather high strung. The feeling of helplessness in dealing with a court case may have pushed him over the edge," writes jfdavis668, invoking a variant of Troubled Genius (+4 Insightful). "The flame that burns brightest also burns quickest," writes hairyfeet (+4 Insightful). An anonymous commenter blames the charges: "Imagine yourself stuck in a case where you are facing 30+ years" (+5 Interesting). User cjjjer doesn't buy it. "Lots of people have been faced with worse . . . and don't kill themselves" (+4 insightful). Jafafa Hots replies that he has never encountered someone who does not harbor demons (+4 Insightful). When an anonymous commenter claims not to harbor any (+0), Jafafa Hots insists that everyone has "issues" (+4 Insightful).

An anonymous commenter argues that one should not allow prison to push one to suicide. In response to the "Imagine . . . 30+ years" comment, he writes

> I don't fucking need to 'imagine' it, you fucking pussy. . . . I once faced 16 years' imprisonment for some trumped-up charges . . . So like most other people who get in trouble with the United States on the federal level, I copped a plea . . . In the end I served 30 months . . . what if I had responded to the prosecutor's BULLSHIT attempts to scare me and killed myself? . . . I can honestly say that prison was a growth experience for me. NEVER EVER GIVE UP, no matter what some bastard is doing to you. If my story is not powerful enough for you, look up the story of [Holocaust survivor] Primo Levi. (+5 Interesting)

To me, the Interesting tag suggests the rating is less for agreement than because raters appreciate authentic stories of personal experience. Cederic disagrees: "Fuck that system, and its suicidal outcome" (+4 Insightful). But Nethead had his own experience: "Been there too. Faced 1024 years. Copped a plea for 4 and the judge made it 6. . . . I lived through it and might just be better for it" (+4 Interesting).

Twice in this thread top comments criticize copyright, a perennial target on *Slashdot*. "All our insane copyrights and patents are doing is making sure that Asia becomes the next superpower," writes hairyfeet in the same post quoted previously (+4 Insightful). Elsewhere in the thread, waterbear takes aim at a copyright system grown beyond its original effectiveness or intent (+4 Insightful).

There are top-rated arguments on both sides about Swartz's guilt. A top-level anonymous comment, titled "Wish I knew why," asks for context (+3 Interesting). User kenh replies with the Lawbreaker frame:

> I read a bit of the indictment [link to the indictment] and I find it hard to believe the charges are 'trumped up' because they are so easy to disprove. . . . I think his passion for his political/legal positions drove him to commit crimes, crimes for which the penalty was so great it may have driven him to suicide . . . (+4 Insightful)

Later in the thread, he defends the same position with a link to a Wired article on the topic (+4 Informative).

An anonymous commenter compares Swartz with Alan Turing, the pioneering British computer scientist who did crucial cryptography work during World War II and committed suicide when forced to take drugs to suppress his homosexuality (+2 Insightful). When challenged, wonkey_monkey and wierd_w defend the comparison. The latter implies Democratic Struggle over copyright: "There was a sharp disconnect between what is ethically sound, and what is legally necessitated." In the case of copyright, "a morally offensive situation is being maintained . . . For the benefit of rentseekers (JSTOR, Elsevier, and all those other publishing house whores), at the detriment of public knowledge and education" (+4 Insightful). He also appeals to Unequal Justice: "big corps" grab control of research, but the public is excluded. Stirling Newberry writes, "Rents kill . . . All of us are losers, except the people with the corrupt rent stream" (+3 Insightful). User dave562 makes a similar argument: "The entire system has been co-opted and subverted to protect the monetary interests of the few. Whenever anyone steps up to threaten those interests, the DoJ and various other law enforcement entities step in to wreak havoc on those who dare to step out of line" (+3 Insightful).

Unequal Justice is invoked in several other threads that compare the treatment of Swartz with that of bankers. An anonymous top-rated comment reads: "Spread knowledge: 30+ years in prison. Kill a person: 10 years in prison. Rape a woman: I don't even know how much. Be a banker and fuck people's lives by investing their well-earned money into bad assets: Earn a bonus" (+5 Insightful). (The attitude towards women's issues is not atypical. By contrast, in the *Times* discussion the implication was that rape sentences are too short.) User waterbear makes a similar injustice argument, but

drawing on copyright law itself: “If Aaron Swartz had not committed suicide, his case would still look like oppressive overreaction by proprietary interests and by the justice system which too often seems to act as if it were their private proxy. . . . Any prison term at all, let alone up to 35 years, looks to me totally disproportionate . . . it deserves to be compared with false claims (made knowingly or recklessly) to copyright in cases where there is none” (+4 Insightful).

Compared to my other cases, top comments on *Slashdot* digress on several topics (death and freedom, copyright, depression and suicide, humor). In my experience, the exceedingly rational tone of discussion, with few expressions of empathy, is characteristic of *Slashdot* discussion. The comment system may contribute to these digressions. The hierarchical format allows a back-and-forth discussion to stray far from the original topic. By default, the site collapses all but top-rated comments to a single line; this allows comments deep within the hierarchy to stand out and attract attention for them and for nearby comments. Together, it seems likely that these design features enable digression and fragmentation of discussion, while also highlighting what readers consider to be top comments.

Ars Technica

The *Ars Technica* article’s title is “Internet pioneer and information activist takes his own life,” but the article’s subtitle better captures the text: “Aaron Swartz faced decades in prison for downloading academic articles” (Lee, 2013). The heart of the article is an explanation of Swartz’s activism, describing his convictions and explaining how he put them into practice.

The article ends with advocacy. While saying that Swartz should not have done what he did, the pursuit of him by the government was a case of Institutional Failure: “whether or not it contributed to his suicide, the federal government’s prosecution of Swartz was a grotesque miscarriage of justice.” The article concludes by urging changes to the Computer Fraud and Abuse Act under which Swartz was charged.

Unlike the other news articles in my study, this is written by someone who clearly knew of Swartz and shared some of his ideals. The author, Timothy B. Lee, says he met Swartz once, and worked with a team that helped make the PACER documents

“liberated” by Swartz publicly available. Lee puts Swartz in the context of other like-minded people (he explains the relationship with Lessig, and links to Lessig’s blog post, also in my study) and activist organizations, including Creative Commons and Swartz’s own Demand Progress. Other articles can leave the impression that Swartz was a fringe activist or lone renegade; this one leaves no doubt that he was not alone.

Still, in this account Swartz is unusual. “Aaron” (Lee consistently calls Swartz by his first name) “accomplished more in his 26 years than most of us will accomplish in our lifetimes.” His informal dress among a crowd of “legal academics” made him “easy to pick out of a crowd”: a “penchant for defying social convention” that “may have been his undoing.”

The core of the article, however, is an explanation of Swartz’s activism. After briefly sketching his professional accomplishments (RSS, reddit), it describes detail how Swartz “threw himself into political activism.” The largest section, titled “Guerilla [sic] open access,” explains Swartz’s PACER exploit. Swartz “reverse-engineered the authentication process . . . spun up some cloud servers and, using credentials purloined from one of the libraries, began scraping documents from PACER.” But Swartz “was also outraged by the high prices charged for scholarly publications.” Lee links to and quotes from Swartz’s “Guerrilla Open Access Manifesto.”

The article explains the prosecution of Swartz. Swartz is said to have been depressed about that; his previous bouts with depression are not mentioned. Though there are quotes from Lessig, Swartz and Swartz’s parents, none are from any of his critics.

The final line of the article is the phone number for the National Suicide Prevention Hotline.

The discussion. *Ars* only allows registered users to comment or vote. The Swartz article attracted 265 reader comments; only one article that day received more. The comment with the most up-votes received +199; the one with the most down-votes received -152. Among the 51 chapters with the most up-votes, the one with the fewest received +24.

The *Ars* discussion system lacks hierarchical threading, so threads of discussion

can only be represented by comments that explicitly reply to previous ones (typically by quoting them). As one might expect, this results in a relatively coherent flow of discussion consistent with the “rolling present” (Xin et al., 2010). It also concentrates attention towards the start of the discussion: of the first 40 comments posted, 32 are among the top 50 by votes up.

The focus of the discussion shifts over time; due to the flat format this means it also shifts spatially from top to bottom. Though comment topics are mixed, in general the discussion begins with the Tragedy frame, then shifts to criticisms of the prosecution. A back-and-forth debate about copyright and open access emerges. Towards the middle of the set of top comments empathy among some low-rated comments is criticized, and much of the conversation turns to depression. Other important topics that appear are Swartz’s tactics (e.g. whether they constituted legitimate civil disobedience) and comparison with the treatment of other cases (e.g. of financial wrongdoing).

Among top comments, discussion begins with expressions of sadness and loss but soon shifts to criticisms of the prosecution. This is interspersed with talk about copyright (there is a substantial back-and-forth thread) and critiques of Swartz’s actions. Some of these comments touch on suicide, particularly criticisms of Swartz for failing to accept punishment and suggestions that prosecutors pushed him too far. The last ten comments among the top 51 are all about suicide, depression, and the lack of empathy expressed in some comments.

The Tragedy frame begins the discussion with a post by tinycritterfromthesea, who writes “Very sad news” and provides links to Lessig’s blog post and the Remember Aaron Swartz site (+64/-3). The next two comments also express Tragedy, but Unequal Justice soon appears. User tuel2006 writes, “50 years for pirating documents??? Murderers and rapists get off so easy!!” (+169/-9). User dorkbert adds criticism of copyright: “I hope they’re REAL PROUD having driven a brilliant man to commit suicide over a non-violent (made up) crime with zero financial gain” (+154/-38). Sixclaws responds that they “are going to use his death as an example to threaten those who dare raise their voices against the system” (+45/-13). The comment with the most votes up (and the highest overall score), by Willie McBride, also invokes Unequal Justice:

Evidently a dangerous criminal. He should’ve limited himself to something

> less serious, like laundering billions of dollars for the narcos. If this doesn't demonstrate how irremediably fucked up is the copyright system I have no idea what else could. (+197/-18)

The reference to drug laundering is presumably to the HSBC case. Although Swartz was not actually charged with copyright infringement, the criticism is specifically of *copyright*, rather than of the justice system more generally. This may be prompted by the primary article's discussion of access to information and Swartz's copyright activism. It is also consistent with frequent stories about debates about copyright on *Ars Technica* and among technical folk in general. User wangstrademeous explicitly cites HSBC while criticizing commenters for a lack of empathy: "I've already noted more 'angst' regarding this situation than in regards to DoJ letting of HSBC for far more egregious actions" (+53/-8).

While ratings point to strong support for the Unequal Justice frame, Swartz is not let off the hook. Ostracus writes, "don't activists understand that not only what they do has consequences" (+77/-152). "Schwartz absolutely deserved to be prosecuted," writes iandanger, invoking the Lawbreaker frame, but then also draws on Unequal Justice: "what he was charged with is in disproportion and seems to be retaliatory for his past activities" (+71/-35). "A crime worthy of decades in prison?" asks dmsilev. "No, not even close. But please lets not assume that this was just a harmless little lark with no costs" (+92/-20).

For some, Swartz's suicide undermines his principles. User mavere writes, "suicide is not a logical argument in and of itself. . . . at best, it's an emotional plea to those who are already partial" (+41/-16). "Emotion bests logic every time, we are humans not computers," retorts LugidCG (+45/-8). TheFerenc replies, "What if Nelson Mandela had just committed suicide, rather than go to jail? Ghandi? If you're going to do the deed, accept the penalty, or be seen as the poser you are" (+30/-119). When confronted by amartens' argument that Swartz's suicide does not "mean that his ideas were wrong" (+91/-6), TheFerenc stands his ground: "It does mean he didn't truly believe in them, though" (+18/115).

Alfonse disagrees with TheFerenc's reasoning: "What it means is that he's a *person*, nothing more" (+66/-6). Yet Afonse also disagrees with the article's conclusion:

there can be no "'miscarriage of justice' until there bas been a *verdict*" (italics in original). The tragedy is that "Swartz couldn't find the strength within himself to push through this period and face what was coming" (+76/-86). This back-and-forth about suicide prompts wangstramedeus to criticize the comments for a lack of empathy. "I simply don't know how to cope with this level of inability to relate to another human being" (+58/-8).

Many top comments address copyright and open access, but without consensus. OTD Razor writes, "Nothing of public record belongs behind paywalls" (+80/-8). "Academic knowledge should be open and available for free," says ngativ (+35/-13). Society gains "[i]nformed discourse," argues c0g, especially since "[a]ntiscience is freely available" (+93/-8), while copernicum links to a White House petition for open access.

On the other side, issor objects to ngativ: "in the real world who foots the bill for publishing?" (+24/-37). When Tim Lee argues that authors are unpaid while Internet distribution is cheap (+74/-16), iandanger argues for copyright as an incentive, replying that destroying the existing business model could hinder access: "JSTOR can't just operate its functions without income . . . a large number of journals . . . would continue selling individual content at high prices" (+27/-2). Greho says to Tim Lee, "Your comment left me speechless." Academic works need editing, formatting and presentation. "[W]here I worked, any 'profits' from publishing sales were immediately reinvested in the organization, to support our membership" (+12/-0). When doppio approaches Democratic Struggle with his statement that publishers with paywalls "fully deserve to be punished" by actions like those of Swartz, he receives little support (+8/-15).

User questions the article's call to action: "who exactly is this 'We'? Tech nerds who post in message boards?" He questions Swartz's actions also: "What does society gain by 'freeing' PACER records and academic papers?" (+55/-33). Usec c0g replies, "Informed discourse. Antiscience is freely available. . . . A better informed, more educated population is a good thing" (+95/-8).

In general, participants in the *Ars* discussion agree with the article's appeal to Institutional Failure and introduce strong appeals to Unequal Justice, yet there is no consensus about Swartz's actions. Suggestions that Swartz was a Lawbreaker attract substantial support, while criticism of copyright hardly rises to the level of Democratic Struggle.

The Volokh Conspiracy

In the week following Swartz's death, on 14 and 18 January, law scholar Orin Kerr published two blog posts there about case. Initial reports had by then given way to analysis and criticism. The timing of his articles (he posted Part 2 on 18 January) thus affected their content and presumably the comments that followed. This may explain some of the differences from other discussions, and between the two blog posts themselves.

Kerr is a specialist on Internet law; his articles are narrowly focused on the case rather than on Swartz's biography or death. His two posts parallel the two phases of a criminal trial: the first deals with the appropriateness of the charges, the second with the penalties. They are by far the most meticulously detailed and reasoned of the articles in my study; the second is about 6,700 words long. Like Lessig, but unlike the news reports, Kerr is explicitly making an argument. He breaks up both posts into sections and subsections. Kerr provides links to and citations of relevant documents, including the indictment, the police report, precedents from case law, the law itself, and Swartz's own blog.

The *Volokh Conspiracy* comment system supports hierarchical threading. Any visitor can vote comments up, while only logged-in users can vote a comment down. Up and down-votes cannot therefore be directly compared.

Part 1: the law. In the first article, "The Criminal Charges Against Aaron Swartz (Part 1: The Law)," Kerr (2013a) sets out to answer the question, "Were the charges against Swartz based on a fair reading of the laws?" This is the more straightforward of Kerr's two articles. From the outset, the question the article addresses is whether there has been an Institutional Failure; even whether Swartz is a lawbreaker is out of scope.

Before laying out his reasoning, he outlines his conclusion:

> I think the charges against Swartz were based on a fair reading of the law. . . . once the decision to charge the case had been made, the charges brought here were pretty much what any good federal prosecutor would have charged. This is different from what a lot of people are hearing on the Internets, so I realize this post isn't going to be popular.[15]

15 The apparently deliberate use of the colloquial and inaccurate plural "Internets" suggests a slightly derogatory vision of a mass online audience.

(Kerr, 2013a)

Kerr first provides some background, including links to Lessig's January 12 blog post and a couple of news stories. He then reviews the facts of the case, though in practice some of this is speculative. In particular, he writes that Swartz "wanted everyone to have access to all of the journals in the [JSTOR] database," a prosecution claim that Lessig says only Swartz knew (and Lessig perhaps, if Swartz told him). There is some ambiguity here. Kerr labels this section "the facts alleged in the indictment," suggesting that he is only assessing whether the alleged facts match up with the charges brought, not attempting to determine their veracity.

Following this outline of facts, Kerr enumerates the prosecution's charges one by one, assessing each one against the facts he has presented, finally arriving at the conclusion that the charges were fair as a matter of law: "what Swartz was alleged to have done fits pretty well with the charges that were brought."

The discussion on part 1. The article received 649 comments. Over the period of 13 and 16 June, only one article received more, and that was Part 2. The comment with the most up-votes received +35; the comment with the most down-votes received -15. While anonymous readers could vote up, only logged-in users could vote down. The cut-off rating for the top 50 comments is +10. Kerr himself contributes ten comments to the discussion. Compared to those in other discussions, participants write many comments but cast few votes. Of the Swartz discussions I studied, these had the most comments per named user.

The discussion revolves around three topics: judgments of facts and law (did the JSTOR archive constitute property? was Swartz's access unauthorized? is Kerr's assessment of facts reasonable?), debates about intellectual property (should the articles be free?), and claims about the morality or otherwise of Swartz's actions. This is true both of the top-rated comments and of a random sample of fifty.

In the article, Kerr makes comparisons with existing case law. Debates about law are often expressed in terms of comparisons, similes and metaphors. Is the JSTOR archive like other kinds of property? Were the technical measures Swartz took to circumvent access controls like picking a lock? With the first comment, ZarcharyMartinez

launches a debate about whether the law should distinguish between overcoming trivial and substantial impediments to access:

> A good analogy would be the difference between someone picking a lock to get through a door (something that requires special skills), and someone opening an unlocked door and walking through it (something anyone can do). I believe that "computer abuse" statutes should only prohibit the first, but not the second. (+7/-4)

Kerr rejects the suggestion, arguing that copying someone's written password is "the quintessential unauthorized access." "It's like making a copy of your key when you're not looking," he writes, "no one would think that I'm then allowed to enter your home just because I gained access by using a key rather than by breaking the door down" (+14/-1).

ZacharyMartinez responds that there was no "lock" on the network's door. "The law shouldn't criminalise people opening unlocked doors and walking through them - if the owner wants to keep people out, rather than turning to criminal law, shouldn't they just buy a lock?" (+5/-3). Bob_from_Ohio rejects that argument with a bit of satire:

> If you leave your back door unlocked, have you consented for people to come in and use your toilet? There is not really any harm other than a small use of your water. You didn't have a sign or anything, just a door. (+13/-1)

In my experience, analogies like this are evident in many online debates about legal cases, particularly where technology is involved. Is MIT's network like a house? Is the IP-blocking instituted by the MIT administrators like a lock on a door? A similar comparison, present in many debates about copyright, is used by prosecutor Ortiz. She asserts that "stealing is stealing," suggesting that copyright infringement is theft. The usual justification for this is that infringement results in lost sales and is therefore indistinguishable from theft.[16] Infringement entails making a copy, however: nothing is actually lost; from this perspective it is unlike stealing. In law, infringement is a different category from theft. But the law does not always use words in their ordinary sense.

Such analogies rationalize the terms involved. The door lock is stripped of its

16 In fact, the connection between infringement and sales is difficult to measure. Any given act of infringement may or may not represent a lost sale. Infringement may even grow an audience or market.

physical characteristics and social context, while the network management practices of MIT are similarly reduced to a simple logic of inclusion and exclusion. In fact, both are technical systems consisting of both human and non-human actors. Claims that one is like another are put forward as judgments of the essential qualities of the systems and activities involved; they also rhetorically construct the meaning of those technologies and actions. A comparison of computer access to picking the lock on the door of a home recontextualizes that access in a way that brings into view phenomena and sentiments associated with the door: ideas of privacy, of property, intrusion, and so forth (Feenberg, 1999).

Analogy is thus used not only as illustrations of ideas, but as a basis for extrapolation. This is both necessary and dangerous. Necessary because the particular must be reconciled with the general, in reason and in law; when general categories do not already exist they must be constructed from categories that do. Dangerous because an analogy, once accepted, entails assumptions that can cut short debate. In this instance, ZacharyMartinez draws attention to some such assumptions. “It is an immense waste of energy for society to look to legal or criminal solutions what are technical problems” (+10/-8), he writes, trying to expand the question beyond matters of law alone.

As I mentioned previously, in the article Kerr repeats an unsubstantiated claim by the prosecution:

> Aaron Swartz decided to “liberate” the entire JSTOR database. He wanted everyone to have access to all of the journals in the database, so he came up with a plan to gain access to the database and copy it so he could make it publicly available to everyone via filesharing networks. (Kerr, 2013a)

Swartz’s intentions remain unknown. Kerr’s explanation is plausible, but whether it is true is something that he could not know. Nevertheless, it underlies much of his argument. Ryan Singel draws attention to this with a high-rated comment: “You say, without any shred of proof, that Aaron was going to ‘liberate’ the entire JSTOR database. . . . Motive is key to this case and you are making a *huge* assumption” (+12/-1). Kerr writes other comments after this one, but none in response to it. In Part 2, where this point is central, Kerr presents evidence and argument to support his contention, but

admits it is not conclusive.

Kerr's argument is tightly focused on a question of whether there was an Institutional Failure: were the charges against Swartz justified in law? This framing is challenged by several of the top comments. "I think this is a misguided effort," writes LaurenGelman (+13/-2). "Separating the discussion of 'the law' from 'the legal process' is a research memo and not an answer to whether the charges were appropriate for Aaron did." For her, the real question is: "Are these the laws we would choose"? "[T]he second post," she says, "is all that matters," anticipating that there Kerr would criticize the lack of prosecutorial discretion.

Other commenters break out of the legal question framed by Kerr. "You could get off with less time for raping someone, or just being HSBC," writes kagil (+16/-2), applying Unequal Justice. "Swartz may have broken the law but it is really law that should be ashamed. As far as I am concerned, the man was on the right side of 'History'," writes jostmey (+32/-8). Chui Tey takes the argument further, invoking Democratic Struggle:

> Ultimately, there is a difference between what is right and what is the law. The civil rights movement, Gandhi marching to make salt, all represents the difference between what a child would know is right and what a lawyer would claim to be right. . . . Research paid for by the largesse of the public belongs to the public. (+17/-3)

The comment is highly rated, but in the thread that follows it is hard to determine who has the upper hand. Most of the discussion remains about law and judgment, not values. Perhaps assertions of value leave too little room for argument.

There are repeated indications of tension between lawyers and non-lawyers, technical folk in particular (it seems likely some technical commenters are not site regulars). "I have not been explicitly authorized to post this comment. . . . Surely this is unauthorized access, as you do not have at terms of service published explicitly authorizing this action," writes Panopticrat (+5/-15). Many comments like this argue for effective meaninglessness on the basis of an ambiguous definition; they appear to have been written by non-lawyers. Respondents have little patience for such thinking. "You are being dense. Courts have very little sympathy for people are being dense," writes Cruxius (+13/-1). David M. Nieporent responds to another comment, "That last is something that would be said only by (a) an engineer who wants to prove he's smarter

than lawyers, or (b) an 'information wants to be free' ideologue" (+16/-1).

DaveJ writes, "Suddenly engineers like me are digging into every detail and nuance of the law and prosecutorial behavior." In response, loki_13 writes, "y'all are doing the same bang-up job digging into every detail and nuance of the law, as the lawyers here would be doing going over your code" (+1/0). Elsewhere, software engineer liberpolly says to Kerr, "I respect your legal expertise . . . What I see lacking, is your technical expertise," arguing that changing an IP or MAC address is not like stealing a password (+4/0).

Paralleling disputes about two technical fields is a clash of cultures around the norms of MIT. Mike argues, "MIT is open to the community . . . to encourage people to bend the rules and test limits. It trusts people to act irresponsibly. It's hard to understand if you're not from that culture." These differences, however, remain within the frame of Institutional Failure.

Part 2: prosecutorial discretion. In "The Criminal Charges Against Aaron Swartz (Part 2: Prosecutorial Discretion)," Kerr (2013b) sets out to answer the question, "Were the prosecutors in this case unfair in how they exercised discretion?" He breaks the question into four: was punishment appropriate; how much was appropriate; who may have been responsible for overzealous prosecutorial tactics; and what, if anything, should be done to change the law. He argues that appropriate punishment would be sufficient to deter future law breaking by Swartz, that any problem of prosecutorial excesses is systemic (the prosecutors in this case should not be scapegoated), and that "Felony liability under the statute is triggered much too easily" (changing this might have led to a different outcome for Swartz, as it was the felony label he refused to accept).

Unlike the legal logic of Part 1, Kerr's argument here is grounded in a moral claim. Kerr argues that, based on Swartz's Manifesto, statements and actions,

> he felt that there was a moral imperative to violate laws that he saw as unjust. . . . Swartz was not acting in the grand tradition of civil disobedience in which one willingly draws punishment to bring attention to the unjustness of the law. . . . Rather, he wanted to change the facts on the ground to make his preferred world a fait accompli. . . . he wanted to make the laws unenforceable, winning the debate unilaterally outside of Congress. (Kerr, 2013b)

Kerr presents Swartz's appeal to Democratic Struggle, but rejects it in favor of the Law Breaker frame. Given the principles at stake, punishment would have to be sufficient to deter further lawbreaking by Swartz. What level of punishment would suffice? For Kerr, "one of the puzzles about Swartz" is that he was committed to civil disobedience, yet although he had lawyers as friends and was deeply interested in legal questions, he was unprepared for the consequent punishment. "[H]e seems to have had his soul crushed by the prospect that he would spend time in jail."

In my view, this "puzzle" reveals much about Kerr's thinking. Punishment is known to be a poor deterrent even to self-acknowledged criminals. Kerr seems to expect Swartz to approach moral action as a game of rational calculation. If Swartz's belief in a "moral imperative" was sincere then he may simply have felt that he had no choice. There is a tragic inevitability to Swartz's confrontation with a legal system unable to even imagine his moral logic.

The discussion on part 2. Kerr's article received 793 comments. The comment with the most up-votes received +47; the one with the most down-votes received -7. The top 50 comments received +10 votes or more. Kerr contributes sixteen comments to the discussion.

Where Part 1 hinges on facts and law, Part 2 centers on ethics. Discussion is accordingly different. The top 50 comments revolve around two topics: prosecutorial discretion and civil disobedience. Both of these entail judgments not only of fact and law, but of values.

The majority of the top 50 comments are about whether the prosecutors employed appropriate discretion. There is broad agreement that this case reflects an Institutional Failure of the justice system. "It's not just this case and some other case, it's every case," writes TwelveInchPianist (+17/-0).

The first comment, by markefield, initiates a long debate about the prosecutors' culpability: "these two prosecutors are blameworthy just as other prosecutors are. Prosecutorial overreach is a big problem in the justice system. The fact that this particular case might become the lever needed to change that isn't a flaw, it's an opportunity" (+32/-7). Kerr replies that it is unfair to "single out two prosecutors who just

happened to be the ones who were assigned this case" (+25/-7). As he sees it, markefield is making the ironic suggestion that the prosecutors should be made an example of because they did the same to Swartz. Others disagree. Oliver Crangle replies to Kerr's comment, "There is no irony there" (+26/-1); Kerr reiterates, "The irony was Mark Field's reaction" (+10/-4). At this, markefield turns from logic to wit: "Not unless you're secretly Alanis Morrissette" (+8/-1). This reference to a song in which Morrissette famously misapplies irony rejects Kerr's claim while also subtly criticizing his understanding. The joke is supported by littlejohnson, who responds with a reference to a song by Britney Spears: "Ooops, you did it again…" (+3/-0). Elsewhere in the same thread, Billy Goat provides a more respectful rebuttal to Kerr, interpreting markefield's comment as a proposal for reform rather than scapegoating (+24/-0). This debate recurs in other threads. In effect, it is about whether to frame criticism of the prosecutors as Bad Actors or as representatives of Institutional Failure.

A number of comments point to systemic problems beyond the case. Jorge Emilio Emrys Landivar argues that inconsistent prosecution constitutes Unequal Justice, replacing the rule of law with "rule-of-don't-piss-off-the-powerful" (+14/-1). Elsewhere, he suggests that this is what Swartz had done. Marie_Antoinette agrees, saying the prosecutor "is a stooge of the same monied interests" as the SOPA supporters undermined by Swartz's activism (+13/-3).

User lllvlll argues that "the law itself is bad": "actual crimes go unpunished" because "we created an unjust society that is gerrymandered by money, corruption and power" (+14/-2). He refers to MF Global, a derivatives broker that failed after taking the money of customers to cover its losses. Chris Tompkins follows up by citing Lessig's comparison to those responsible for the financial crisis. Arthur Kirkland makes a comparison to lax prosecution of financial firms, citing HSBC.

The question of civil disobedience (and thus of Democratic Struggle), raised by Kerr, appears in many of the top comments. One of the longest threads on the topic follows from Arthur Kirkland's HSBC comment. Debate arises over whether Swartz's action constituted legitimate civil disobedience given that he attempted to avoid punishment. Critics of Swartz support Kerr's contention that Swartz was wrong to circumvent democratic processes. Supporters argue that the law and democracy have

failed. Justin Levine writes, “a growing number of people no longer have faith in the democratic process when it comes to intellectual property issues” (+13/-1). Both arguments appear in top-rated comments, but user ratings give critics the edge. I will explore the topic of civil disobedience in detail later.

Comparison of Swartz Case Frames

The articles about Swartz invoke various frames: Troubled Genius for *CNN* and the *Times* and Institutional Failure on *Ars Technica.* Lawrence Lessig proposes *Unequal Justice*, while Orin Kerr challenges the applicability of Institutional Failure and Democratic Struggle. *Slashdot*, with its brief news blurb, claims little.

Comments in every discussion contain arguments or frames not present in the primary article. What follows is a comparison of how top comments frame the Swartz story differently in different discussions.

Table 7 compares the occurrence of frames among top-50 comments across discussions about Swartz. The frames are Troubled Genius, Law Breaker, Bad Actor, Institutional Failure, Unequal Justice and Democratic Struggle. To highlight absences, they are indicated with a dash (–).

Table 7: *Frames in Top 50 Swartz Comments*

Site	LB	BA	IF	UJ	DS
CNN	3	2	8	–	3
New York Times	–	8	18	19	4
Ars Technica	4	1	11	3	1
Slashdot	3	–	9	5	2
Lessig	–	4	10	11	13
Volokh 1	28	–	13	3	1
Volokh 2	8	10	15	8	7

The arbitrary choice of 50 top comments could potentially mislead about the prominence of frames in a discussion. Table 8 shows frame occurrence for only the top 10 comments in each (for *Slashdot* I picked comments with scores of +5, of which there are more than 10 in each discussion).

Table 8: *Frames in Top 10 Swartz Comments*

Site	LB	BA	IF	UJ	DS
CNN	–	1	4	–	–
New York Times	–	1	6	4	1
Ars Technica	1	1	5	1	–
Slashdot	–	–	1	1	–
Lessig	–	1	1	3	4
Volokh 1	8	–	3	1	–
Volokh 2	1	2	4	2	3

These numbers tell several stories consistent with a close reading of the top comments. Top comments rarely blame Swartz, though the legal discussion on the *Volokh Conspiracy* does focus on whether Swartz was a Law Breaker—unsurprisingly, since that is the topic of the first article. Lessig was Swartz's friend; comments on his blog see Swartz's action as consistent with a larger Democratic Struggle. Claims of Institutional Failure are strong in all or nearly all discussions, outweighing invocations of Bad Actor or Troubled Genius. Gamson (1992) warns that individualization can depoliticize issues; that is not happening here. While Democratic Struggle is present in each discussion, it is not a major focus of attention: but Unequal Justice is present across the board, except at *CNN*, where much of the discussion is about something else (depression and suicide).

Appeals to Unequal Justice and Democratic Struggle demonstrate a radical interpretation of the story in all cases except perhaps *CNN*. Lessig makes such an argument; Orin Kerr on the *Volokh Conspiracy* considers and rejects it. The other articles make no such suggestion: yet commenters do. As for *CNN*, commenters make other important contributions not captured by these statistics by providing information not present in the article.

I present these statistics with some misgivings. I have not performed intercoder reliability testing. Even had I done so, there are strong reasons to treat numeric comparisons among comments with caution. Comments are drastically unequal (in terms of visibility, ratings, and role in discussion). As comments are often driven by disagreement, and commenters are a different group from readers and raters, the frequency with which something is expressed might not be related to the strength with

which it is held. My later analysis of empathy will show just how misleading frequency can be. Top comments that digress on other topics skew the numbers, as they do not express any of these frames. Even if intercoder reliability testing found results varying by twenty percent or more, it would have little impact on the analysis. What matters most is the presence or absence of a frame. I see the numbers not as an assessment of the relative strength of these frames in different discussions, but as a map of those discussions, one that can aid closer qualitative analysis.

Gamson identifies three components of collective action frames: injustice, agency, and identity. He argues that many of these are present in the conversations among working people in his study: “more than three-fourths of the groups had conversations on at least one issue in which someone articulated an injustice frame, expressed moral indignation about it, and was supported by others” (Gamson, 1992, pp. 57–8). Considering agency, he finds that “a majority of the groups . . . had a sympathetic discussion of collective action on at least one issue, and most . . . brought such matters into the conversation in some form” (Gamson, 1992, p. 82). Identity entails an “us” in contrast to some “them”; “some 86 percent of groups used an adversarial frame on at least one issue” (Gamson, 1992, p. 108). Although these elements were present in most conversations, only rarely (less than a quarter of the time) were they integrated into complete collective action frames (Gamson, 1992, p. 111). How do these findings compare with the presence of collective action frames and these precursors in the discussions I looked at?

For Gamson, injustice does not necessarily implicate antagonism: it could be the result of stupidity or “misplaced priorities” (Gamson, 1992, p. 111). By this definition, Institutional Failure is an injustice frame. The Unequal Justice frame introduces antagonism, encompassing injustice and identity (there is an “us” denied the same justice as “them”). The Democratic Struggle frame entails agency (Swartz’s action was part of a larger struggle).

Institutional Failure is present (and frequently dominant) among top comments in every discussion about Swartz. This is not surprising: nearly every article about Swartz contemplates injustice. *CNN* quotes Swartz’s family blaming the prosecutor for his death. The *Times* quotes Carl Malamud saying that what Swartz did should never have

been illegal. *Ars Technica* characterizes the prosecution as “a grotesque miscarriage of justice.” Lessig agrees; the *Volokh Conspiracy* articles deal directly with this question.

Absent from the articles themselves (Lessig excepted) is the us-them antagonism present in the Unequal Justice frame: but it is present in top comments for every discussion except for *CNN*. Even there, where there is a long digression about the nature of depression, Blake writes, “When information becomes cost prohibitive then you have class based education” (+41/-1). User handleym describes the existing law as “evil,” implying inequality and antagonism through comparisons to Gandhi and Rosa Parks. Democratic Struggle and an agency frame are present in top comments for all discussions.

Top comments present the components for collective action frames, but they do not go beyond talk to action. A comparison with eulogies on the *Remember Aaron Swartz* (2013) site illustrates how much stronger collective action frames could be. Of a random sample of 50 eulogies, 44% (22) express a shared ideal, 28% (14) refer explicitly to a shared identity or community (e.g. “the cyber community”, “nous sommes tous «Aaron Swartz»”), and 24% (12) argue for continuing his work. “We’re gonna keep working to make the world a better place more open and free,” writes Maximillian Gerschmann. Nor are these all friends of Swartz: only 14 in the sample indicate that they knew Swartz; those who said they knew him focused on personal memories and were much less likely to write about his ideals. In all, 27 of 36 posts (75%) by people who did not say they knew him refer to a shared ideal, a collective identity, the need to continue his work, or characterize him as a role model. Dana Weber, who did not know Swartz, wrote of having a son with autism:

> his work changed my son’s life for the better, forever. . . . My son’s best help came from the RSS feeds and papers that I could get access to that offered the truth and science about autism. Every morning I read the RSS feeds from academic journals world wide to find out more . . . I interned at Elsevier. I have seen all sides of the academic pay wall and I have felt my ignorance around my neck like a boulder . . . Until we see that the populace will never be scientifically aware UNLESS we have access to the information we will not be able to go from a people of belief to a people of ideas. . . . I will do my best to see that your ideas are not forgotten.

This is a particularly powerful example, but it is not alone among eulogies on the memorial site. In contrast, calls or commitment to action are extremely rare in the reader comments I studied.

Framing might explain the difference. Eulogies constitute a genre framed narrowly by practices and taboos. Criticism of Swartz, or of anyone else, including the prosecutors, is almost nonexistent; the Unequal Justice frame is consequently absent. Only three eulogies in 50 place any blame at all, often obliquely. Paul Ford writes, “To be treated so unfairly was an awful injustice.” Institutional Failure is similarly scarce. Eulogies instead find positive things to talk about. They are focused on personal experience, memory and feelings rather than critique or argument.

Another example illustrates the rarity of collective action in reader comments. In 2008, the Canadian government proposed a copyright bill, dubbed the “Canadian DMCA” (Digital Millennium Copyright Act), which was strongly opposed online. A *Slashdot* discussion (timothy, 2008) became a site of coordinated protest. “Please write your MP,” writes whisper_jeff (+5), providing a link to a government list of email addresses. I posted a comment myself encouraging others to contact mainstream media sites (+5); digitrev replied “Don’t forget the Ottawa Sun” (+4). Other commenters quote government replies so that they can be rebutted. “I already phoned in to my MP,” writes Tiberius_Fel, telling others to do the same providing a link with phone number and email information (+4). Many participants in this discussion post letters that they have written to Members of Parliament or to news media, explicitly or implicitly encouraging others to do the same using their letters for inspiration. One participant late in the discussion even submits proposed changes to the law: “Here Is a redefinition of section 41.1 I am working on” (+1).

In my experience and examination of the *Slashdot* record this kind of coordination is extremely rare in reader comment discussions. I believe I have only encountered a couple of times on news sites: for the “Canadian DMCA” on *Slashdot* and for SOPA/PIPA on *Ars Technica*. Even the original (U.S.) DMCA did not receive similar treatment on *Slashdot*. This particular protest benefited from a number of factors: awareness of the issue among hackers had grown dramatically since the DMCA (Coleman, 2013); copyright had become a front-page story; Michael Geist, linked to the

in original article, was a prominent opponent using social media to encourage protest across the country; *Slashdot* was a gathering place for hackers who were consistently opposed to the bill (support for it there was virtually nonexistent); protest organizations were active and protests were already taking place.

In this exceptional case the use of comments for collective action took place under extremely favorable conditions. It is perhaps not surprising that ongoing organized protest at a critical juncture coordinated through on Geist's blog and social media overflowed into another popular venue for discussion among like-minded people.

What the Swartz comment discussions do, however, is address a public and a world that is very different from that presented in most of the articles (Lessig's excepted). There are repeated references to the HSBC case and comparisons with punishment for rape. Where news outlets posit a world of at worst imperfect justice, many commenters project the Swartz case into a world in which justice is reserved for the few, whether they be football players or bankers.

As Warner (2002) argues, commenters go in search of a public that shares this world, calling to it and calling it into being. Judging by the ratings they receive, they succeed. Commenters responding to the *CNN* and *Times* articles explicitly reject the world and framing of the article. (Lessig and the *Volokh Conspiracy* are exceptions. Lessig himself adopts an antagonistic stance; Kerr considers and rejects it when explicitly foregrounds the assumptions in his framing.) The objective journalism of the articles becomes the basis for normative judgments. Rather than explicitly reframing the story, they embed it in a larger context.

Chapter 8: Edward Snowden Revealed

The Snowden story broke on 6 June 2013 when the *Guardian* (Greenwald, 2013a) published an article revealing that the American National Security Agency (NSA) is engaged in bulk surveillance of domestic telephone call metadata. The metadata collected by the NSA includes information such as the identities of callers and the duration of conversations: information that can be used to build a map of relationship networks, potentially revealing more about individuals than would the actual content of conversations. This was only the first in a series of leaks revealing activities by the NSA that appear to be contrary to U.S. law or to constitutional guarantees. Some of this spying had been secretly authorized by the Foreign Intelligence Surveillance (FISA) court; others (such as NSA personnel spying on their partners) was not.

One 9 June, the *Guardian* followed up with a story identifying the leaker as Edward Snowden, a security specialist with Booz Allen Hamilton, a contractor for the U.S. government. Snowden, who had fled to Hong Kong, agreed to be interviewed and reveal his identity in order, he said, to “inform the public” (Greenwald, 2013b). The cases I have chosen are articles and discussions in response to the initial revelation of Snowden as the leaker, at which point there was far less context for this story than for the Swartz one. The Snowden story is still ongoing, but I will not describe subsequent developments as they do not concern my cases.

Snowden Case Statistics

The tables below provide statistics about comments and ratings for the Snowden discussions. I have already explained these statistics when discussing the Swartz story.

Table 9: *Snowden Discussion Comments and Commenters*

Site	N	Named Users	Anonymous Comments	Mean Posts	Median Posts	Reply %
Ars Technica	459	193	–	2.4	1	–
CNN	13,989	2,062	1522	6.0	2	63
New York Times	1,799	967	–	1.9	1	54
Guardian	6,189	2,217	–	2.8	1	59
Slashdot	860	243	309	2.3	1	89

Table 10: *Snowden Discussion Votes*

Site	N	Top Rating	Most Votes	Most Up	Most Down	Mean Up	Median Up	Votes /N	Votes/ User	Most Up/N
Ars	459	269	271	270	214	16.9	6	25.0	59.5	0.59
CNN	13,989	861	906	874	103	1.9	1	2.8	19.1	0.06
Times	1,799	1,204	1,204	1,204	-	14.5	3	14.5	27.0	0.55
Guardian	6,189	6,242	6,242	6,242	-	14.8	3	14.8	41.3	1.01
Slashdot	860	5	-	-	-	1.0	1	-	3.6	-

Note: The Most Up/Comments statistic is calculated by dividing the most votes up by the total number of comments reported for the discussion. In two cases this was greater than the number of comments I actually retrieved: 13,997 for *CNN*, and 2,182 for the *Times*.

Snowden Frame Definitions

Two frames that apply to the Snowden articles, as well as articles about many other news stories, are Objectivity and Advocacy:

Objectivity. It is the journalist's role to report facts and to balance opposing viewpoints while excluding her own subjective opinions. Quotations from officials and others are objective because they truly reflect what was said. Most stories have two sides, each of which should be given a roughly equal chance to make its case.

Advocacy. The author has an opinion and is writing to persuade readers that it is correct, drawing on supporting evidence and presenting arguments against opponents.

In order to draw a little more detail out, I describe several facets of Objectivity:

Facts. The article addresses the question, "What happened?" It attempts to inform readers of the salient facts of the case, including: events before Snowden came forward, Snowden's actions and statements, and the reactions of government.

Analysis. The article addresses the question, "What are the implications?" In addition to stating the facts of the case, the article attempts to assess the significance of the story and possible outcomes. How important were Snowden's revelations? What will happen to him? Sometimes, as in the case of *CNN*, precedents are introduced.

Balance. The article addresses the question, “What do opposing parties have to say?” The government declared Snowden’s leaks illegal; Snowden claims he was acting in the public interest. To answer this question without appearing to lose objectivity, articles rely heavily on statements by officials or by Snowden himself.

Compared to comments, articles are long and combine several of these sub-frame facets, especially Facts and Analysis.

Discussions about Snowden are dominated by very different frames from the articles they respond to. For the Swartz story, discussion is dominated by the debate over who is to blame. Comments on the Snowden story divide much more clearly into two camps, for and against. I propose that the question that frames the comments of participants is:

Where do I you stand?

This often corresponds to the Advocacy frame for articles, but implies a greater emphasis on the commenter’s position. Comments responding to this frame take a position about Snowden’s actions: moreover, they identify the commenter with that position. The answer is necessarily subjective. This contrasts with the framing question I identified for the Swartz discussions (Who is to blame?), which (like the legal issues in the story) implies some objectivity. “Where do I stand” can be, and often is, answered with a bare statement of support or opposition.

Solidarity. The comment expresses support for Snowden, and establishes common ground with others who feel likewise. This is explicit support, distinct from the implicit agreement in comments addressing the issue rather than Snowden himself. Claims that he is brave or a hero fall into this category.

Renegade. Snowden rejected legitimate authority; his actions were therefore wrong. The authority in question could be the government or his employer, to whom he had made commitments; the American people or nation, which he betrayed; or the democratic process, which he circumvented.

Democratic Legitimacy. Something is wrong with American society or governance. This is encompasses policy failures (blanket surveillance does not prevent

terrorism), inappropriate persecution of Snowden, excessive government, even corruption. I include individual actions (e.g. by presidents Obama and Bush) in this category. The frame implies a structural problem and a consequent injustice.

Questionable Character. Comments expressing this frame bring Snowden's character into question, although they do not necessarily explicitly state that it is flawed. Typical claims are about the cost of his hotel room in Hong Kong, cowardice or his choice to flee to an undemocratic China competing with United States. This frame does not include claims about the inherent rightness or wrongness leaking the information. While it is conceivable that this frame could avoid the question Where do I stand?, in practice it is used to support or imply criticism of Snowden's actions.

Partisan Politics. The commenter affirms alignment with a political party or ideology, using the Snowden story to justify the position. Examples might include saying this is a consequence of voting for Obama, or for arguing that the fault really lies with Bush.

As with the Swartz frames, I developed these over time in response to my analysis of the data. Although there is some overlap with the Swartz frames (e.g. instances of Democratic Legitimacy sometimes occur together with Democratic Struggle and Unequal Justice), I chose to analyze the Snowden story on its own terms. I therefore avoided collapsing frames for the two stories together for fear that I might try to force what is said about one to fit the categories I created for the other. The one exception is Solidarity, which is certainly present in Swartz discussions, though not nearly as prominent as with Snowden.

CNN

The *CNN* article (Smith, 2013) clearly aims to be Objective. It tries to provide Facts about the case, and Balance arguments by Snowden and his supporters with criticisms of his opponents. Of the articles I looked at, I would say the *CNN* one is the most comprehensive in this regard.

The article makes Snowden the main actor in the story. Snowden speaks first and last. The title announces one of his main claims: "NSA leaker comes forward, warns of agency's 'existential threat'." The article presents a narrative beginning with the

background leading up to this point: what Snowden did, some of what the leaks revealed (though with little detail), and presents some of his motivations. “In a world where there’s no privacy,” he says, there would be “no room for intellectual exploration and creativity.” The article then asks, “Will he be extradited?”, shifting forward in time to the current reactions of the government. Officials and elected representatives from both major parties criticize Snowden, call for his prosecution and defend the NSA programs, giving credit to the spying program for two convictions. The article then balances these with challenges by Glenn Greenwald and Democratic Senator Mark Udall, both of whom argue for the importance of privacy and for greater transparency about NSA activities. Finally, the article returns to Snowden, gives a bit of his personal history, and considers his current situation. A final quote from him ends the article: “If they want to get you, over time they will.”

The debate between two opposing sides constitutes the core of the article. The public that the *CNN* article appears to address is one interested in a map of the facts of the case and the sides in the debate. This is very much the traditional journalistic approach with which readers would likely be familiar and comfortable, captured by Fox News’s slogan, “we report, you decide.” At the same time, the public are implicated: it is their security and privacy at stake. Mentions of terrorism, Americans, and Snowden’s aim “to inform the public” make it implicitly clear that American readers are not simply disinterested observers.

The discussion. The *CNN* site reported 13,997 comments, but I only found 13,989. This is the most comments for any story that day; the next closest is had 9,764 (see Appendix A). The comment with the most up-votes received +861. Anonymous readers could vote up; only logged-in users could vote down. The comment with the most votes down received -13. Among the 50 comments with the most up-votes, the comment with the fewest received +22.

It is doubtful that any one person has read all of the comments, yet there are replies throughout, indicating that most or all of them have nonetheless been read. As the default order shows older comments first (requiring many clicks to fetch them all—it took me over ten minutes to retrieve everything), some users must have chosen to sort comments by newest first or by highest-rated first.

Solidarity and support for Snowden are very strong among top comments. Of the top 50, 28 express support for Snowden or are critical of the government. “Just one word… Hero,” writes an anonymous commenter (+86/-8). Brief remarks like this are common among top comments. “Our current Government is the exact same Government our founding fathers fled from,” writes daverrrrrr (+30/-2).

Most top-rated comments fall into two groups. The first is isolates: top-level comments that do not begin threads with any highly-rated comments in them. Responding to the article, rather than to other comments, they are almost all one or two sentences. Many express Solidarity with Snowden. “Edward Snowden is an American hero!” writes jimweix (+41/-3). “This man will go down as a martyr,” writes an anonymous commenter (+86/-8). Others make brief judgments. “Our current Government is the same exact Government our founding fathers fled from,” writes daverrrrrr (+30/-2). “Thomas Jefferson would not approve of Big Brother America,” writes TampaJoey (+29/-1). (Two other top-rated comments also allude to Orwell.)

Three other top-50 comments are in their own thread, but the majority of top 50 comments are in a thread descending from a single comment by Elvisthree16, who writes “I’m not surprised . . . all phone calls are recorded and stored . . . without the hassle of a pesky warrant” (+871/-35). An Any Mouse underlines the message, writing briefly that “this is old news” (+386/-21). Elvisthree16’s comment has the second most up-votes; this thread is the first one seen by readers viewing best comments first.

User aphi confirms the assertion: “I worked on projects to do things like this back in the ’90s” (+59/-2). There are two obvious intents to such statements. One, which I have seen elsewhere, is to defend the spying: this is old news so there is nothing to worry about. A comment later by Gilsharkey takes this approach:

> this sudden outrage reminds me of that scene in Casablanca, where Claude Rains tells Humphrey Bogart “I’m shocked, shocked that gambling’s taking place in this establishment,” and then the croupier comes up and says, “your earning, sir.” See, this sort of government spying on select individuals has been going on since 1776 . . . there’s simply always been a tension between security, freedom, power and privacy. All just part of the social experiment called democracy (+45/-0)

There is a less sanguine interpretation of the “old news” claim, however,

indicated when Elivesthree16 refers to “a pesky warrant.” By saying that Snowden’s revelations are not new, commenters like Evilsthree16 are making a statement of their existing sentiments towards government. For them, Snowden only confirmed what they already knew or suspected about their government. This frame of Democratic Legitimacy appears to be the preferred interpretation of “old news.” This frame forms a matrix for the discussion in this thread with its many top comments, for it recurs repeatedly: it is both a background assumption among Snowden’s supporters, who dominate the top comments, and a theme to which they return following digressions.

At several points in the discussion, comments supporting Snowden provoke denunciations. When My Son Valets states that “Snowden is a hero” (+97/-11), GunnyNinja retorts that “he committed treason” (+70/-28). Three comments among the top 50 are critical of Snowden; all accuse him of treason. The highest rated is by stevedumford, who writes:

> I do love my Country and greatly appreciate that they are doing whatever possible to keep me and all my fellow citizens safe from terrorist attacks. What I do not love are traitors, both private and journalistic . . . (+236/-103)

Only GunnyNinja attracts more up-votes than do any subsequent replies (+70/-28), and not by much (compared to +59/-5), while also replying to a more popular pro-Snowden comment (+97/-11).

One example of the gravitational force exerted by the Democratic Legitimacy frame begins when reb362 suggests that Snowden was right to disobey “an unlawful order” (+196/-12). User newell london makes this a question of law: “It’s not unlawful as long as the President . . . . issued the proper Executive Orders” (+6/-0). In the thread that follows, newell london cites legal precedents including “the internment of Japanese Americans in WW2” (+12/-1), to which TYED responds, “The Internment of Japanese Americans in WW2 was legal and WRONG” (+35/-2).

A similar pattern begins when Carpe Diem blames President Obama (+14/-15). Dany Rioux responds that President Bush was responsible for the Patriot Act (+69/9). But Darth Homiiz rejects this partisan dispute: “Both Bush and Obama serve the same master. The next administration will take it a step further, no matter what party” (+44/-4, a

high score considering the comment was written over 2 hours later).

This non-partisan critique recurs in a series of comments about the state of citizen participation in American democracy. Thought Police writes, “We are all just too lazy to protest” (+874/-13). Gilsharkey’s comment about Casablanca, quoted previously, downplays concerns, but the third highest-rated post, by Sarah’s song, disagrees with the assumption:

> It is not that we are too lazy but the government is so divided down political lines, protesting would mean nothing! Haven’t you noticed that congress only listens to lobbyist and they only listen to people with big bucks. Does Occupy Wall Street ring a bell? (+412/-18)

Eye of Sauron explicitly ties this back to a rejection of partisanship: “People forget that they are Americans first, then maybe liberals or conservatives . . . these people clearly betrayed your trust and forgot about that they are granted the privilege by WE THE PEOPLE to server” (+305/-9). One user’s name even expresses this: antirepublicanantidemocrat. The most partisan top-rated post is by Badger who places the blame on Cheney and Bush for the Patriot Act (+39/-6), to which nmtaxes (the user name presumably meaning “no more taxes”) replies that the Democrats are just as guilty: “IT is both parties fool” (+34/-1). In contrast to this non-partisan critique among top comments, I noticed a number of instances of Democrat and Republican name-calling among low-rated comments elsewhere in the discussion.

Partisanship is thus a part of the Democratic Legitimacy concern that unites many commenters. It goes hand in hand with a sense in many of the top comments that there is an “us” and a “them,” like Eye of Sauron’s “WE THE PEOPLE” or the supporters of a hero whose trust is abused by his government. When stevedumford writes “I do love my country” he is endeavoring to define that identity as a “we” opposed to an external “them” terrorists. When toomuchhoopla calls Snowden a “likely spy and a traitor” who fled to China (+127/-53), reb362 responds by identifying antagonists within the country, rather than without, comparing the American government with the Nazis:

> This spying by our own government against its own people without a warrant is unlawful . . . No American should sacrifice freedom or liberty for security. The Nazis went that far with the Enabling Act. But at least the Nazis came right out and said it! (+196/-12)

This is a very high rating for a radical claim. In many online discussions, comparisons to Nazis are rejected outright on the basis of Godwin's Law (Godwin, 1994). Not here. Top comments expressing the Democratic Legitimacy frame do not bother to challenge the *CNN* article's presentation of government officials and elected representatives as legitimate: they simply start with a different premise.

Among top comments there is no rejection of the article at all. This is compatible with the objective framing of the *CNN*, which eschews judgment. After presenting Snowden's position fairly clearly, it leaves that up to readers. In the comments they do so, hailing Snowden as a hero but taking the critique farther than he himself has done.

The comments establish broad shared concern for what Snowden revealed and about government failures (e.g. of the two-party system). By expressing these views publicly, participants in discussion imagine and establish the existence of a constituency that agrees with them, whether through further comments or through votes. In Gamson's terms, injustice and identity are developed. Agency, however, remains stunted.

The Guardian

The *Guardian* article, "Edward Snowden: the whistleblower behind the NSA surveillance revelations", by Glenn Greenwald, Ewen MacAskill and Laura Poitras (2013), makes no pretense of impartiality. "Snowden will go down in history as one of America's most consequential whistleblowers, alongside Daniel Ellsberg and Bradley Manning," they write. Snowden is at the center of the article. A grid of photographs of his face heads it, and much of the text is told in his own words: he is quoted thirty times. There is a link to a video of an interview with him, while the long and detailed text of the article mostly summarizes what he has to say. It begins with his decision to come forward and what he has sacrificed to do so. It then describes his activities since publication of the leaks began: his flight to Hong Kong, and his concerns about reprisals by the United States. After this, the article steps back in time, giving a biography of Snowden up to the point when he decided to leak the NSA documents. Finally, it turns to his motivations. "For him, it is a matter of principle." It concludes with a quote from him: "I feel satisfied that this was all worth it. I have no regrets." At no point does the article challenge Snowden's statements; nor does it explore the substance of the leaks

themselves.

The authors' remarks amount to Advocacy, but the article itself is essentially a mouthpiece for Snowden. It implicitly addresses a public who sympathize with Snowden's story and aims. There is very little in the article that addresses arguments his critics might put forth. To the extent that identification between the paper and the story was already being forged in the early days of the leaks, readers of the *Guardian* might reasonably have felt that they were among like-minded people.

The discussion. The story attracted 6,189 comments, more than any other *Guardian* story that day (the next closest, also about the NSA spying, attracted 2,935). Anonymous users could vote. The highest rating was +6,242. Of a sample of the 50 comments with the highest ratings, the lowest rating +277. Of the 6,189 comments in the discussion, 14 are "Guardian Picks" highlighted by the newspaper's staff.

Comments on *The Guardian* story overwhelmingly support Snowden's leaks: of a random sample of 50 comments, 19 express support for Snowden or criticism of government spying; only 4 take a contrary position. A few of the remainder are ambiguous; most address other topics (e.g. expressing partisan political positions or discussing Julian Assange). Support is even greater among top-rated comments. Of the top 50 comments, 43 take Snowden's side explicitly or implicitly; not a single one criticizes Snowden or supports the actions of the NSA.

Reader ratings thus appear to magnify an existing bias in reader opinion, reducing diversity among the top comments. This winner-take all outcome is not surprising given the steep preferential attachment curve of reader ratings. Here, reader comments give an appearance of consensus. This is consistent with the content and tone of the comments. Furthermore, compared to the other discussions in my study, there is little in the way of debate or detailed argument. Among top comments, the dominant frame is Solidarity.

Expressions of Solidarity are typically brief, amounting to as little as a single word. The most common response is an expression of gratitude or praise. The first comment on the article, by whyohwhy1, with the highest rating (+6,242), reads only "Thank you!" Others say only "Hero" (+1,390) or "Brave man" (+1,060). Fully nineteen of

the top 50 posts express only one or more of these three themes: thanks, heroism or courage. Another eleven include similar remarks. Many of these are directed at Snowden himself, as though Snowden himself had written the article. “I am in awe of you,” writes evenharpier (623). “I can’t be anything other than proud of you. Good luck,” writes Laura9999. In all, eleven of the top 50 address Snowden as “you.” Remarks like these present are like applause following the performance of his speech (which is almost how the article reads) and deeds.

In addition to expressing Solidarity, many comments describe Snowden as brave or courageous. Some explain this in terms of the punishment the United States government is likely to inflict on him; the implication is that it is bravery that has made Snowden a hero. Contrast this with Swartz, whom many critics characterized as cowardly because he attempted to avoid punishment. Framing the issue in this way, as courageous heroism, focuses on the personal characteristics of the individual rather than the substantive importance of what he has done. This is consistent with the article, however, which is all about Snowden, not the leaks.

Although no top comments criticize Snowden, others that do are nonetheless present in the reactions they provoke among top comments. In response to whyohwhy1’s top-rated “Thank you!” comment, BonkIfYouHonk writes, “thank you indeed, but I wonder how long the guardian will keep flogging this story” (+227, just missing the top-50 cut-off of +277). This provokes a flurry of responses. “Please tell me you were paid by the NSA to write that,” responds RadicalLivre (+1,902). “I’d rather they ‘flogged’ this story than Thatcher’s funeral,” writes GRSmith300 (+1,667). Six of the top 50 comments are responses to that one rather mild remark by BonkIfYouHonk.

The pattern repeats when Cathy Henry writes, “This is a low-level person who runs to China. Real-nice, Glen Greenwald” (+42). Her criticism attracts rebuttals. User ardennespate writes only “Eh?” (+435). “Are you really incapable of understanding the issues involved in this?” responds msulzer (+554). Charles Driver steps in to defend Cathy Henry, arguing that Snowden is of Questionable Character:

> Ignore all those ignorant naysayers . . . The government was not doing anything illegal . . . All this fool did was expose, legal activities and then run and hide behind communist skirts, not giving one care about all the lives of people he may have put into danger. (+40)

Snowden's defenders shoot back at Charles Driver. Snowden reviewed the documents "to make sure he was not putting anyone in danger. . . . Are you sacred of an informed electorate?", writes Christopher Zemp (+303). User rrheard argues that the government was breaking the law, excerpting a legal analysis from the Electronic Privacy Information Center, a Washington non-profit (+325).

Several comments not among the top 50 challenge Cathy Henry's legitimacy. "Her first and only comment," writes siff (+230). "Just another drive-by Abraxas profile," writes marxmarv (+46, it is not clear what "Abraxas" refers to). User skullaria cautions "to everyone else": "be wary of sock puppets trying to make it look like the majority are against this guy" (+181). Cathy Henry's user page shows that she joined on 15 June 2012. During June and July 2013 she posted 89 comments, all of which appear to be criticisms of Snowden. Her account then falls silent until a comment about a different topic in April 2015. Was Cathy Henry a sincere participant in discussion or a strategic actor working on behalf of someone else who was detected by other commenters? Either is possible; judging from the ratings and the response, her remark had minimal impact.

These threads with their top-rated comments illustrate that when the Solidarity frame favored by discussion participants is threatened, they respond by defending Snowden and the article. The warning by skullaria about giving a false impression of majority views suggests concern that comments should be an accurate reflection of popular opinion: that a public exists whose authenticity legitimacy must be defended.

Despite the overwhelming consensus among most commenters, expressions of collective identity are curiously absent. User exturpicausa says "We're behind you Ed!" (+680) and two minutes later asks, "What can we do to help Ed?" (+311). In the short thread that follows, other than a brief disagreement with Charles Driver, the only answer to that question, by LandOfConfusion, is, "Create awareness" (+9). Top comments affirm shared values but not shared identities. Participants take a stand; they do not mobilize. Yet the comments are prolific; votes even more so: this is the only Snowden discussion I studied in which the top rating (+6,242) exceeds the number of comments (6,189). There is public action here: commenters express their views even if those views are already present (and can be voted up). This discussion is primarily one of expression rather than

debate or deliberation.

The New York Times

The *Times* article, “Edward Snowden, Ex-C.I.A. Worker, Says He Disclosed U.S. Surveillance,” presents the story as an objective outlaying of the facts of the story (Mazzetti, 2013). Interpretation, however, is left almost entirely to government officials and contractors: Snowden is less of an independent actor than an event, a stimulus to which they must respond; it is the government and the President who are expected to exercise future agency. The article is framed as an Objective Analysis of the situation in which they find themselves.

The *Times* article, begins with the revelation that Snowden was responsible for the leak and places it in the context of a concerned government. The persecution of Snowden by the U.S. government is presented as an imperative: “the United States must set up a strategy for prosecuting a man whom many will see as a hero for provoking a debate that President Obama himself has said he welcomes.” The “many” who admire Snowden are absent from the article. They appear implicitly when it explains that the revelations added to an already difficult situation for the government, which was “grappling” with “fallout” from other leak investigations.

The article quotes Snowden twice, once explaining his disillusionment with his time in the military, once about his motivations:

> If you realize that that’s the world you helped create and it is going to get worse with the next generation . . . and extend the capabilities of this architecture of oppression, you realize that you might be willing to accept any risks and it doesn’t matter what the outcome is. (Mazzetti, 2013)

This concern that the world would “get worse” is not explained. The article does not deal with the substance of Snowden’s leaks and only briefly alludes to his motivations. This quote is immediately followed by a brief discussion of the problem that intelligence leaks like this one pose for officials and contractors. The article continues to discuss prospects for Snowden’s arrest and the “awkward” situation for Booz Allen Hamilton and its vice chairman Mike McConnell, former head of the NSA. The article ends with McConnell explaining that intelligence is difficult when it is misunderstood by

groups who lack expertise:

> what makes it hard is that everyone has an opinion. There's very little appreciation for the threat, and there are so many special interests, particularly civil liberty groups with privacy concerns. That mix keeps us from getting to the crux of the national issue. (Mazzetti, 2013)

The main questions the article asks are about how the government will respond. Will Hong Kong complicate extradition? How can Booz Allen Hamilton defend their security and reputation? How does this play in the media? What steps is the government taking to satisfy journalists? The story is framed as problem that "presents both international and domestic political difficulties for the Obama administration." The public who are affected by the spying (the "many," the "fallout") are only alluded to in passing ("domestic difficulties").

The discussion. While the *Times* article presents an Objective Analysis frame for the Snowden story, concerning itself with facts, official statements and likely outcomes, commenters take a very different tack. They are preoccupied with and divided over the ethics of Snowden's actions.

Of the 2,182 comments on this story I was only able to process the first 1,799 due to a bug in one of the paper's comment pages. The missing comments are from later in the discussion; they are likely to have correspondingly lower scores and fewer responses. I do not feel that their absence is a significant problem. The next closest article in terms of comments that day attracted about a third as many (730). Only logged-in users could vote. The comment with the most up-votes received +1,204. The site does not support down-votes. Of the 50 comments with the highest ratings, the lowest rating is 109. The *Times* comment system hides replies by default; as a result, replies tend to attract few votes. Of the top fifty comments, only three are replies. Replies and dialog do exist, but they are far less prominent.

Most top comments are engaged in vigorous argument about the morality of Snowden's actions. The story itself only touches on ethics: though it indicates that Snowden had ethical motivations, it does not say what they are. Brief assertions by officials that his leaks harm security go unsupported and unquestioned. Top comments, in contrast, directly address the question of whether Snowden did right or wrong: and

argue for why the answer should be one or the other (although Tom Scharf argues that Snowden "deserves both praise and jail time" (+395)).

Among top comments supporting Snowden, the Democratic Legitimacy and Solidarity frames dominate. For example, the highest-rated comment, by robert bloom, praises Snowden and the *Guardian*, and references George Orwell's warnings for the future (+1,204). This attracts 23 responses. (None of which has a rating higher than +6, as is typical for replies in *Times* discussion.) Many of these replies disagree with robert bloom's support for Snowden or invocation of Owell. Dave, for example, invokes the Questionable Character frame: "Courage . . . is following your conscience, then staying here in the U.S. to face the consequences. Leaking information and then going to a foreign country is cowardice," replies Dave (+0). "'Wisdom'? 'Principles'? 'Courage'? . . . This man is a copycat leaker," writes borntorun45 (+4). Others agree with robert bloom: "There is no security with the government lying to us," writes show me (+0). A comment by arydberg takes Orwell as a cue to criticize the lack of labeling for genetically modified food (+1). Other comments revert to partisanship. "Does anyone seriously believe this kind of abuse would have occurred under a President Romney? . . . Thank you Democrat Party for showing us what American Fascism looks like," writes tpaine (+0). The sole reply to another reply in the thread is by Bill in Vermont, who disagrees with tpaine. These are low-rated replies, however. They resemble a random selection top-level comments more than they do top-rated ones. The problem of low ratings for replies is one to which I will return.

Partisan Politics shows up several times among top comments supporting Snowden, usually as criticism of Obama. "Mr. Obama should thank him [Snowden] for providing the transparency he himself promised but failed to give us," writes Brett Wharton (+774, the third-highest score). "A true American hero, and so unlike Obama," writes Adalberto (+196). Republicans come in for criticism too, however. Simon opposes "any politician who stands by George Bush, Jr. (Obama)" (+133). Brian Sussman blames "people like Dick Cheney and Don Rumsfeld," and argues that Obama should institute different policies (+116).

Given that much of the discussion is effectively a debate about the ethics of Snowden's actions, how do these two competing sides fare? Both pro- and anti-

Snowden camps receive substantial support from ratings. Of the top 50 comments on the *Times* story, 30 support Snowden or criticize government spying or lying, while 12 criticize Snowden's leaks. Analyzing exchanges is uninformative due to low ratings for replies. In order to assess the relative strength of pro- and anti-Snowden camps, I looked for a comment to compare with robert bloom's top-rated (+1,204) pro-Snowden remark. An earlier top-level comment by B. Dillon characterizes Snowden's action as a violation of trust and argues that those arguing for a right to the information "do more harm than good to this country" (+115). A later post by C. Henry, from Virginia, characterizes Snowden as a "traitor who sold out our national security to our enemies and then fled to China" (+377). The highest-rated criticism of Snowden is by blasmaic from Washington D.C., who calls Snowden "judge, jury and executioner" who took it upon himself to decide what was right and wrong, breaking the law without demonstrating that the NSA's actions are illegal (+650).

This Renegade argument appears several times in top comments: whether or not the NSA spying is justified, it is not Snowden's place to do an end-run around procedures put in place by elected representatives. Overall, however, there are more comments in the top 50 supporting Snowden than criticizing him. These numbers suggest that while both sides have a substantial following, supporters outnumber critics. The third highest-rated comment (by Brett Wharton, previously quoted) takes a moderate position: on the one hand he says he supports the surveillance program; on the other, he is opposed to the government secrecy surrounding it and argues that Snowden has performed an important service (+774).

NYT Picks. The *Times* site design and the practices of editors appear to strongly influence comment ratings. Of the top fifty comments on the story, only three are replies: and these three are NYT picks. The prevalence of top-level comments is disproportionate. Of the 1,799 comments in my sample, over half (964) are replies. As I have described previously, design choices on the site relegate replies to near invisibility. This inability to achieve high ratings does not seem to inhibit them: this suggests that commenters are writing for some other reason than the desire to achieve ratings or reach a large audience. On the one hand, it therefore seems unlikely that replies to *New York Times* comments are by public relations operatives engaged in strategic

communication. On the other, the practice of writing without being read bolsters Dean's (2005) argument that contributions replace communications.

There is an extremely skewed distribution of replies: fewer than 20% of comments attracted any replies at all. One reason for this is surely the two-level thread system: replies cannot subsequently be replied to. For comparison, the distribution of replies for *CNN* (which features multiple levels of threading) is less skewed, with over 40% of comments receiving replies.

Staff picks are prominent among comments, so it is worth examining how they relate to other comments in the discussion. One of the clearest indicators is how they are distributed between supporters of Snowden and opponents. The distribution of pro- and anti-Snowden positions among staff picks in the top 50 is nearly even, and equal to the distribution in the random sample. However, it diverges widely from the positions among other top-rated comments, as indicated in Table 12.

Table *11: Pro- and Anti-Snowden Stances in the Times*

Sample	**n**	**pro+ anti**	**pro**	**anti**	**pro/ pro+anti**
Random sample	50	29	15	14	52%
Top 50	50	42	30	12	71%
Staff picks in top 50	28	20	11	9	55%
Non-picks in top 50	22	22	19	3	86%

These sub-sample sizes are fairly small; the results may not be representative of *Times* practice. Taking a larger sample is problematic as only so many comments can be highly rated. Nevertheless, the difference between picks and non-picks is dramatic. If this result is significant, two obvious explanations present themselves. The first is that staff picks are chosen without regard to rating and therefore reflect the overall distribution of views. This is unconvincing, however, given the small number of replies among the staff picks described previously. It seems more likely that staff are choosing among already highly-rated comments.

The second explanation is that staff are deliberately choosing comments so as to balance out competing viewpoints. This would be consistent with journalistic norms of objectivity, which emphasize granting equal weight to two opposing sides in a disagreement. Cunningham (2004, p. 289) reports a paper doing this with letters to the

editor. Nielsen (2014, p. 482-3) found that journalists avoid participating in comments, preferring to maintain their professional distance and expertise. It would not be surprising if journalists extended their norms of professional practice to the management of reader comments.

Journalistic objectivity has been criticized on numerous grounds: for legitimizing marginal arguments (e.g. on climate change), for squeezing out pluralistic views (balance is almost always between exactly two viewpoints), for privileging authority, for failing to provide critical context, to name just a few (Cunningham, 2004; Hackett & Zhao, 1998). The standards of objectivity were used by journalists and commenters alike to criticize Glenn Greenwald for his handling of the NSA leaks: such critics claimed that because Greenwald clearly had a political position of his own, he was not a real journalist and his work constituted advocacy rather than legitimate journalism. The issue then became not whether the leaks were true, or whether the democratic arguments against NSA practice were valid, but whether Greenwald was the right person to be telling the story. The privilege granted to official statements by the *New York Times* story is likely another manifestation of this objective stance. The *Times* does not appear to take sides in the story, it merely quotes authorities.

Ordinary citizens are absent from the *Times* article: excluding the substance of Snowden's leaks from the *Times* report leaves out those affected. The article suggests a world in which officials are legitimate actors, in which Snowden's action was deviant, and in which all others are relatively disinterested observers.

But commenters do not adhere to the article's framing of the issue. For most of them, the morality and implications of Snowden's actions are at the heart of the story. There is no pretense to objectivity or balance, nor even the recognition that these are being rejected. By praising or denouncing Snowden, they construct a world of dynamic conflict with an "us" and a "them." The dominant frame is one of debate about the rightness or wrongness of Snowden's actions. The public imagined by many commenters is one of interested parties like themselves who may have disagreements over judgments and values, but who are engaged in the same debate—a debate absent from the article that prompted them to write. Ratings indicate that this debate and this public are not restricted to those who write comments.

The impact of staff picks on this discussion is ambiguous. The *Times* staff may be attempting to extend their journalistic principles of balance to the comment discussion. With their prominence, those picks have the potential to weaken challenges to the boundaries of debate and legitimacy of official narratives. Support for Snowden, which appears to be dominant among top comments, is made less apparent. On the other hand, staff picks must work with the material that is already there: most staff picks reflect the framing of the discussion, not the article, and highlight the conflict therein.

This does leave one question, however. Why are the pro- and anti-Snowden factions so evenly balanced in the random sample, while pro-Snowden supporters dominate among top rated comments that are not editors' picks? I see two possible explanations. The first is that the voting system magnifies small differences in opinion. I think this helps explain a similar discrepancy in comments on the *Guardian* story, but in this case staff picks are likely to reduce any such effect by balancing opposing views. The second is that the volume of comments does not reflect the views of raters. I later present evidence that replies tend to disagree with their antecedents: if top-rated comments support Snowden, then they are likely to attract replies that disagree (notice the much higher representation of replies in the random sample). Those replies, however, due in part to the site's design, are likely to be low-rated. Of course both of these factors may contribute, but evidence from analyses of other discussions (particularly empathy in Swartz discussions, discussed later) points to a difference between ratings and comment volume.

Slashdot

Slashdot articles are seldom more than a blurb. This one (samzenpus, 2013) is no exception. Titled "NSA WhistleBlower Outs Himself (from the man-behind-the-curtain dept.)," it reads simply:

> An anonymous reader writes "The individual responsible for one of the most significant leaks in US political history is Edward Snowden, a 29-year-old former technical assistant for the CIA and current employee of the defense contractor Booz Allen Hamilton. Snowden has been working at the National Security Agency for the last four years as an employee of various outside contractors, including Booz Allen and Dell. The *Guardian*, after several days of interviews, is revealing his identity at his request. From the moment he decided to disclose numerous top-secret documents

> to the public, he was determined not to opt for the protection of anonymity. 'I have no intention of hiding who I am because I know I have done nothing wrong,' he said." (2013)

More detail is available through hyperlinks to the *Guardian* article, a *Slashdot* submission about NSA access to Verizon data, and a Dice article on the topic (Dice is *Slashdot*'s parent company). The blurb takes an Objective Facts stance, standing as a placeholder for a story about which many *Slashdot* readers were already aware.

The discussion. The *Slashdot* article attracted 860 comments, more than any other article that day, of which 64 achieved scores of +4 or +5.

Slashdot commenters overwhelmingly support Snowden. In the very first comment, Jah-Wren Ryel writes, "This man may well be our Jesus. The government is going to crucify him in their fury" (+5 Interesting). Of the top 64, just one of is critical of Snowden: not because the author is opposed to the leaks, but because he believes that by outing himself Snowden put himself out of the fight. In a random sample of 50 comments, only one opposes Snowden—and even then only in the context of the author's other remarks in the thread (the comment says only, "And you're just a retarded leftist") (+1). In both samples there are several comments rebutting critics of Snowden.

Nearly all comments support Snowden and criticize the NSA's spying; Democratic Legitimacy constitutes a taken-for-granted consensus. DarkOx carefully lays out the argument that the NSA programs are undemocratic:

> The whole point of national security is to protect the nation. Part of the nation is our republican system of government. Well you can't have a representative government that is in any way democratic if people can't use the ballot box to judge the actions of the incumbents. People can't make good judgements when so much of what government actually does is classified and kept secret. Frankly I don't think its unfair or out of line to call what the folks at NSA, CIA, DOJ, 1600 Penn. are doing "un-American activities". (+4 Insightful)

Although arguments like this are only repeated a few times in top comments, my sense is that they are recognized as representative of shared assumptions underlying criticism of the NSA.

Several top comments express solidarity. "Dude thanks, what you've done

requires real courage and people like you change the world for the better," writes an anonymous commenter (+5 Insightful). "This man is a hero," writes X.25 (+4 Insightful).

The largest thread begins with the first comment, by Jah-Wren Ryel, previously quoted, who compares Snowden with Jesus. But it is the first reply to this that really gets the thread going when Confusedent writes,

> here's hoping the sacrifice isn't completely wasted. The fact that this stuff hasn't led to protesting in the streets really reflects just how complacent the US population is . . . I for one am ashamed I voted for Obama in 2008 (+5 Insightful)

Confusedent's criticism of Obama is not taken as partisan. DoofusOfDeath writes, "I'm not sure you should be ashamed for having voted for Obama in 2008. Try to remember the (realistic) alternatives we faced" (+4 Insightful). Both parties are blamed for the spying. "Yes, Bush started it and gets blame where blame is due, but Obama ran on a platform that included dismantling this program . . . he deserves all the blame we can throw at him," writes an anonymous commenter (+5 Insightful). User beamdriver argues that problem goes farther back than Bush: "most of this stuff, the basis for it anyway, goes back to Eisenhower. . . . the PATRIOT act . . . was just another step down the road we've been on for a long time" (+5 Informative). Later, an anonymous commenter suggests that "it was probably only the Nixon debacle which primed the public to actually reject this kind of snooping. Today we might just roll over" (+5 Interesting).

Commenters reject partisanship and critique the current two-party system in the United States. When commenters argue about the relative failings of presidents Obama and George W. Bush, they do not indicate support for either. "Is it so difficult to see that the two major parties are not, in fact, diametrically opposed on many things and that the things they share most are jack-booted hunger for power, oppression of dissenters, and authoritarianism?" writes KGIII (+4 Interesting). "Please understand that the 'left vs right' thing is just a distraction. Both parties are happily taking our liberties away," writes tukang (+5 Informative). "Bush and Obama are/were both on board with this program. In this regard they're equally evil," writes kelemvor4 (+4 Insightful). Confusedent goes further: "Voting for the proverbial 'lesser of two evils' is the mandate we give them to get away with all this crap. Neither side is less evil than the other anymore" (+5 Insightful).

User tgd suggests that the parties themselves are not the root of the problem: “the people involved in creating programs like this transcend any particular election cycle. . . . Its the inertia of huge organizations following misguided policies . . . most likely created by people who really believed it was the best thing for the country” (+5 Insightful).

There is some defense of Obama. When Grishnakh blames him for failing to take action to avoid war, end the drug war, and stop prosecuting whistleblowers (+5 Insightful), Alomex takes him to task, arguing that Obama has improved the first two but not the third (+5 Insightful, raters support equally high ratings for these two opposing viewpoints). There are a few suggestions of some third party alternative, but they are not the muscular claims of partisans. For example, the anonymous commenter who mentioned Nixon ponders the potential of the Tea Party to be a force for change:

> there are too many apologists and protectors of the police state, and not enough level-headed people willing to reject it. Perhaps the Tea Party radicals might actually be worth something, or maybe they’ll just provide an easy excuse to ignore the naysayers as conspiracy theorists. (+5 Interesting)

To which Trepidity responds,

> I do think that could be one possible positive outcome of the Tea Party, if it could be channeled into an anti-surveillance political force. An engagement with techno-libertarian issues has historically been a weakness of American libertarianism, which is to a large extent based on imagining sparsely populated frontier localism: no taxes, let me keep my rifle, I’ll fight off the government with my militia when they come, etc. (+5 Interesting)

I have encountered the non-partisan “pox on both their houses” stance repeatedly on *Slashhdot* over the years. Analyzing discussions about copyright I read similar sentiments about Republicans and Democrats; the only instances I found of vigorous support for one of the two major parties was in a story from 2003 (implying that the Iraq war may have been a turning point). As in this example, some enthusiasm for libertarians does appear occasionally, though often it is also strongly critiqued. There has also, in my subjective view, been increasing tolerance of left-wing thinking. In this discussion, Grishnakh writes, “Obama’s no centrist, he’s thoroughly right-wing. Unfortunately, the Republicans are extreme right-wing, so your choices are 1) right-wing,

and 2) even more right-wing" (+5 Insightful).

Returning to the issues raised by Confusedent, if party politics is not the answer, what is? In the same thread, Taco Cowboy writes,

> If the Arabs are so brave as to stand up against their tyrannical leaders, if the Turks are so brave to tell their "elected dictator" to fuck off, why can't we . . . Have we, the Americans, become pussies ? . . . Mr. Snowden has given me the hope, that my country is worth fighting for . . . No matter they are Democrats or Republicans . . . if they fuck my Constitution, I am going to fuck them back . . . I have . . . the DUTY, as an American citizen, to take back my government from those motherfucking tyrants !!! (+5 Insightful)[17]

Stiletto replies that there are no avenues for change, providing links to articles about how armed resistance, civil disobedience and protest marches have all been crushed, concluding simply "Vote? LOL" (+4 Insightful). "So long as the US government ensures most of the people have something to lose they won't revolt," writes currently_awake (+5 Insightful). User symbolset argues that martyrs are important. "It is the brutality of the oppression of the martyr that incites the rebellion, not his call for social change"—oppression is not desirable, but that is what will arouse resistance (+4 Insightful).

Beginning another thread, scottbomb writes, "This dude has balls of steel and I think deserves our help. If a fund is established, I'll gladly chip in a few bucks" (+4 Informative). One reply supports the call, but the discussion otherwise digresses around whether this would be effective and whether it would make supporters targets of government action.

There is sexist language in several of the comments I have quoted. I have observed sexism as a major failing of *Slashdot* discussion over the years, particularly when essential sex differences are said to explain why there are so few women programmers. Such tendencies are perhaps not surprising given the overwhelmingly male readership and the widespread use of this kind of language throughout society and on the Internet. That said, in recent years I have noticed a shift. Claims of systemic sexism in society are given much more credit than they were in the past.

17 Taco Cowboy's user ID is 5327, indicating that he has been a member of the site almost since the beginning. His user name appears to be a combination of Cmdr Taco (the user name of Rob Malda, *Slashdot*'s founder), and Cowboy Neal (one of the site's well-known editors).

Commenters on both the *Guardian* and *Slashdot* are overwhelmingly supportive of Snowden. But whereas top comments on the *Guardian* are wrapped up in expressions of Solidarity and debates about ethics, on *Slashdot* the discussion is not about two sides competing, or even about the specifics of the story itself, but about the implications and causes of the problems the Snowden case revealed. Two factors appear to contribute to this difference. The first is culture and framing: on *Slashdot*, Democratic Legitimacy is taken for granted as a frame for discussion. There is simply no need to argue ethics or express Solidarity, as these things are assumed. This consensus may in turn be a result of a relatively closed forum-like readership compared with news articles subject to larger popularity surges. The second is technical design. The *Guardian* only supports one level of replies, which limits digressions. *Slashdot*'s unlimited hierarchy, in contrast, encourages digression. Many of the high-scoring comments about how the U.S. got here and what can be done about it take place in digressions that stray from the core story of Snowden's identity. Yet while these comments are deeply embedded in the hierarchy, *Slashdot*'s default presentation highlights them by collapsing comments that have not achieved the threshold score: so new readers and participants are likely to be attracted to these somewhat off-topic threads.

Ars Technica

The *Ars Technica* headline, "Whistleblower who exposed NSA mass-surveillance revealed by The Guardian," is accompanied by a sub-headline, "29-year-old Edward Snowden wanted to reveal 'an existential threat to democracy'" (Geuss, 2013). The article presents the story almost entirely from Snowden's perspective, excerpting repeatedly from the *Guardian* interview. It begins by outlining the substance of what Snowden revealed, "one of the biggest exposures of privacy invading actions taken by the government without the public's knowledge," with embedded hyperlinks to four previous articles about the leaks on *Ars* and the interview at the *Guardian*. The article reviews the story to date, including a few quotes from Snowden that provide some insight into his thinking. His motivations are explained briefly using Snowden's own words: "what the NSA is doing poses 'an existential threat to democracy' because 'the government has granted itself power it is not entitled to. There is no public oversight'."

The article concludes with Snowden arguing that he, unlike Bradley Manning, “evaluated every single document I disclosed to ensure that each was legitimately in the public interest.”

The main text of the article is followed by a brief update linked to and quoting from a statement by Booz Allen Hamilton about Snowden’s employment. The use of updates is standard practice on *Ars Technica*: rather than revising the text of an already-published article (which appears to be a common practice on mainstream news sites), story authors and editors add labeled updates making clear what has changed and what has not.

While Snowden’s point of view dominates the article, *Ars* nonetheless applies and Objective Facts frame, using statements of fact and quotes from Snowden to tell the story without actually taking a position itself (this contrasts with the *Ars* coverage of Swartz’s death, which advocated policy changes). The article is partial by implication, but not explicitly. The added statement from Booze Allen Hamilton adds some Balance; nonetheless, the article is effectively one of Advocacy.

The discussion. The article attracted 459 comments, more than other articles over a comparable period on June 9. Only logged-in users could vote. The highest number of up-votes is +270, the highest number of down-votes is -214. Of the 50 comments with the most up-votes, the one with the least has +32.

While the *Ars* comment system does not permit hierarchical threading, there is a reply button on each comment for starting a new comment with a quoted excerpt. Of the top 50 comments, 24 are replies of this sort. In a random sample of 50 comments, 32 are replies; the higher number is likely because replies are concentrated later in discussion while early comments attract more votes.

Sentiment is overwhelmingly on Snowden’s side. Of the 50 top comments, 35 state or imply support, while two do the opposite. In a random sample of 50 comments, 19 are in favor of Snowden than the NSA program, while four are critical of him.

The first three comments express Solidarity. Doctor Hoot writes, “people have a right to be a whistleblower” (+212/-15). “I really applaud his bravery,” follows MAFIAAfire (+251/-9). “He will be vilified by government agencies . . . who will try and paint him as a

‘traitor to freedom and safety’, but history will vindicate him,” writes Drakkenmencsh (+203/-8).

The first sentence of the fourth comment appears to reference Doctor Hoot’s claim about the right to be a whistleblower. User metalliqaz writes, “There’s no right to be a whistleblower. The US/CIA tends to call that ‘treason’ and they have a knack for making people disappear” (+36/-155). This is an ambiguous remark: it could equally be criticizing the government or Snowden. Raters appear to take the latter view; while in my top 50 by votes up, the comment is hidden due to the large number of down-votes.

The theme of courage, mentioned in MAFIAAfire’s early comment, continues through many of the top comments. Graham J writes, “He is, simply, a hero” (+126/-19). EchtoGamut writes, “What he did is the very definition of what it is to be an American” (+34/-5). “This man has bigger balls than I do and I applaud him for it,” writes polarism (+93/-6). “This guy is one brave dude. . . . I wonder how many of us would/could do the right thing,” says MAFIAAfire in another comment (+251/-9).

Snowden’s courage is highlighted against an anticipated aggressive government response. Several commenters propose that Snowden revealed his identity in order to make covert assassination on imprisonment more difficult. “The attention will protect him for a little while, but once the news cycle moves on, this guy’s toast. Governments, and especially security agencies, do not forget and do not forgive,” writes byrningman (+120/-2). FFabian alludes to the sex investigation against WikiLeaks founder Julian Assange: “Unlikely that they’ll use rape again, too obvious—maybe childporn? doing drugs? spy for the commies?” (+124/-54). “His fear for his family is legitimate. He will be painted a traitor and will be hunted down,” writes zishbu (+212/-5).

Concerns about Democratic Legitimacy run through the discussion, echoing the concerns of Snowden himself. “I’m pretty sure that the very notion of a secret law is incompatible with a democratic government,” writes Facekhan (+39/-2). User knutsi asks, “what happens to the balance of power between a state and its people?” (+112/-12). User 00000 argues, “Governments that can arbitrarily dictate policy is conduct unbecoming a democracy. . . . If you ask me: the terrorists won” (+36/-2). User stragen001 writes, “if you don’t already know, go look up Orwellian and Oligarchy, then try to tell me thats not where the US is heading” (+38/-6). Several commenters, including

stragen001, express the view that government is supposed to be (but currently is not) "the servant of the people", as knutsi writes (+112/-12). Despite the often harsh language, many commenters make more narrow criticisms, focusing on specific policies and activities; government can remain legitimate even when certain of its activities are not.

As on *Slashdot*, a few commenters ask what is to be done about this state of affairs. EchtoGammut writes,

> What he did is the very definition of what it is to be an American. End of story. What we need to be focusing on now, is what are we as Americans going to do about this. . . . are we going to make a stand and demand an end to government violating our rights? (+34/-5)

"If ever there was a time for protest, this is it . . . I for one will attend any protests on this I can," writes harteman (+52/-8). BunnyPunch argues for the need to "protest" and "fight back," because "This is not democracy any more" (+57/-2). Despite evident support from raters, calls like these do not appear to lead to actual mobilization. A comparison with mobilization around copyright in Canada (as I found occurred on *Slashdot* in my preliminary study) or the SOPA/PIPA fight might help reveal what was missing in this case.

Comments critical of Snowden have a large impact of discussion despite their small numbers. Sobad writes one of the more highly rated criticisms of Snowden: "there are jobs In the military and government that critically require discretion. People violating that are also violating the trust of the American people" (+13/-38). The claim that Snowden is a whistleblower is challenged several times by commenters who see him as a Renegade. Technician1382 writes, "Leaking top secret court orders . . . does not make you a whistleblower. . . . The fact that he fears prosecution is evidence that he is aware of the penalty for disclosing top secret information (also called treason)" (+58/-173). This attracted more up-votes than any other comment critical of Snowden; it appears it may represent a substantial contingent of readers who have chosen not to contribute to a discussion where they would be a minority. This comment and others like it prompted many responses among Snowden's supporters, including eighteen of the top fifty comments.

Some of these engage in detail with the critics to whom they respond. “Call it treason if you want, but I see it as an obligation that every citizen needs to keep in mind, especially in a democratic society,” writes bluefinger in a comment arguing that the importance for secrecy must be balanced against the need for whistleblowers (+95/-4). “Please explain how we’re to vote out politicians based on policies they keep secret?” writes stewpidbarnes (+59/-2). Others respond harshly. “People like you . . . have forgotten who they work for and what they are trying to protect,” says shady28 (+34/-6). One thread of discussion, initiated by Intangible 360, expresses more general concern about Snowden’s critics:

> What I find most depressing are the folks in the Ars Technica community (ostensibly more intelligent than your average internet commenters) who look at this and say, “well he broke a law, therefore what he did was wrong.” Did these infants never progress out of stage 5 of Kohlberg’s Stages of Moral Development ? Do you really not understand that illegal and immoral are separate concepts. (+107/-14)

In discussions elsewhere (e.g. *CNN*), commenters often disparage one another as ignorant nobodies from the Internet. Here, Intangible 360 lays claim to a more positive (and restrictive) group identity. This comment also expects a high degree of consensus. Depending on how one assesses the data, support for Snowden is either dominant or overwhelming. The votes on the Technician1382 comment (+58/-173) indicate that criticism of Snowden could be running as high as one in four, but the actual numbers of comments and the votes on other comments suggest that Snowden’s critics may not even be that numerous. User zishbu wrote just two minutes and four comments before Technician1382, “This is exactly the kind of person I want working in the government,” gaining the third most votes up (+212/-5). Voters who saw Technician1382’s comment almost certainly saw zishbu’s also.

Does Intangible 360 expect the (supposed) intelligence of *Ars* commenters to lead to a very high level of agreement? Or does s/he perceive greater opposition to Snowden than I do? Gaining understanding of the opinions of others can be difficult for researchers and participants alike. Vote distributions on a steep curve and scattered comments are difficult for readers to assess collectively. Disagreeable comments may stand out. My data about empathy in the *CNN* story on Swartz, which I delve into later, provides more evidence that participants in comment discussions do not necessarily

have an accurate perception of the prevalence of others' views.

As in other cases, the ostensibly Objective stance of the article serves as a platform for ethical claims and opinions by commenters. Among the top 50 comments, there is one, by adminfoo, that critiques the *Ars* and *Guardian* articles:

> so far, what we have are kind of softball questions about motivations and fears . . . He seems to be articulate and very well capable of describing these things, technically, in great detail. I wante to see literally *hours* of videos where he speaks specifically about these things. . . . The news articles so far, feel like news articles from nontechnical people often feel: overly and wrongly summarized. Let's have the deep dive (+51/-0)

Comparison of Snowden Frames

In general, the discussions about Snowden are more polarized and less nuanced in their arguments than are those about Swartz. This is reflected in the framing question (Where do I stand?) and the frames I coded. The components of collective action frames are present to a lesser extent. The following table shows the distribution of the Solidarity, Democratic Legitimacy, Renegade, Questionable Character and Partisan Politics frames across the discussions.

Table 12: *Frames in Top 10 Snowden Comments*

Site	SO	DL	RE	QC	PP
Ars	5	4	–	–	–
CNN	2	6	2	1	–
Guardian	8	2	–	–	–
Times	5	5	3	3	3
Slashdot	3	4	–	–	–

These numbers indicate greater criticism of Snowden on the two mainstream American news sites, and imply that the *Guardian* discussion resembles those of *Ars Technica* and *Slashdot*.

A comparison with *Slashdot* makes it clear that this is not the case. Although both discussions express overwhelming support for Snowden, most top-rated comments on the *Guardian* are mainly about expressing values. A solidarity frame is strong. There is little analysis or deliberation; moreover, expressions of agency and shared identity are scarce. Top-rated comments in the *Guardian* discussion are shorter than average (21 vs

32 words), and short in absolute terms; top rated-comments in the *Slashdot* discussion are longer than average (median length of 72 words compared to 48), and long in absolute terms.

In my discussion of *Slashdot*, I proposed that the technical design of the sites might have an impact. Some of the differences might also be attributable to the *Guardian*'s nationality: British commenters might be less likely to engage with the nitty-gritty of American politics (although *The Guardian* attracts many American readers, while *Slashdot* attracts many from other countries). But I think there is more at work here. As a long time *Slashdot* reader, I find it unsurprising that Snowden support is so strong or that there is little need to justify it. What I see here is the reaction of a public that has already constituted itself around shared values, which may not be particularly surprised by the revelations (for years there have been *Slashdot* comments claiming that this kind of spying was routine), and which therefore proceeds to discuss the next logical questions: how did it happen, and what can we do about it?

The strength of the Solidarity and Democratic Legitimacy frame in all discussions indicates the presence of Gamson's (1992) identity and injustice frames. Agency is also present in discussion on *Slashdot* and *Ars Technica*. Like the *Guardian*, top comments on these sites strongly support Snowden. But where *Guardian* comments are largely restricted to expressions of Solidarity, top comments on *Slashdot* and *Ars Technica* bring up the matter of agency. *Ars*, like *Slashdot*, represents a narrower audience with established values. Rather than taking purely Objective stances, articles on both sites often speak to those values, as *Ars* does in the *Swartz* article when it argues that he was the victim of Institutional Failure, or the role *Ars* took in the SOPA struggle. It makes sense, then, that these would be platforms for taking (or at least proposing) political action on issues where there is near consensus.

Chapter 9: Public Judgments

Most of the on-topic comments in my cases address a framing question: “Who is to blame?” and “Where to I stand?” for the Swartz and Snowden discussions, respectively. There is one thing these questions share: they judge. Judgment, I propose, is a frame that encompasses discussions about both stories, answering my original question about what beyond marginality relative to a primary article unites the reader comment discussions I examine. This judgment in turn defines political space.

Before making this argument directly, I deal with the question of publicness. Are commenters conscious of and do they write for a public audience of unknown strangers, or are they focused on personal dialog? To answer this, I consider whom comments address and provide evidence that these discussions are not echo chambers: most dialog is among parties who disagree.

I address the significance of ratings, which are key to my research method. Do ratings make a meaningful contribution to discussion, and are they perceived this way by participants? I examine how ratings illuminate empathy in the Swartz case, and analyze instances in which commenters talk explicitly about ratings.

I then draw out key features of each of the cases in my study indicating that comments say things that the article to which they respond do not.

I return to the relationship between comments and a primary article. News articles written in accord with journalistic objectivity attempt to stand nowhere, taking no position. Almost every comment responds by standing somewhere. Judgment is one of the distinctive contributions of comments.

I illustrate judgment in detail by looking at how commenters on both stories discuss political legitimacy in general and civil disobedience in particular.

Finally, drawing on Arendt’s theories of action and judgment, I argue that the judgment of commenters defines the boundaries of a space of politics and legitimacy, and discuss the parallel physical space of the margin in which comment discussion takes place.

Public Discourse or Personal Dialog?

I argue that reader comment discussion is public yet fragmented, possessing no single narrative. What if comments and exchanges are more personal than public? If commenters are only speaking to one another, rather than to an audience or public, then discussion may be more like an aggregation of isolated statements and exchanges than a territory with an external boundary.

For Warner (2002, p. 67), a public is notional: participants address an indefinite imagined audience, co-creating it with those who, granting their attention, choose to be members. In practice, he says, people imagine *a* public as *the* public at large ("The general body of mankind" (Public, 1913, p. 1159). But just as two people might hold a personal discussion in a public space, taking no account of who else might hear, it is conceivable that reader comment discussion could be *in* public without being *for* public consumption. If so, like a network of correspondence among individuals, what may appear from the outside and after the fact to resemble a shared public landscape is for its participants dissolved into a myriad of overlapping contributions (to use Dean's (2005) term). One way to address this question is to determine to whom commenters are writing. Are participants actually talking to individuals rather than a wider audience? To whom do they think they are talking?

Comments themselves seldom indicate this directly. Many comments address others as "you." Remarks indicating personal acquaintance are extremely rare (certainly among top comments). This contrasts with Orin Kerr's efforts to talk to commenters. "Thanks, Bruce," he says in one comment; "Welcome to the Volokh Conspiracy, liverpolly," in another; in a third, "Mike . . . in response to your Tweet". He recognizes someone: "Ryan . . . I know you have strong views". Address like this suggests personal conversations rather than public discourse. But Kerr is the article author, known to all because of his privileged role. Personalized comments or courtesies like "thanks" and "welcome" are virtually nonexistent in comments by other users. When Omnivore1 responds to andreasma on Lessig's blog, he writes: "I shared your entire comment on my FB page . . . I don't know your real name but I linked to this and gave you credit." This even though andreasma is a prolific contributor (at least on the topic of Swartz), who wrote nine comments in this discussion, one on *CNN*, and one on *Boing Boing*.

Warner (2002, p. 72) points out that the addressee may be rhetorical: one may name an individual while in fact targeting a wider public. This seems likely for Kerr, who is surely conscious of his wide audience. A few similar comments are also directed at article authors. When one commenter writes, “Mr. Lessig, this looks to be . . .”, it is conceivable that Lessig is the only intended target (to use Warner’s term). But when a commenter in the *Guardian* discussion writes, “Glenn you have accomplished more . . .”, the comment appears to be a public acknowledgement: while it is addressed to Greenwald, its target is other readers.

In the case of Snowden, while there are plenty of comments praising him in the third person (“This man is a hero” writes X.25 on *Slashdot*), there are also many that speak to him directly. The top-scoring comment on the *Guardian* story, by whyohwhy1, reads simply “Thank you!” (+6242). It is one of many: evenharpier writes “I am in awe of you” (+623); TheIneffableSwede writes, “Thank you Mr. Snowden. You’re a brave man and a hero in my eyes” (+403). The top comment in the *Times* discussion is similar: “Thank you, Mr. Snowden, for your wisdom, your principles, and your courage,” writes robert bloom (+1204). “Thank you Mr. Snowden” writes dmarcoot on *Ars* (+67/-6). These writers presumably do not expect Snowden himself to see their comments. The target of their comments is surely readers of and participants in discussion. Why do it then? Second person address is undoubtedly more personal, more intimate, than a statement in the third person. Perhaps by portraying a closer connection with Snowden commenters aim to express greater Solidarity.

Article authors apart, when comments name another commenter it is almost always in passing: the emphasis is on what that person had to say, not any identifying characteristic of the individual or history of activity in other discussions. “Oaths go two ways, Alex,” writes Rev. E.M. Camarena, Ph.D. in *Times* discussion about Snowden, responding to an argument by Alex. “Are you seriously arguing that,” writes David M. Nieporent on the *Volokh Conspiracy*. Often any reference to the previous commenter is omitted, going straight to the point of discussion. Consistent with Gómez et al.’s (Gómez et al., 2008) finding for *Slashdot* that there is little evidence of sustained relationships among *Slashdot* users, there is an absence of personal chit-chat that would indicate existing personal relationships or the development of new ones.

Even if commenters are not personally known to one another, it is possible that they are only speaking to those of like mind. Some scholars worry that the Internet can promote echo chambers rather than inclusive discussion. Wright and Street (Wright & Street, 2007), for example, found that in some discussion forums replies are rare. In almost every one of the discussions I looked at, replies constitute the majority of comments. If commenters are speaking to acquaintances, one might expect them to share some common ground, whereas if they are speaking to strangers, they are more likely to disagree.

Halavais (2001) and Gómez et al. (2008) found that dialog on *Slashdot* often corresponds to disagreement. In my analysis of individual discussions I found that many top-rated comments contain within them a kernel of other comments that are not so highly regarded by readers. Numerous defenses of Snowden on *Ars Technica* are written in response to remarks by his critics: though most of the critics are voted down, the defenses are voted up. Similarly, empathy for Swartz on *CNN* and *Slashdot* is provoked by comments taking a hostile or callous stance. A clear example is rachel's top-rated (+116) reply on the *Times* (where replies usually receive low votes) rejecting a comment accusing Swartz of stealing. Unpopular comments can thus act as triggers or attractors for popular ones, without which many of the latter might not exist. The absence of this dynamic may partly explain the lack of detailed comments in the *Guardian*'s discussion of Snowden: with so few critical comments and only Snowden's position in the article itself, participants in discussion have nothing to argue against. (*Guardian* readers also expressed themselves by granting the top comment more votes than there are comments in the entire discussion.)

Is there evidence in my cases that replies are motivated by disagreement: that they judge, express diversity and juxtapose opposing views? To test whether disagreement is more common than agreement in my cases, I drew random samples of 30 replies from each hierarchical discussion (i.e., all but *Ars*).

I did not code agreement, as it is difficult to assess the extent to which a reply agrees with its antecedent. Almost every comment adds something other than agreement to the discussion: the more substantial the addition, the more negotiated any apparent agreement. Rather than attempting to evaluate shades of agreement, I coded

opposition: comments that focus on or appear to be motivated by a point of difference, criticism or correction. This category encompasses differences of opinion, corrections or refutations of fact, spelling, even name-calling. I am not concerned with incidental disagreement; opposition must be central to a comment for it to qualify. Many comments that I did not code as oppositional nonetheless are not in straightforward agreement: some ask questions, record personal experiences, express empathy, digress or use a comment as a jumping off point for talking about something else. Many such comments imply difference without expressing opposition. I have not attempted to operationalize these other kinds of response. As a result, my measure of opposition is likely to under-represent difference.

A further weakness of this quantitative approach is that it does not distinguish kinds of opposition. For example, in *Slashdot* discussions most opposition is substantive, whereas on *CNN* rude and disruptive responses are more common.

Table 13: *Oppositional Replies Across Discussions*

Story	**Discussion**	**Oppose**	**Oppose %**
Swartz	*CNN*	15	52
Swartz	*Slashdot*	21	70
Swartz	*New York Times*	18	64
Swartz	Lessig	17	56
Swartz	*Volokh* 1	20	67
Swartz	*Volokh* 2	21	70
Snowden	*CNN*	19	63
Snowden	*Guardian*	12	41
Snowden	*Slashdot*	11	37
Snowden	*New York Times*	13	43

Note: In discussions with two-level comment hierarchies (the *Times* and the *Guardian*), comments at the second level often reply to one another. In such cases I compared with the actual antecedent, not the parent in the hierarchy. A few comments from the sample were excluded entirely: three because they were by a user replying to him or herself, one because it replied to a comment that no longer exists (it was deleted by a moderator).

There is a high level of opposition in all discussions, though less for those about

Snowden, where broad consensus existed. Much of the *CNN* discussion about Swartz was about depression; this may explain the lower level of opposition compared the Snowden discussion. The high level of disagreement (combined with a general lack of tact) suggests that commenters are not speaking to like-minds or acquaintances.

That alone, however, may not account for the high level of opposition replies. In the *Guardian* discussion about Snowden, a small number of critics provoke numerous replies in support of Snowden. In that case, majority opinion is piling on to criticize a minority. But participants in discussion can have difficulty assessing the reactions of the public of which they are a part, as my subsequent analysis of empathy will demonstrate. Exponents of majority viewpoints, not realizing their dominance, may not feel content to rest on their laurels. In an oft-cited *XKCD* cartoon (Munroe, 2008), a man sits at his computer while the voice of his partner calls him away. He replies that he must write this one post, because “someone is WRONG on the Internet.”

On the one hand, this seems likely to discourage the expression of minority viewpoints. If one knows one’s views will meet with be met with a wave of opposition or disappear among a mass of down-voted comments, there may seem to be no point to expressing them (e.g. in the *Ars* discussion about Snowden, criticism seems to be expressed more through down-votes than in comments). Uncertainty leaves hope that one’s words might have an impact. If that viewpoint is not visible, all the more reason: supporters may simply be waiting for a comment on which to hang their votes. Conversely, provoking down-votes or criticism can be taken as a mark of distinction. Cheng, Danescu-Niculescu-Mizil and Leskovec (2014) found that down-votes encouraged their authors to be more prolific, not less, and their subsequent writing was perceived to be of lower quality.

My experience is that agreement can suppress dialog. For example, in the discussion on Lessig’s blog lissack takes the unpopular position that Aaron was the author of his own misfortune (0/-2). Killswitch, who had written a four paragraph comment along the same lines, responds, “Well said. I would not have posted my reply if I had seen this before. You said it more eloquently than I did” (0/-1).

I myself do not comment unless I have something to say that has not already been said: if it *has* already been said, I simply vote for it. Sometimes I feel compelled:

perhaps there are others out there like me just waiting to vote for a comment we can support; or perhaps I have finally found the right argument. There are always the unknown others who might yet be reached, who may only be glimpsed through a comment here or a vote there. Arendt argues that human natality makes possible the unpredictability of public action. This unpredictability propels participation while leaving discussion in an unsettled state: perhaps it is this unpredictability that drives participants to comment.

A comparison with other formats (e.g. forums or social media) would give a clearer picture of whether opposition is distinctive for reader comments. Still, in my view this data is consistent with my experience that discussion often focuses on disagreement and opposition, and that difference is a driving force of dialog. A reader can often predict that replies to a comment will present oppositional or alternative views. The data illustrates how replies, in contrast to ratings, foster the extension of discussion in space while differentiating the terrain. In a hierarchical discussion, ratings and replies are bound together on the page. A popular comment is a magnet for attention, and for opposing replies that stay with it in the space of the page. Even in a discussion exhibiting strong consensus, unpopular views can ride highly-rated comments to the top.

The lack of personal address and the prevalence of opposition and disagreement among replies suggest that the discussions I have looked at are indeed diverse, as I claim, not echo chambers or conversations among friends.

The Significance of Ratings

Reader comments allow two main forms of participation: writing and rating. My evidence is that commenters reply to strangers with whom they often disagree. Raters are like a public audience: real but unknown. Are commenters conscious of the audience of raters? Do ratings make a significant contribution to discussion?

I have claimed that comments are radically unequal, and that ratings provide one way to take this into account. The assumption that ratings are meaningful is critical to my method. If it is true, then highly-rated comments should, on average, be qualitatively different. One of the best examples of this is the treatment of empathy.

Large-scale comment discussions in general have acquired a reputation for

exhibiting a lack of empathy. It is not hard to find insensitive remarks about Swartz's death. "He wanted data to be free. Now he is free" writes one anonymous commenter on *Slashdot*. "12 years past his prime, then?" writes another (-1). On *CNN*, where moderator intervention is rare, many commenters accuse him of cowardice. User cflisthebest writes, "coward. couldn't face the reality of his problem so he instead takes the cowardly way out" (+1/-4). Joe Bell writes, "What a looser. Thank you for saving us the theatrics. . . . prison bed sheets will do the same job" (0/0). On *Ars*, Q1DM6 says, "Not only was he a criminal, he was a weak coward to suicide. If you want to hold a weak coward up on a pedestal, you go for it" (+16/-43). Bad Monkey! comments, "What a drama queen" (+10/-53).

Remarks like these foster a perception of a pervasive lack of empathy. User benjfowler on *Slashdot* writes, "A young man took his own life. And so far, I'm only reading sick jokes and flamebait. . . . The first posters to this discussion should take a long, hard look at themselves" (+5). Lilyfromthevalley on *CNN* says, "There's a lot of cruelty out there in the remarks made" (+8/0). Kelly Rusinack writes, "insensitivity - the American passtime" (+22/-3). On *Ars*, wangstramedeous finds "the lack of empathy" in comments that calmly analyze the case's merits "astounding and soul corrupting . . . I simply don't know how to cope with this level of inability to relate to another human being and not feel sick" (+52/-8).

Yet the ratings on comments like these give the lie to perceptions of unfeeling readers. Empathetic comments consistently out-score callous ones. This is most apparent in threads about suicide and depression on *Slashdot*, *Ars Technica* and *CNN*. On *CNN*, Eddie Malloy writes, "why does it seam that the average reader here has the emotional maturity of an 8 year old? . . . . Suicide is an act of desperation to escape unbearable pain. What we lack collectively more than anything is empathy" (+47/-2). Idean50 says, "I found my father 3 months ago with the gun still in his hand . . . We cannot comprehend the 'aloneness' that they feel . . . if . . . there is a heaven - I believe suicides get to be the first in line" (+51/-2). On *Ars*, Goofball_Jones writes that "calling suicide victims 'cowards' . . . shows they have no understanding of the matter . . . nor have they any experience" (+39/-2).

Sometimes odious remarks provoke similar responses. On *CNN* Indigo Wizard,

writes of Swartz's death, "I am glad that one more socially inept fan-boy is not breathing my air or using my bandwidth" (+17/-17). TheAmused responds, "Swartz was a genius who probably did more for your ability to use the Internet than you will ever deserve. He never used YOUR bandwidth or air, but you probably sucked up HIS and wasted it" (+84/-1). This particular response is similar to the polarizing effect of incivility (Scheufele et al., 2013), but it is not typical of responses to insensitive comments in the sample. When on *Ars* Q1DM6 writes of depressed people, "Some of these folks need to man up (woman up?)" (+6/-41). HalationEffect shoots back, "even assholes like you deserve better than to be surrounded by assholes like you" (+32/-3).

Expressions of empathy were most scarce on the *Volokh Conspiracy* where depression and suicide were only minor topics of discussion. Of the top comments, the strongest expression is by bernard, in the second discussion, who decries the prosecutors' threats:

> It's very easy for you, with a great deal of experience in these matters, to shrug it off . . . For Ortiz and Heymann this was all in a day's work. They could see frineds, take time off . . . Not so Swartz. He lived with these threats every hour. (+20/-2).

Jack Dee responded almost immediately,

> It's the prosecutors job to make your life miserable so you will be deterred in the future . . . and so that you will agree to a plea bargain. . . . It's not the prosecutors fault that Swartz was mentally frail and could not stand the pressure. If you can't do the time, then don't do the crime. (+15/-2)

Though there is little discussion of empathy here, these ratings suggest it may be weaker than in the other discussions I looked at. Might this be influenced by the technical framing of law?

In the case of empathy, there appears to be mismatch between the preferences of raters and the perception of discussion by participants, some of whom seem not to take numeric ratings into account. Low-rated comments provoke rebuttals in the comments themselves: commenters express empathy because they perceive its lack. Yet the consistency of ratings also demonstrates that more is going on that simply discussion among commenters. There exists an audience of raters whose views are not reflected by comments alone. They judge comments and viewpoints by voting on them;

in so doing they influence the space of discussion physically (through sorting and filtering mechanisms that rely on ratings). They reveal agreement among participants, shaping the conceptual topography of discussion.

It seems likely that most commenters must have some awareness of ratings. After all, the act of rating is more common than the act of commenting: all discussions had more votes cast than comments written, from a low ratio of 2.8 to 1 in *CNN*'s discussion of Snowden to a high of 29 to 1 in the *Times*'s discussion of Swartz. In some discussions more people voted for the top-rated comment than wrote comments (on *Ars*, *Lessig*, the *New York Times* and the *Guardian*). Comments with high ratings tend to attract more replies. Ratings influence the sort order of comments, which are shown and which are not, which are chosen as editors picks and which are hidden by down-votes. Judgments expressed through ratings shape both the physical space of comments and the conceptual space of convergence and agreement.

Lacking access to the experiences of participants, I examined comments in which discussion participants talk about ratings. In the cases I studied, there are few top comments about ratings rather than the issue at hand. I had to seek comments about ratings out by searching for relevant terms: "vote" and "recommend" on news sites and "mod" on *Slashdot*. Such comments occur everywhere except on Lessig's blog and the *New York Times* story about Swartz.

As previously described, existing research on *Slashdot* confirms that most readers there are conscious of and value ratings. A typical *Slashdot* comment by an anonymous user reads, "Parent should not be modded as Funny, but Insightful" (+0). This is a common strategy on the site: an anonymous reader or user who currently lacks mod points can instead write comments to draw the attention of moderators. "Mod parent up" is a common comment, alerting moderators that a post is worth evaluating. As it happens, the post to which this comment refers is now moderated Insightful.

A couple of the comments I found use vote totals as an indicator of general opinion. A *Guardian* commenter writes in the Snowden discussion, "4364 comments as of now, and the first comment of all - 'Thank you!' gets 4837 recommends! These are the biggest amounts of comments and recommends I've ever seen on Cif so this must chap's actions must have touched a very deep collective nerve" CultRitual (+17). On

Slashdot there is even discussion of ratings on another site: ohnocitizen writes that a *reddit* article about Swartz has “1294 votes and climbing” (+5 Informative). On Lessig’s blog, when Dantes writes to andreasma, “Who cares what you think?” (+6/-13), andreasma shoots back, “As of last count, 158 people cared what I think. Please refer to the little counter at the bottom of my comment, next to the up arrow. It’s the same place that your comment has a ‘1’” (+32/-4).

In the *Guardian* discussion several comments suggest the NSA might spy on rating behavior. “Don’t recommend this or you will probably end up on one of CIA’s lists.” writes HenryDeCommentarius (+8). On the *Times* story about Snowden Julie asks, “Would even clicking ‘recommend’ qualify my name and contact information to be handed over to a government agency?” (+17).

Some comments are directed at anonymous raters. “Down voters be damned,” writes Q1DM6 on *Ars* (+17/-41). LionessLover on *CNN* writes, “go vote this down, humorless commenters” (+3/-2). On *CNN*, USAPeasant declares that “People who hate facts vote down” (USAPeasant) on *CNN* (+0/-1). I found six other similar remarks by the same user, in one instance requesting that voters “Please explain the vote down and why you dislike the truth” (+1/-1). An anonymous commenter on *CNN* writes that “whoever downvoted this” is naive (+0/-0). On *Ars*, Cervus edits an earlier post in order to respond to down-votes. “Downvotes? What I’m saying here is that they won’t get away with it because of whistleblowers like Mr. Snowden” (+90/-18); judging by the rating, the clarification worked. On the *Volokh Conspiracy*, in a rare instance of a commenter expressing acquaintance with another, ChrisTS defends another commenter: “To the person who gave you a downvote: Pers was teasing me . . . We are political opposites in many respects, but we also respect and enjoy one another.”

As these comments illustrate, down-votes provoke more comment than up-votes. “So much unliking these days,” writes Dark Star in the *CNN* discussion about Swartz (+4/-5), to which barbie1311 replies, “That unlike button really empowered some of the nimrods.” In the *Slashdot* discussion about Snowden, ebno-10db cites the *Slashdot* norm that comments should not be voted down based on personal disagreement: “(-1: Post disagrees with my already-settled worldview) is not a valid mod option” (+1).

In general, comments about ratings are rare and themselves receive low ratings:

but on *Ars* criticism of down-votes blows up into a full-scale thread. This is initiated by Sobad, who takes aim at Snowden (and people "with their liberal arts degrees or jobs at Starbucks"), writing, "At the rate with people are being down voted for having opposing viewpoints, it feels like being a PC fan in one the Apple posts" (+13/-38). This is a genuine problem: VictorChan complains that one of his comments "got down voted to oblivion" (+19/-2). Although Sobad's comment is hidden due to its low overall rating, it is quoted in six other comments that continue the thread. Though other comments and raters on *Ars* appear to disagree with Sobad's views about Snowden, they take his complaint about down-votes seriously. Quiet Desperation supports Snowden, but worries that

> Expressing an opinion one side gets you up voted to the skies. Having an opinion on the other side, no matter how presented, gets you modded to a collapsed comment. . . . I see . . . people defending the right of a whistleblower to speak out at the same time they do their best to silence any opposition (+20/-10)

"The intended result," replies grimlog, "is to drive greater engagement between the site and readers. The user voting system is the metric by which the editors quantify 'engagement'. And going by the number of votes, users are engaged" (+14/-0). User astie doesn't buy it: "unpopular views, even if expressed logically and reasonably, are quickly lost and the debate becomes substantially one-sided." Users will self-censor unpopular opinions, reducing engagement (+12/-2). Both grimlog and astie propose that unpopularity should not be sufficient to hide comments.

Concerns about viewpoints being hidden appear to be well founded. In a random sample of Snowden discussion comments on *Ars*, I found just four that were critical of Snowden: all of which were hidden due to low overall ratings. Yet all four had received up-votes, with ratings of +8/-164, +5/-25, +3/-23 and +13/-38 (this last is the comment in which Sobad challenges the use of down-votes). For a casual reader, these viewpoints can vanish, visible only when quoted in comments that refute them.

These down-voted comments on *Ars Technica* were initially missing from my analysis. *Ars* hides sufficiently down-voted comments entirely from readers who are not logged in. This led me to the erroneous conclusion that virtually no critics of Snowden received any up-votes. Finding quotes from comments that did not exist, I figured the

missing comments had been removed by moderators. I discovered them only by chance when I happened to log in. As Greg Elmer (2015) explains, most data obtained from online discussions is necessarily obtained from a first-person perspective by researchers or by web scraping software to which they delegate. Sites customize the data to the reader, in this case more dramatically than by sorting or collapsing comments. The significance of ratings is inescapable.

Nevertheless, for most discussions the comment system itself is a side-issue. But sometimes the system itself is the topic. When a CBC article announced a new "comment space" in 2013, one of the top concerns in the over 2,000 comments was removal of the capability to vote comments down (CBC, 2013). I have witnessed similar discussions elsewhere. While seldom discussed, this evidence suggests discussion participants are quite conscious of the significance of ratings and their influence on the discussion space.

Key Case Features

Reader comments seem to be written for an unknown public of readers, raters and other commenters. Replies often bring together contrasting viewpoints. Despite the diversity within each discussion in my study, is there a practice or phenomenon that characterizes and unites them?

Below are brief summaries of key distinguishing features of cases in my study. These disregard the details and fragmentation of the reader comment discussions, instead focusing on a few notable features. One theme that emerges is that many discussions do something that the primary article does not. The first summaries are for the Swartz cases:

- Lessig argues that Swartz's case should not be particularized as the unique story of an individual, but instead understood as symptomatic of a larger problem of Unequal Justice. Commenters embrace this interpretation, comparing Swartz to other principled protesters and interpreting his actions in as part of a larger Democratic Struggle.
- Just as Lessig had warned, the *New York Times* article particularizes Swartz's

story as that of a Troubled Genius. Commenters quietly reject this frame, reframing the story instead as an instance of Unequal Justice. Although the article provides minimal detail about Swartz's motivations, commenters engage in debate about the substance of his activism for open access to scholarly research.

- *CNN* also treats Swartz as a Troubled Genius. Commenters accept the article's focus on depression, but situate it in a broader context. They fill gaps in the story, engage with his activism and connect his depression to the struggles of others.
- The *Slashdot* discussion is dominated by digressions on several topics, possibly due to the hierarchical thread structure. Compared to other discussions, the tone is exceedingly rational (dealing with questions like the nature of freedom after death) to the point of callousness.
- Despite broad agreement with the *Ars Technica* article's criticism of Swartz's prosecution as an Institutional Failure, commenters there engage in substantial debate about his cause of open access. With the linear forum-like discussion format, topics shifts from one to another over time.
- The first *Volokh Conspiracy* article asks and answers questions of law. Top comments accept the frame of the question, and debate about the answers; only a few try to put Swartz's action in a larger context of motivations and values.
- The second *Volokh Conspiracy* article addresses questions of value; comments accordingly engage in a vigorous discussion. They express judgments of Swartz and the ethics of the prosecution, and debate whether his actions constituted legitimate civil disobedience.

Below are brief descriptions of distinguishing features of the Snowden cases:

- The frame of the *CNN* article is journalistic Objectivity and it makes no value judgments about the story. Top comments respond with judgment of his actions and solidarity with him.
- The *Guardian* article embraces Advocacy, presenting Snowden's own words. In response, commenters eschew debate for expressions of solidarity with him and

with one another.

- The *New York Times* article takes an Objective stance, avoiding Snowden's motivations and the substance of his leaks, but these are the center of the debate in the comments. The *Times* editors respond by picking comments that balance opposing views, possibly extending journalistic balance to the comment section.
- *Slashdot* commenters take their support for Snowden for granted. Rather than debating the ethics of his actions, they proceed to digress and analyze the causes of the problems he revealed and exploring implications and possible remedies.
- Like the article to which they respond, most *Ars Technica* comments support Snowden, but there is more debate about this and comments fail to develop the implications to the extent that *Slashdot* commenters do. Discussion is strongly influenced by Snowden critics to whom commenters respond and against whom they argue.

In general, comments bring to these stories things that the articles do not. In most cases they implicitly reject or negotiate the framing of the primary article, embedding the story in a larger context than it presents, whether that means linking Swartz to cases of depression in their everyday lives or injecting debates about values where articles avoided them. They call on frames like Unequal Justice, Democratic Struggle and Democratic Legitimacy even though most articles do not. And where most articles take objective stances, comments almost invariably judge.

Authentic Commenters, Objective Journalists

The subjectivity of judgment in reader comments stands in stark contrast to the objective stance of many of the articles to which they respond. What is the relationship between the objectivity of the articles on the one hand, and the subjectivity and judgment of the comments on the other?

Primary articles are powerful influences on the discussions that follow. *CNN* focuses on Swartz's depression; so do many of the comments. The two *Volokh*

Conspiracy articles emphasize law and values respectively; the comments follow suit. The *Guardian* presents Snowden as a historic figure; the comments hail him as a hero. The *New York Times* privileges negative official reactions to Snowden's actions, attracting more critical comments than articles on other sites.

The framing of the primary article is often spurned by commenters. *Times* and *CNN* commenters, for example, do not share the just-the-facts Objective frames of those sites' articles on Snowden: they raise questions of Democratic Legitimacy, of morality, of substantive policy issues. Few commenters responding to the *Times* and *CNN* articles about Swartz take up the Troubled Genius frame and its particularization of his story. Almost no top comments on either story adopt a disinterested stance. In short, comments are generally more subjective, more judgmental, and in many cases more radical than the articles to which they respond.

As described by Goffman (1974), frames are mental structures used to organize social experience, shaping the expectations of participants. Framing encompasses facts, claims and arguments; it also creates expectations about tone, affect and performance. The framing of a news article is not the same as the framing of its comments. Just as the audience of a theater performance operate according to different expectations than do the actors (one of Goffman's examples), so do commenters follow different norms and rules than bloggers and journalists.

News is usually framed by the standards and practices that dominate the journalism field. Among these is an ideal of objectivity according to which journalists present themselves as neutral, unbiased observers aloof from what they report on: neither participants in dialog nor judges of competing views. Of course journalists have personal opinions, but they are expected to exclude these from their work. While not all journalism adheres to this standard, deviations are expected to be clearly quarantined from ordinary reporting, indicated by placement and labels (op-ed articles, opinion pieces, individual columns). Violations of this norm are criticized; I have frequently encountered reader comments taking aim at journalists for perceived lapses in objectivity. Objectivity is also influential for many bloggers who see their work as a kind of journalism. With two exceptions, the *Guardian* article by Greenwald (who was subsequently criticized) and Lessig's reaction to Swartz's death, the news articles and

blog posts I examine here all exhibit the hallmarks of objectivity.

As scholars have detailed (Hackett & Zhao, 1998), the ideal of objectivity is unachievable in theory and problematic in application. It can justify flawed practices, such as presenting false balance between two opposing sides (regardless of whether they are two in number of equivalent in significance); privileging elite actors whose interested statements are masked by the neutrality of reports; and excluding context and critical analysis. In the United States, where the Fairness Doctrine once enforced balance on television, an expectation of objectivity has become entrenched among journalists and the public; an objective presentation is seen to distinguish legitimate journalism from propaganda. But far from being neutral, dogmatic adherence to objectivity tends to privilege the interests of elites and the professional status of journalists over the role of journalism as a basis for democratic discourse.

Regardless, journalists are expected to perform objectivity even if they cannot achieve it: to the extent that they hold personal views, they are expected not to express them sincerely. This is not to say they are insincere: like the actors in a play, their objective stance is understood as performance. Unlike Arendt's public individual whose action arises from the human capacity for originality and reveals who she is, the journalist is expected to suppress this potential, aspiring instead to a universal norm. Even when a journalist fails, the performance itself forecloses on the possibility of the glory and expression that Arendt describes. The judgment of journalists is expected to be of the rational sort.

Commenters are not expected to be objective. Instead, three expectations guide nearly all reader comment discussions: dialog, sincerity and authenticity. A comment responds to something that has already been said, and it is expected to express the honest thoughts, reactions, or experiences of the commenter. For my purposes, satire and sarcasm are sincere so long as they proceed from and reflect genuine sentiments rather than being employed to mislead (though such usages can cause confusion; according to Poe's Law (Poe, 2005), authorial intent is difficult to assess online).

Most comments manifest the expectation of dialog. In the discussions I examined, nearly all are responses to other comments or the primary article. For dialog to take place, there must exist some agreement among participants with regards to what

is being discussed and how, setting bounds on what is worth saying and why. Reader comments respond to an article, which thus has the first opportunity to frame discussion, addressing some matters and silently excluding others.

The expectations of sincerity and authenticity are highlighted by bad actors: spammers, trolls, sock puppets and astroturfers. Spam rejects both authenticity and dialog, being only about itself ("I earned $20,000 working from home!"). Alternatively, it can feign dialog but be insincere; "Great post!" is a typical example. Trolls aim to increase dialog through insincere expressions designed to provoke reaction. It is the lack of sincerity that makes a troll: someone who wrote a similar comment out of personal conviction might be disruptive, but is not a troll. Sock puppetry is the practice of commenting from multiple user accounts, usually in order to give the impression that others agree with what one has to say. It thus manages to be dishonest about identity among mutually anonymous participants. Sock puppets are usually on-topic, but their dialog is inauthentic because they mislead about who they are. Similarly, astroturfers pretend to represent themselves when in fact they speak for third parties.

Commenters are free to say things that journalists may not. Indeed, they are encouraged to do so: to express their thoughts ("Share your thoughts," prompts the *Times*), to vote up or down, to agree or disagree. It is not surprising that commenters express radical views not present in the articles to which they respond. Every article I looked at provoked highly-rated responses more radical than itself. In contrast, the articles themselves are consistently moderate. The *Times* echoes the government line. *CNN* provides biography. Even Glenn Greenwald in the *Guardian* avoids too much personal entanglement by giving the floor to Snowden, who avoids radical claims. With the exception of Greenwald's personal views, these are all consistent with norms of objectivity.

One imagines that the constraints imposed on journalists may well produce a hunger for judgment in comments. Like the performers and audience in the theater, the expectations of journalists and commenters can be seen as different but complementary. The judgment and subjectivity of commenters is invited to fill the void left by the article.

How do journalists cope with radicalism in the comment discussions they foster and maintain? They are in a privileged position to shape discussion: by direct

participation, through editor's picks, through technical design and by defining social norms. Although those who make such choices are expected not to take sides, they must unavoidably choose to promote some views over others. Many of the most popular of these views (as represented by reader ratings) are anything but objective. While journalists cannot sincerely express their views, their readership can and will. The journalist is put in the position of enabling (or promoting) what she cannot say herself.

Comments offer an escape from the straitjacket of conventional objectivity. Editors for the *Times* story on Snowden appear to address the problem by balancing highly-rated opposing views. Other journalists must make their own accommodations. On sites where journalists do not pick comments, writing an article lays the groundwork for (often predictable) alternative responses. For articles where such views are not desired, commenting can be turned off. When a choice can be made, it must be made. Whatever their attempts to sustain objectivity, their practice inevitably produces a record that is opinionated, passionate and engaged. Perhaps comments, like op-ed pages, are a safety valve that allows preserves the appearance of objectivity by allowing opinion elsewhere. Or perhaps it draws journalists in as accomplices, implicating them in practices that clash with their professional norms.

Judging Legitimacy

Commenters in these discussions make judgments about the actions of Snowden, Swartz, and other actors. Was Snowden engaged in legitimate civil disobedience? Was Swartz a thief? The answers demand political judgment, not merely the rational assignment of facts to defined categories. When the facts and categories are given, there is only one logical conclusion. Arendt argues that politics must be open-ended, preserving the potential of human freedom to reveal something new (Zerilli, 2005, p. 144). Categories must be open to definition or challenge. For her, this non-cognitive aspect of judgment is political, a manifestation of a plurality of individuals exercising their freedom.

Nowhere in my cases is this construction of categories more apparent than when participants in discussion debate matters of legitimacy: of the actions of government and of Swartz and Snowden themselves. Making judgments about these cases entails

deciding what counts as legitimate and why, implicating values as well as logic. This is perhaps best illustrated through two related themes in the comments, courage and civil disobedience.

Courage appears again and again in discussions about both Snowden and Swartz, as a measure of character and of the rightness of a cause. Numerous commenters hail Snowden's courage for acting against the most powerful government in the world. He knew that his freedom or even his life might be forfeit: concerns certainly not allayed when U.S. officials and lawmakers later said he should be tried for treason or (supposedly jokingly) killed (Everett, 2013; Sasso, 2013).

Snowden is the "bravest person in the world," writes anagama in the *Guardian* discussion (+1110). "I admire Mr. Snowden's courage . . . He will likely go to prison for many years," writes Barry in the *Times* comments (+242). "This guy is brave. He's at risk of death by polonium poisining or worse" thinks Super America POWER on *CNN* (+187/-9). *Times* commenter mje writes, "By making the decision to reveal himself he has shown incredible integrity and bravery. . . . he has taken any pretense of the moral high ground from Obama and the government. Sadly his reward will likely be many years in prison or worse" (+210). Snowden's is a masculine courage. "This guy is a hero. He has an epic set of balls," comments Rindan on *Ars* (+174/-12). On *Slashdot*, scottbomb writes, "This dude has balls of steel" (+4 Informative). Wondering why actions like those of Snowden are rare, Taco Cowboy asks, "Have we, the Americans, become pussies?" (+5 Insightful).

Praise for courage easily slides into talk of martyrdom. The very first comment of the *Slashdot* discussion, by Jah-Wren Ryel reads, "This man may well be our Jesus. The government is going to crucify him in their fury" (+5 Interesting). Taco Cowboy writes, "I cringe because of that 'I can't do nothing' feeling that is being felt by so many people today." "Why should Mr. Snowden become the sacrificial lamb?" he asks, arguing that it is the duty of Americans "to take back my government from those motherfucking tyrants," the "fuckers in Washington" who "fuck my Constitution" (+5 Insightful).

In response, symbolset confronts the logic of martyrdom directly, arguing that martyrs are needed to effect social change:

> . . . without the unpleasant consequences of martyrdom the standard social inertia cannot be overcome. It is the brutality of the oppression of the martyr that incites the rebellion, not his call for social change. . . . I'm not saying that he should be punished - only that he will. (+4 Insightful)

The logic of martyrdom suffuses the story of Swartz also. As described earlier, some critics accuse him of cowardice for resisting punishment or for committing suicide. His friend danah boyd (boyd, 2013) worries that Swartz's life is being reduced to one dimension. She writes on her blog, "Aaron will be turned into a martyr, an abstraction of a geek activist." Instead, "we need to look for an approach to change-making that doesn't result in brilliant people being held up as examples so that they can be tormented by power." Matt Stoller, another friend of Swartz who blogged about his death on *Naked Capitalism* (Stoller, 2013), responds in boyd's comments:

> But Aaron is a martyr, the state made sure of that when it destroyed him. And the state made sure it was an us versus them battle when it began attacking "us". I loved Aaron. But he would recognize that this is about power. He would also recognize that you've asked for people to disarm in the face of what just happened. No, no, and no. (boyd, 2013)

What sybolset, boyd, and Stoller all recognize is that in the narrative of martyrdom, suffering legitimates action. The tension between principle and punishment can tear people apart: "If you have strong moral convictions but not the will to expose yourself to punishment," symbolset writes, "you should avoid this situation because the internal conflict between your will to do the right thing and your fear of punishment can drive you insane." Praise for Snowden's bravery adheres to the same principle: it is his courage in the face of likely suffering that makes him a hero.

A martyr is an Arendtian hero (Arendt, 1959): like Achilles, a hero in an unjust war, whose deeds matter more than the cause for which he fights. He reveals who he is —not to himself, but to the witnesses to his deeds, who remember and tell the story. The commenters who praise Snowden's courage in the *Guardian*, who address their thanks to him, not to the paper, are witnesses playing their part in his performance. Their judgment defines the story's meaning. They are not simply voting against his critics, nor are they following up his action with activism of their own: they are memorializing his deed, and the voters who vote their comments up are doing likewise.

Suffering thus becomes a criterion for judging the legitimacy of political action: or indeed whether given action qualifies as political at all. The legitimacy of civil disobedience is a theme in online discussions about many instances of activism, including the Occupy movement, Wikileaks, Edward Snowden, and Aaron Swartz.

Swartz himself raised the issue in his *Guerilla Open Access Manifesto*, where he argued that scholars and others should engage in civil disobedience by distributing journal articles regardless of copyright restrictions. This theme was picked up in many discussions about his death, though it was relatively scarce the day the news broke.

In the *CNN* discussion, Bill criticizes Swartz: "there have been plenty of people who aided the advancement of technology without breaking the law" (+47/-18). Fatemeh Khatibloo replies, "when our laws cannot keep up with the pace of technology . . . we need people who are willing to break them" (+210/-9). User handleym invokes Democratic Struggle in response to Bill:

> So you think Gandhi should never have broken the law? Likewise for Rosa Parks? Likewise for the Stonewall demonstrators? Likewise for Mohamed Bouazizi? In the face of evil laws, resistance is the moral response. . . . We are talking about work that was done by public universities, all paid for by US taxpayers . . . but which is kept locked up by parasites like Elsevier which contribute nothing to the actual creation of this knowledge. (+88/-2)

Paul Shuster on Lessig's blog asserts that Swartz's action was civil disobedience (+34/0). A comment by Bob Younger in the *Times* discussion about Swartz's death is an NYT pick, though not a top comment by ratings:

> we have 'protesters' such as Assange and others who conduct illegal activities, wanting to attract attention to what they consider wrong laws, but never anticipate having to pay the price for those activities. . . . this generation of 'protesters' are more like the thief who breaks into a house . . . He does not plan for nor expect to have the laws regarding breaking, entering, and theft repealed. They do not want to nor plan to engage with the court system in order to expose the incorrectness of those laws. Civil disobedience is a strategy for changing laws; but the legal consequences are an integral element of its success. (+16, staff pick)

The only other discussions of the issue among top posts are in a vigorous argument on the *Volokh Conspiracy*. In the first discussion there is one comment in the

sample endorsing the "accept the consequences" position (+17/-4); there is no high-scoring rebuttal in the thread that follows. In the second discussion, Orin Kerr comments:

> We live in a system in which we are all equal and all get to choose the laws. Martin Luther King wasn't trying to circumvent the democratic process; he was trying to shine a light on a problem to change public opinion so that the democratically-elected laws would change. Swartz's manifesto expressly rejects King's approach as insufficient, and instead it tries to make the change he wants a fait accompli. In the South in the 1960s, the side that was trying to circumvent the democratic process by making its preferred world a reality by working outside law was not MLK's. (+10/-4)

User markregan attracts a higher rating when he responds, "To say 'we live in a system in which we are all equal and all get to choose the laws' sounds good, but the maldistribution in this country of political and economic power makes that more of an aspiration than a consistent reality" (+21/-1). This comment does not address the question of whether Swartz's actions in particular should qualify. That discussion appears later, when Kerr repeats his argument about democracy. Justin Levine responds,

> a growing number of people no longer have faith in the democratic process when it comes to intellectual property issues. How can you expect them to when they are increasingly governed by a set of international treaties that are negotiated in secrecy? When you and the lobbying power of certain special interests that have already warped copyright law beyond anything morally justifiable (+13/-1)

Kerr replies that the democratic process does indeed work, as demonstrated by the protests against SOPA and Swartz's part in them (+16/-2).

In a later exchange, Anon Y. Mous suggests that the Underground Railroad presents another model for civil disobedience—one that does not involve getting caught or accepting responsibility (+21/-2). Scott Harrison defines the category differently, arguing that this was a legitimate tactic, but was "closer to insurrection than to civil disobedience" (+23/-0).

On 18 January, *Slashdot* posted an article titled, "Hacktivism: Civil Disobedience Or Cyber Crime?", citing Bradley Manning and Aaron Swartz as examples (Soulskill, 2013b). The first of 243 comments, by alen, presents the civil rights movement as the

standard for civil disobedience. He emphasizes that “MLK and friends went to jail”—they did not simply break the law (+5). A week later, in the *Slashdot* article “Have Questions For MIT’s Aaron Swartz Review?” (Soulskill, 2013c), GovCheese repeats the point with a quote from Martin Luther King, Junior’s “Letter from a Birmingham Jail”: “One who breaks an unjust law must do so openly, lovingly, and with a willingness to accept the penalty” (+5 Insightful). Millennium repeats the point later in the discussion: “The real protest in civil disobedience starts when you pay the price, not when you do the deed. This is what gets the dialogue started, this is how you draw sympathy to your cause” (+4).

In a top comment on the *Times* discussion about Snowden, Khal Spencer also quotes from Martin Luther King’s “Letter from Birmingham Jail”:

> an individual who breaks a law that conscience tells him is unjust, and willingly accepts the penalty by staying in jail in order to arouse the conscience of the community over its injustice, is in reality expressing the very highest respect for the law. (+314)

The comment (and therefore its rating) is ambiguous, potentially supporting opposed arguments: that Snowden was engaged in civil disobedience, or that his action was illegitimate because he did not accept the penalty. Elsewhere in the discussion, Khal Spencer makes his support for Snowden clear. Of the seven responses to this comment, only one argues that Snowden was right to flee, while five argue that his flight to Hong Kong demonstrates his contempt for the law. This Questionable Character argument that Snowden’s flight undermines the legitimacy of his action is pervasive in criticisms of him.

The Birmingham Jail argument is powerful. Here and elsewhere I have seen the civil rights movement cited as the gold standard for civil disobedience and for legitimate law–breaking dissent. Judgments of legitimacy in the Swartz and Snowden cases are often grounded in an ideal of character as courage, willingness to take risks and accept consequences, and character as the basis on which moral judgments are reached. Stripped of politics, this understanding of civil disobedience is consistent with the logic of courage, martyrdom and sacrifice. Even in disagreement, participants in discussion accept the relevance of these categories (and others such as theft and treason).

Yet despite this appearance, the categories are not stable. In the many discussions about Aaron Swartz on *Slashdot* and *Ars Technica*, I found that the "accept the consequences" argument attracted more support than those who supported Swartz's claim to civil disobedience. In the case of Edward Snowden, however, I have observed the opposite trend: those who condemn his flight to Russia are in a minority, criticized by others who say that accepting punishment would be pointless, while the NSA story is what matters. Perhaps this reflects shifts in attitudes (Bradley Manning is often held up as an example of why submitting to punishment would achieve nothing) or a tendency to rationalize or differences between the particulars of the two cases. Without disagreeing about the material facts of each case, participants nonetheless differ in judgment about what these facts mean, which categories are most relevant, and how the categories themselves are defined. No matter how purely rational argument may appear, participants and readers are confronted with plurality; they deal with it by making judgments.

Judgment and Space

News articles written in accord with journalistic objectivity attempt to stand nowhere. Almost every comment responds by standing somewhere. They judge. In my study, comments place articles in context. They refuse to allow Swartz's story to be particularized as that of a Troubled Genius. Invoking Unequal Justice, they identify him as one of "us"; invoking Democratic Struggle they locate him in tradition of political resistance. Commenters on Snowden express Solidarity and debate questions of Democratic Legitimacy. They identify with the judgments they make. "Thank you," they say to Snowden; addressing him in the second person they implicate themselves. They take a stand.

Zerilli (2005) explains how judgment anchors us in a shared world: Arendt rejects objectivity as a chimera that takes us out of the world. We achieve a shared world not by striving for objectivity or erasing difference, but through a plurality of perspectives. Judgment in this sense is not merely a rational ("cognitive") calculation: for if it were, our choices would be the necessary products of reason, assigning phenomena to existing categories, rather than the unpredictable outcome of exercising our freedom.

For Arendt, political judgment is instead about acknowledging the plurality of the perspectives of others. “The public realm is constituted by the critics and the spectators,” she writes (1982, p. 63). We are irreducibly plural, but that condition alone is not enough: we must act, exercising our freedom and performing our uniqueness in the eyes of others. Their witness is necessary because we cannot see ourselves, the scope of our relationships, or the effects of our actions on others who act in turn, in a cascade that is unpredictable and ongoing. Witnesses acknowledge our action: doing so, they give it meaning. Acknowledgement is crucial: regardless of whether we agree with it or not, for action or speech to count as political it must be recognized as such. Due to human plurality, there will always be fundamental disagreements that cannot be resolved by understanding. The goal of politics should not be consensus, but acknowledgement that a plurality of perspectives are legitimate. Imagination allows us to step into the shoes of others enough to understand and recognize. Through this process we construct a shared world. Zerilli describes Arendt’s view:

> Arguments are valuable not when they produce agreement—though they may well do so—but when they enable us to see from standpoints other than our own and deepen our sense of what is shared or real. This shared sense of worldly reality is the condition of anything we could call communicable and valid, and it is unthinkable apart from the plurality of viewpoints from which an object or event is seen. (Zerilli, 2005, p. 140)

Beiner (1983, p. 156) places judgment at the heart of human life and politics. “Judgment is . . . the constitutive medium of political life and political discourse, the medium of politics.”

Comments respond to action in a primary article. The article itself may constitute action, like the advocacy of Lessig or Greenwald. In such instances authenticity and sincerity are important. Consistent with Arendt’s requirement that action should be non-instrumental, action is undermined if the author is pursuing an ulterior goal. Journalism adhering to norms of objectivity lacks the freedom necessary for action; it more often describes the action of others, as in the case of the *CNN* and *Times* articles I studied.

Commenters judge actions. With them come an army of raters whose votes validate the judgments of commenters. While I may reject the perspective of a comment, the votes push me to recognize its legitimacy. The judgments of comments and voters

make politics: for they assign meaning to actions, and it is meaning that allows actions to become political.

Participants in the *Volokh Conspiracy* discussions demonstrate judgment of this sort when they debate which metaphors are appropriate to describe Swartz's actions. Was circumventing blocks on MIT's network like breaking the lock on a door? The law provides categories for judgment, but the use of metaphor by commenters illustrates that they are constructing categories as much as applying them, while disagreement suggests that this is not simply a matter of exercising reason.

The discussions of civil disobedience illustrate how commenters judge actors who risk, suffer and make sacrifices. To make a statement of suffering is the strategy of the civilly disobedient: but that statement counts for nothing unless it is witnessed and acknowledged. To argue that Swartz was not engaged in legitimate civil disobedience is to implicitly recognize the claim that he was: it enables the question to be a political one of values, rather than limiting it to a cognitive matter of satisfying a prior definition.

Authenticity and sincerity are thus crucial for reader comments. One has responsibility for one's judgments (Beiner, 1983). They reveal. To make a judgment is to say, "I stand here." To acknowledge a judgment to recognize both the space it occupies and the difference of the speaker who utters it. Neither the speaker nor the space she occupies is negated simply by virtue of being (or being thought to be) wrong (in terms of politics, ethics, values, or facts). Insincerity and inauthenticity, on the other hand, undermine plurality. An inauthentic claim is one made by nobody; an insincere one occupies no space. Whatever its other merits, even if it be true, an insincere or inauthentic assertion has no perspective, no place to stand. It can only demand recognition through subterfuge; its subterfuge perceived, it demands no recognition. The most devastating accusation that can be leveled at another commenter is not that she is wrong but that she is inauthentic, in effect, that she is not real. To call someone a shill is not to say that she is wrong: it is to unmake her speech and excise it from the discourse.

Plurality in reader comments occupies conceptual space; judgments define it. "Opening up the world in a political sense requires a public space defined by equality," writes Zerilli (2005, p. 146): an equality made possible by the acknowledgment of participants and their plurality, "counting one's interlocutors *as* interlocutors." Excepting

moderators and article authors, participants in reader comment discussions are generally treated equally. According to Gómez, Kaltenbrunner and López (2008), reputation and prior relationships have little impact on whether a *Slashdot* comment attracts replies. My study finds little evidence that commenters address one another with familiarity. All are equally strangers.

As with other social structures, equality at one level gives rise to differentiation at another. The equality of participants enables them to occupy and define space through the positions they take. But while participants may be equal in their plurality, the space they occupy is not. Attention and significance are given to positions rather than persons, but they are given unevenly. Claims are supported and opposed. Not all are supported equally, or acknowledged equally. Perspectives are included by acknowledgement, excluded in silence. This space is not undifferentiated. It has a topography.

Space is a recurring theme in theories of publics. Like Zerilli and Arendt, Warner (2002, p. 67) speaks of a "space of discourse," Habermas (1991) of a public *sphere*; his historical analysis describes private and public physical spaces in the home, salons and coffee houses. Arendt discusses the public space of the *polis*, to which the home is the private counterpart.

The defining feature of reader comment discussions is marginality, a spatial relation. They are at the edges of the page, below the fold (the bottom of the browser window, so that readers must scroll to see them), behind a link or a button. Comments exist in the space of another, a space commenters cannot control or plan for. Lacking control over space, they are at the mercy of time. High-volume discussions come and go in hours in response to articles that arrive and are then forgotten as attention shifts elsewhere. The intense surges of participation in these discussions are brief, outside the control of commenters who can only respond tactically, without control of discussion or their destinies. Comments possess a spatial logic. Their temporality is demanding but fragmented. Commenters repeat one another and digress. Threads of conversation are broken up; discussion as a whole lacks flow, anchored to the primary article rather than to the rolling present of what has gone before. Comments scatter around the article offering divergent opinions and claims. Where the article possesses unity, they are plural. They are simultaneous, distributed in space.

To read comments is to explore a space. Most comment systems present no single linear view of discussion, no standpoint from which everyone sees the same thing. The volume of comments is often such that no-one reads them all, instead sorting and filtering, expanding a thread here, a thread there. For de Certeau (1984, p. 87), reading is a tactical practice of interpretation and remembrance, playing on memory of past stories and experiences. In the multidimensional space of comments, like de Certeau's pedestrian, every reader chooses her own path.

Nor is there any reason to expect most readers to be members of a community. Some pop in briefly, read a scattering of comments, and move on. Meanwhile, the landscape changes. Within threads, new comments appear interlaced with older ones. Ratings increase over time, altering the contours of discussion. When comments are hierarchical or sortable, there is no cursor pointing to *now*, no point to which the discussion has arrived so far. In contrast with the temporal ordering of the typical discussion forum, which guides all readers through the same linear sequence of places (comments, pages, threads), comment discussions are spatial, characterized by a simultaneous plurality of paths and experiences.

Among the plurality of comments the article stands singular and alone, an island in a sea of comments. A margin is not a territory: it is an edge, a narrow band between the land and the sea. The article is marginal to the comments too. That sea of discourse verges on many articles. As commenters play on memories of past events and past stories, injecting new claims and new frames, they situate their articles also.

Reader comments and ratings thus alter the topography of the public space in which they take place. Unlike a Cartesian coordinate system (but like other spaces of meaning or social interaction) that space is not uniform. Its density varies. In the conceptual space of comments, some perspectives are given more prominence or weight than others. In the physical space, expanses of low-rated and little-seen comments abut dense nodes of reading, writing, rating and replying.

Chapter 10: Conclusion

In this thesis I explore several large-scale reader comment discussions, drawing on Warner's (2002) theory of publics and Arendt's (1959) theory of action to develop a description of what participants do and say. I argue that the defining feature of reader comments is their marginality relative to a primary article. This is evident in their genealogy: comments appear to have emerged as adjuncts to blog posts rather than as an evolution of existing independent discussion formats. (This is underscored by the design differences of the *Ars* system, which does appear to have developed from discussion forums.) This evolution is still underway: since I collected my data, the *Guardian* and *CNN* have revamped their comment systems, while the *New York Times* has literally moved comments to a margin that appears on demand.

As a result of this subordinate relationship, reader comments are subject to the same surges of attention as their primary article. Lessig's blog posts about Swartz illustrate this clearly: his first post, cited widely, attracted fourteen times as many comments as his second one a week later. It is also apparent in the huge volumes of comments for some news articles: over 6,000 for the *Guardian* artcicle about Snowden, nearly 14,000 for *CNN*'s coverage. Most of these comments were made within the first 24 hours.

When hundreds or thousands of commenters converge on a single discussion over a brief period of time, most dialog is effectively anonymous. There is little evidence in my cases that commenters are familiar with one another; they seem to be addressing an impersonal public audience. Comments are motivated most often by points of difference rather than agreement. Many top comments are responses to lower-rated comments with which they disagree. With so much dialog driven by disagreement, discussion tends to be fragmented, containing diverse arguments and points of view.

Most discussions on the sites in my study attract smaller volumes of comments than the cases I looked at, as the comment counts for contemporaneous stories in Appendix A indicate. Yet high comment volumes are a predicable occurrence, and sites must be equipped to cope with them. Several features of reader comment systems facilitate participation in comments when there are too many to read. While these features are not unique to reader comments, they are characteristic of them.

One of the most widespread is rating mechanisms. I intentionally chose discussion systems that included them. These provide a means of participation in addition to writing comments; they also help to differentiate comments and focus attention on those that readers have chosen to promote. Commenters appear to be conscious of ratings, in several instances discussing them explicitly or using them as evidence for claims about broader public opinion. Ratings are a well-used feature; in every discussion I looked at, the number of votes exceeded the number of comments, often by an order of magnitude.

Every system I looked at features a means to highlight top-rated comments, and all but one (*Ars Technica*) allows comments to be sorted by rating. Ratings follow a long-tail distribution, indicating a winner-takes all dynamic in which high ratings compound, suggesting that ratings succeed in attracting reader attention. Ratings thus perform a moderation function, promoting comments that raters perceive as better or more agreeable, and if down-votes are supported, demoting those that raters dislike or perceive as bad. This moderation function is important as intervention by human moderators appears to be rare on many of these sites. I found that, in general, top-rated comments are qualitatively different from randomly-chosen ones. While ratings produce many false negatives (comments I subjectively think might be deserving fail to get voted up), I did not find them producing many false positives (irrelevant or inappropriate comments in the data were almost never voted up).

Every site but *Ars* supports some kind of hierarchical reply structure, enabling replies to be shown close to their antecedents on the page. Those with full hierarchical threading appear to be particularly prone to multi-comment digressions from the topic of the article (as opposed to the lone off-topic comments present in all discussions). These digressions contribute to the fragmentation of discussion, yet they also permit topics and debates to be developed in detail. Comments on systems with only one level of reply hierarchy tend to focus more on the story covered in the primary article.

My method of analyzing exchanges takes advantage of the tension between the agreement represented by ratings, and reply patterns that bring together divergent points of view. While the long-tail distribution of ratings with no upper limit means ratings cannot be compared on a standard scale, pairs are comparable when they are close

together in space and time so that raters have had an opportunity to view and rate both. This permits me to compare support among raters for contrasting viewpoints.

Analysis using this technique produced consistent results. Where there were multiple such pairs dealing with the same issue within the same discussion, their relative ratings tended to be consistent. Moreover, many views showed consistency across discussions on different sites: support for Snowden, for example, was high in every discussion; comments expressing empathy scored highly almost everywhere. Exchange comparisons appear to be a reliable indicator. They also seem to be a valid one: despite the fragmentation of discussion, highly-rated comments deal with consistent themes across discussions, such as civil disobedience or criticism of partisanship in American politics, that appear to reflect widespread sentiment across discussions and sites.

What is the significance of these results? Reader comments are often criticized as irrational and uncivil. Participants seldom express consensus or appear to change their minds. Rationality and civility are valued less than dialog and authenticity (with its correlate sincerity). This does not seem to be a Habermassian sphere in which individuals bracket their differences in pursuit of convergence and compromise. It better resembles Arendt's (1959) theory of agonistic public action. Arendt argues that the role of political judgment is to produce recognition, not consensus, defining space and making action possible (Arendt, 1982; Zerilli, 2005). So long as commenters are perceived as authentic and their political claims as sincere, ratings and replies acknowledge them, granting legitimacy even when they are thought to be wrong. Participants express their plurality and acknowledge that of others.

In most social contexts, North Americans avoid political conflict. Despite strong views and political polarization, politics is not a topic for polite dinner table conversation among people with conflicting views. Politicians, officials, and citizens alike avoid political disagreement and conflict (Eliasoph, 1998). When political talk does arise, consensus and harmony are valued over critique and argument. In my experience, this also holds for most discussion forums. While political debate occurs, it is also constrained in order to preserve community. When I have seen intense disagreements emerge in forums or on social media, they tend to be destructive. There are very few venues where ordinary people come together to give free rein to expressing political

differences. In large-scale reader comment discussions, in contrast, such intense disagreements are routine.

Large-scale reader comment discussions thus present a rare public space in which strangers with opposing political views are thrown together in conflict where there is little expectation that opinions will be tempered. So long as commenters are arguing, rather than denouncing one another as illegitimate shills or trolls, they are necessarily revealing and acknowledging contrary positions. Sincere comments offer the opportunity to uncover what people really think, the assumptions they make and the world they perceive.

What is it that they reveal? Following Gamson (Gamson, 1992), I use frame analysis to draw out the expectations and assumptions that guide discussion, developing and refining frames based on my reading of the comments. Each frame implicitly embeds the world to which a comment responds, defining context for the story. The Troubled Genius frame about Swartz, for example, limits that context to Swartz himself, while the Solidarity frame identifies a commenter with Snowden's actions and principles. While these frames cannot capture the full diversity of discussions, they help to map them, highlighting lines of agreement and conflict, and revealing differences between comments and the articles to which they respond.

The world these frames describe is a very different one from what is presented in the mainstream news articles in my study. Where the *CNN* and *Times* articles place events and individuals in the context of official institutions, in comment discussions the soundness and legitimacy of those institutions is a key point of debate. Some of the highest-rated comments depict a United States where institutions are failing (and partisan politics doesn't work), where the government doles out justice unequally, and where people are struggling for democracy. These are seldom consensus views: but they arise repeatedly among top comments in nearly every discussion (top comments on *CNN*'s article about Swartz do not express the Unequal Justice frame). Their ratings and the debates around them reveal that these issues are recognized as shared objects of difference. Even those who believe that American democracy is just must acknowledge critics.

Top comments repeatedly disregard the limited frames of articles to which they

respond, placing stories in a larger context. On *CNN*, a top comment fills in gaps in the Swartz story, crucial context which the article does not provide. When the *New York Times* treats Swartz as an isolated case, commenters expand the debate to his ideals of open access. Commenters focus on the substance of Snowden's leaks when the *Times* ignores them. Even Lessig's appeal to Unequal Justice is placed in a broader Democratic Struggle frame by the top comment, which casts Swartz as the latest in a long line of activists for democracy and social justice.

Goffman (1974) describes frames as socially prescribed or expected guides to action. Individual comments express various of the frames I define. For these frames to be commensurable, they must share the context of a larger frame whose expectations shape them all. Comments are always written in response to something else: they all address the same article and story. More than this, the comment frames I developed each answer a question. For the Swartz story, frames such as Institutional Failure and Lawbreaker answer the question, "Who is to blame?" For the Snowden story, frames like Solidarity (with Snowden) and Renegade answer the question, "Where do you stand?" These framing questions call for commenters to judge. This expectation (frame) of judgment encompasses every discussion I looked at; in my experience it is common to most large-scale reader comment discussions.

Commenters exercise judgment and express values, unlike journalists, whose professional standards demand they perform objectivity. The *New York Times* staff may even extend these norms to comments by promoting and balancing opposing views about Snowden. Commenters express subjective, often passionate viewpoints that journalists upholding the standards of their profession cannot. They connect stories to their everyday concerns and experiences. When *CNN* and the *Times* treat Swartz's death as the individual tragedy of a Troubled Genius, commenters engage with it as an instance of depression, something they too have encountered in their personal lives.

While journalists control the physical space of the page, conceptually they are expected to occupy an objective no-place. Participants in discussion, in contrast, stake out positions by making public judgments with their votes and comments. For visitors to a site, lacking the control of space necessary to plan, their speech is opportunistic and tactical (de Certeau, 1984). Speaking before an imagined but real audience of strangers,

their multiplicity of responses is diverse and unpredictable, irreducible to a single narrative.

Arendt argues that the role of political judgment is to produce recognition, not consensus, defining space and making action possible (Zerilli, 2005). So long as the political claims of other commenters are perceived as sincere, ratings and responses acknowledge them as legitimately political. In this way, the judgments of commenters are part of and help define a space of discourse surrounding the story told in the article. While the article provides the impetus for comment dialog, the comments reveal a context into which the article fits. Collectively, the recognition and differences expressed in comments construct a shared world. The coinciding and diverging judgments they make give political meaning to the public actions of others, organizing a public "space of discourse" (Warner, 2002, p. 67) around the primary article.

The use of the term "space" is more than metaphorical. Incompatible frames are, as a mathematician might say, linearly independent. Frames like Renegade and Solidarity are opposed in discussions about Snowden, but they are not opposites that can be reduced to one another. Meeting at an origin point with the story told in the article, they stake out different territories. The stories in my study may appear to be simple conflicts between two sides; in fact, they are multidimensional. For some commenters, democracy is the issue, for others it is law, for still others it is challenges of mental health. The space revealed in comments does not necessarily change the minds of participants but it defines the boundaries of what is recognized as relevant and legitimate.

In some sense, the significance of reader comments inverts their marginality. On the one hand, their marginal relationship to a primary article is essential. Commenters cannot plan nor control this space, they can only respond tactically and contingently to the authors, publishers and editors who do. The primary article can exist without comments, but they cannot exist without it. On the other hand, where journalists stop short of making judgments, comment discussions develop meaning: the most important element of all. They place the article in context, binding it into its world. Without the prospect for meaning and judgment, the primary article would have no reason to exist. The world and its meanings represented in comments are primary; in that sense, it is the

article that is marginal.

The history of reader comments is marginal in another sense. They developed as supplements to blog posts and news articles: nice-to-have add-ons, but which can be safely removed if they cause problems. This may not always be the case. News outlets increasingly make use of comments in their coverage of stories. The *New York Times* takes their comments seriously. “A core group of commenters have helped to transform The New York Times for a digital era,” writes Bassey Etim (2015) in an article which profiles several “of the most popular commenters,” complete with portraits and interviews. Comments are cheap, with the potential to substitute technology for high-priced paid labor. They may yet shift from the margins of the news business to the core.

Any such shift is likely to change them. So long as reader comments are perceived as unimportant, they are unlikely to be vehicles for strategic action by publishers or participants. When their remarks are seen and remembered by few, commenters are most likely to express themselves authentically and with sincerity. When profits are unaffected, publishers can afford to allow any meanings to dominate. With impact on the world (or its perception) come incentives to game and manipulate discussions, for better and for worse. Along with them come the network dynamics that have led to centralization and monopoly in other areas of online communication. Reader comments may be more vulnerable than other formats (such as forums) because they are already designed to scale.

Regardless, reader comments will remain marginal in the sense that they are dependent on and respond to a primary article. For most participants, they will remain places for tactical action in a space controlled by others. That marginality will still have an impact on the dynamics of attention and on the framing of discussion. I have mentioned that the popularity and influence of *Slashdot* have declined over time. When I encountered former users who give their reason for leaving, one explanation came up repeatedly: they had moved on to other sites with more timely, accurate and complete coverage of technology news. There are still discussions on *Slashdot*; the comment system works as well as it ever did. But it alone was not enough to sustain the site, which today is a shadow of its former self. Marginality is not a weakness of reader comments; it is their strength. The relationship with a primary article cannot simply be

left behind.

The story I have told about reader comments in the twelve discussions and two stories in my study is only a small slice of the phenomenon seen through the lens of my own experience and subjectivity. The most glaring limitation of my approach is the absence of the experiences of participants. Most systems allow various views of the comments, so that each reader can make her own path through them. I have focused on top-rated comments, but while there is evidence that users concentrate on those, there is also evidence that some do not (e.g. impressions of empathy in discussion inconsistent with ratings, the fact that many low-rated comments attract replies and votes). Latour (2005) and de Certeau (1984) explain how the experiences of actors can only be captured by tracing their paths, not by looking at a map, as it were, as I have done. Ethnographic research is necessary to ground the practices of commenting, rating, reading and replying in the experience of ordinary people.

The discussions I studied concerned the actions of two middle class white men whose concerns they shared with people like themselves.[18] Few of the comments were as vile as some discussions I have seen. Where stories are about the oppression of social groups, such as women or minorities, hateful comments often follow. Such discussion excludes diversity rather than encouraging it.

I made a conscious choice to examine topics that demonstrate the potential of reader comments. I did not seek out sex biases: nevertheless, I found them. Snowden has “balls of steel”; alternatively, he ran to “hide behind communist skirts.” Swartz should “man up”. Manliness in the face of punishment and suffering is the measure of legitimate politics. The readership of two of the sites I studied is dominated by men (*Ars Technica* and *Slashdot*). Counting gendered names on the *Times*, which requests participants use real names, I found male names outnumbered female ones two to one. The adversarial dynamic of disagreement and opposition that drives replies favors traditionally male aggression and the dominance of privileged actors. These discussions bracket much, right down to the identities of participants, yet indications of exclusion break through.

I have said that comments define boundaries of legitimacy. This is a space of

18 Privacy, digital surveillance and access to knowledge are key political issues, but they are understandably not top-of-mind for people who encounter more immediate injustices in their daily lives. To be impacted directly by these issues is to some degree to be privileged.

exclusion as well as inclusion. The overwhelming support for Snowden in the *Guardian* and on *Ars Technica* illustrates how a dominant consensus can push dissenting views aside. The winner-takes-all dynamic of ratings can magnify biases, including biases against certain kinds of difference. In large discussions on major news sites, many comments are motivated by opposition. In my experience, readers who are content remain silent: comments are the province of the angry, the passionate the disaffected. They are a place of populist passion. Whether the topic is climate change, capitalism, immigration or taxes, comment discussions are likely to be full of discontent. They can foster celebration of Snowden's rebellion and misogynist anger alike. Hate and anger are inescapable due to the same dynamics that juxtapose radical interpretations with conventional journalism. In a hierarchical discussion the only escape from bad actors is good moderation.

Yet it is too easy to dismiss reader comments as uncouth and rife with prejudice. Frase (2014) describes of the behavior of misogynist right-wing trolls among video gamers as "the fascistic street-fighting tactics of the troll brigade" who shut down dissent. "Gaming doesn't have a problem; capitalism has a problem" he writes. Comments can promote bad behavior, but they may also be a valuable indicator of it. Bad comments are not simply the expressions of a few bad apples. Blaming the medium of reader comments for vile behavior points to the importance of good moderation and good technical design. I earlier mentioned how *Jezebel* came under attack from sexist trolls. *Jezebel* did not solve their problem by shutting down reader comments: they instead pressured Gawker Media, of which they are a part, to tweak the technical design so that effective moderation became practical (Coen, 2014). Displacing blame onto technology gives structural and cultural injustices a pass.

Comment systems can fail when poor technical design or moderation lead to destructive interactions among participants. I believe my cases show that such dysfunction is not inevitable. Contrary to the views of some, reader comments are not a failed format. The alternative to comments is to switch them off, leaving the authority of the owners and authors of a site unchallenged. This can be a reasonable response in particular circumstances, such as when there are limited resources for moderation or when discussions degenerate repeatedly. I do not think that it should be the general rule. The impulse to secure authority by turning off comments is not censorship, as some

extreme supporters of comments claim. But as a general principle it is an admission that technocratic discourse is preferable to public engagement. I agree with Arendt that it is not enough to satisfy social needs (assuming technocratic governance would actually do so): we must also express human freedom. Regardless of whether the Internet is the ideal place, if the comment systems we have are not doing that, then we need to make them better.

In my own mind, behind this research was a desire to discover how to listen to and learn from what people have to say in comments. There remain three important steps in doing this. The first is to compare these exceptional cases with instances where comment discussions are not so successful, where anger and hate dominate. The second is to get a grip on the significance of comment discussion. Can comments be used as reliable indicators of some aspect of public sentiment? Finally, and most importantly, I have only examined the discourse itself. Appreciating the role that reader comments play in the lives of commenters, raters and readers requires ethnographic research, and that I have not done. This is critically important for understanding the relationship between these three groups, for understanding what they do and for appreciating the effects.

Is Dean (2005) correct that comments only disperse dissent harmlessly? Are they the maggots beside the steak? Beiner's (1983, p. 166-7) discussion of judgment, dating before the popular Internet, offers an alternative explanation. “The contemporary political world,” he writes, “allows very little outlet for genuine political activity . . . attending to the faculty of judgment may be a way of recouping one's status as a citizen.” Perhaps political comments are less a cause of political disengagement than consequence of a lack of outlets.

For Arendt, Beiner (1982, p. 153) suggests, judgment represents a reservoir of hope. It allows us to “sustain ourselves in the present and retain hope for the future only by reflecting on the miraculousness of human freedom as instantiated in particular moments of the past” (Beiner, 1982, p. 154). I began with an excerpt from the *Guardian* discussion about Snowden, a fragment of thread by zchabj5, StrawBear, days of hope and LandOfConfusion that I stumbled upon while exploring my data. They share quiet, anonymous judgment, solidarity and hope:

> No one is going to read this comment, and in the greater scheme of this man's life, it may be nothing, futile, spitting in the wind. But thank you, for humbling the world's great power of the time, and in so doing annihilate its claims to love or defend freedom.
>
> I read it, but I'm quite unimportant. Though I do agree with it.
>
> We're all insignificant. That's our importance.
>
> It has gladdened my heart to see the post count soar on this thread.

Limitations and Future Research

The most significant limitation of my research is that my data is restricted to the comments themselves. I lack information about readers, writers, and raters, their experiences or the context in which they engage in discussion.

I lack demographic data beyond the general claims of the news sites I studied. I do not know whether participants are representative of article readers. Without this information, it is not possible to make claims about reader comments as indicators of public opinion or the views of any particular group. Nor is there data about demographic characteristics such as age or sex; the latter is important given the level of sexism in many online discussions.

The experiences of readers are also key. I apply methods for studying the discourse participants produce. To put the significance of that discourse in context, however, it is necessary to study what it means to those who produce it and those who read it. Comments themselves provide some circumstantial evidence about the thoughts and experiences of commenters. Raters, however, leave very little evidence behind, other than the implication that they share some affinity with the comments that they rate. Readers are entirely invisible. Other research, such as that by Kareklas et al. (2015), provides clues, but more needs to be done in this area.

Without ethnographic data, it is not possible to put reader comments into the broader context of everyday life in general and political talk in particular. Reader comments say things that are absent from articles to which they respond. Does this reflect or influence broader public opinion? Discussion data alone cannot answer this

question.

Several questions are particularly important to assess the significance of my findings: What is the relationship between article readers, comment readers, comment writers and comment raters? What proportion of each group engages in the other activities? Why do they participate? How representative are they of the readers of primary articles? How do participants perceive the audience or public of reader comment discussion? Do they see comments as representative of general public opinion? Are they? How much attention do readers and writers pay to ratings? Can the effect of attention surges driven by the primary article be quantified and qualified?

I have attempted to develop methods for taking advantage of the mass of reader comment data on the Internet. I think that ordinary people have important things to say about politics, that they need to be listened to and that reader comments appear to be one way to do so. I see my research as a foundation for future studies of discussion participants' views about specific topics and stories. The next step, in my view, the missing piece for this study and for future those potential studies, is ethnographic research into the experience of commenters themselves. It seems clear to me that large-scale reader comment discussions are significant, but without an analysis of experience it is not clear what that significance *is*.

References

Aharony, N. (2010). LIS blog comments: an exploratory analysis. *Libri*, *60*(1), 65–77.

Anderson, A. A., Brossard, D., Scheufele, D. A. & Xenos, M. A. (2012). Online talk: How exposure to disagreement in online comments affects beliefs in the promise of controversial science. *Citizen Voices: Performing Public Participation in Science and Environment Communication*, 121–136.

Anderson, N. (2013). Opening arguments in the court of public opinion after Aaron Swartz' death. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/2013/01/opening-arguments-in-the-trial-of-public-opinion-after-aaron-swartz-death/

Arendt, H. (1959). *The human condition*. New York: Doubleday Anchor.

Arendt, H. (1982). *Lectures on Kant's political philosophy*. R. Beiner (Ed.). Chicago: University of Chicago Press.

Arthur, C. (2014). The future for Microsoft, Android and Windows phone - analysis. *The Guardian*. Retrieved from http://www.theguardian.com/technology/2014/feb/11/microsoft-android-windows-phone

Balasubramanyan, R., Cohen, W. W., Pierce, D., & Redlawsk, D. P. (2011). What pushes their buttons?: Predicting comment polarity from the content of political blog posts. In *Proceedings of the workshop on languages in social media* (pp. 12–19). Association for Computational Linguistics.

Bakardjieva, M. (2008). Bulgarian online forums as carnival: Popular political forms and new media. In F. Sudweeks & C. Ess (Eds.), *Cultural attitudes towards technology and communication*. Australia: School of Information Technology, Murdoch University.

Bakker, T. P. (2013). Citizens as political participants: The myth of the active online audience?. *America*, *76*(3), 502-521.

Barabási, A.-L. (2003). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life*. New York: Plume.

Bardini, T. (2000). *Bootstrapping: Douglas Engelbart, coevolution, and the origins of*

personal computing. Stanford University Press.

Bassey, Etim. (2015). Meet some of our top commenters. *New York Times*. Retrieved from http://www.nytimes.com/interactive/2015/11/23/nytnow/23commenters.html

Beiner, R. (1983). *Political judgment.* Methuen: London.

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A., & Etling, B. (2013). Social mobilization and the networked public sphere: Mapping the SOPA-PIPA debate. *Berkman Center Research Publication*, (2013-16).

boyd, d. (2013). Processing the loss of Aaron Swartz. Blog. Retrieved from http://www.zephoria.org/thoughts/archives/2013/01/13/aaron-swartz.html

Brossard, D., & Scheufele, D. A. (2013). This story stinks. *New York Times*. Retrieved from http://www.nytimes.com/2013/03/03/opinion/sunday/this-story-stinks.html

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, *67*, 97–102.

Burke, K. (1969). *A rhetoric of motives*. Berkeley, CA: University of California.

CBC. (2013). CBC gets improved commenting space. Retrieved from http://www.cbc.ca/newsblogs/yourcommunity/2013/06/cbc-community-gets-improved-commenting-space.html

Chan, A. J. (2002). *Collaborative news networks: Distributed editing, collective action, and the construction of online news on slashdot.org* (PhD thesis). Massachusetts Institute of Technology.

Chen, Z., & Berger, J. (2013). When, why, and how controversy causes conversation. *Journal of Consumer Research*, *40*(3), 580-593.

Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2014). How community feedback shapes user behavior. *arXiv preprint arXiv:1405.1429*.

Chilton, P. (1987). Metaphor, euphemism and the militarization of language. *Current Researhc on Peace and Violence*, *10*(1), 7–19.

CNN. (n.d). Advertise. Retrieved from

http://www.cnn.com/services/advertise/audience_profile.html

Coen, J. (2014, August). What gawker media is doing about our rape gif problem. *Jezebel*. Retrieved from http://jezebel.com/what-gawker-media-is-doing-about-our-rape-gif-problem-1620742504

Coleman, G. (2013). *Coding freedom: The ethics and aesthetics of hacking*. Princeton, NJ: Princeton University Press.

Conspiracy, V. (n.d). Editorial independence. Retrieved from http://www.washingtonpost.com/news/volokh-conspiracy/about/

Crompton, T. (2010). Common cause: The case for working with our cultural values. WWF-UK.

Cunningham, B. (2004). In R. W. McChesney & B. Scott (Eds.), *Our unfree press: 100 years of radical media criticism* (pp. 287–300). New York: New Press.

Cyrus, F. (2013). UK spies continue 'quantum insert' attack via LinkedIn, Slashdot pages. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/2013/11/uk-spies-continue-quantum-insert-attack-via-linkedin-slashdot-pages/

Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., & Lee, L. (2009). How opinions are received by online communities: A case study on amazon.com helpfulness votes. In *Proceedings of the 18th international conference on world wide web* (pp. 141–150). ACM.

De Certeau, M. (1984). *The practice of everyday life*. Berkeley, CA: University of California Press.

Dean, J. (2002). The lure of the public. *Theory & Event*, *6*(3).

Dean, J. (2005). Communicative capitalism: Circulation and the foreclosure of politics. *Cultural Politics*, *1*(1), 51–74.

Diakopoulos, N., & Naaman, M. (2011, March). Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 133-142). ACM.

Domingo, D. (2008). Interactivity in the daily routines of online newsrooms: dealing with

an uncomfortable myth. *Journal of Computer-Mediated Communication*, *13*, 680–704. doi:doi:10.1111/j.1083-6101.2008.00415.x

Eliasoph, N. (1998). *Avoiding politics: How Americans produce apathy in everyday life*. Cambridge, UK: Cambridge University Press.

Elmer, G. (2015). Scraping the first person. In Elmer, G, Longlois, G. & Redden, J. (Eds.), *Compromised data: From social media to big data* (112–125). USA: Bloomsbury Publishing.

Ensmenger, N. (2010). *The computer boys take over: Computers, programmers, and the politics of technical expertise*. Cambridge, MA.

Everett, B. (2013). On Edward Snowden, senators talk treason. *Politico*. Retrieved from http://www.politico.com/story/2013/06/edward-snowden-senators-talk-treason-92547.html

Farhi, P. (2014). Washington Post, New York Times and Mozilla team up for new web site comment system. *Washington Post*. Retrieved from http://www.washingtonpost.com/lifestyle/style/washington-post-new-york-times-and-mozilla-team-up-for-new-web-site-comment-system/2014/06/19/fa836e90-f71e-11e3-8aa9-dad2ec039789_story.html

Feenberg, A. (1999). *Questining technology*. London: Routledge.

Fisher, K. (2000). Posting guidelines and general information. *Ars Technica*. Retrieved from http://arstechnica.com/staff/2000/01/postguide/

Fisher, K. (2013). The great Ars survey: The results are in. Retrieved from http://arstechnica.com/staff/2013/01/the-great-ars-survey-the-results-are-in/

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, *12*(2), 219–245.

Force, I. E. T. (n.d.). Getting started in the IETF. Retrieved from http://www.ietf.org/newcomers.html#participation

Frase, P. (2014, Sept.). Gamer's revanche. blog. Retrieved from http://www.peterfrase.com/2014/09/gamers-revanche/

Fraser, N. (1992). Rethinking the public sphere: A contribution to the critique of actual

existing democracy. In C. J. Calhoun (Ed.), *Habermas and the public sphere* (pp. 109–42). Cambridge, MA: MIT Press.

Freelon, D. (2013). Discourse architecture, ideology, and democratic norms in online political discussion. *New Media & Society*, 1461444813513259.

Frick, K., Guertler, D., & Gloor, P. A. (2013). Coolhunting for the world's thought leaders. *arXiv preprint arXiv:1308.1160*.

Gamson, W. A. (1992). *Talking politics*. US: Cambridge University Press.

Geuss, M. (2013). Whistleblower who exposed NSA mass-surveillance revealed by the Guardian. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/2013/06/whistleblower-who-exposed-nsa-mass-surveillance-revealed-by-the-guardian/

Godwin, M. (1994). Meme, counter-meme. *Wired*. Retrieved from http://archive.wired.com/wired/archive/2.10/godwin.if_pr.html

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. New York: Harper & Row.

Gonzalez-Bailon, S., Kaltenbrunner, A., & E Banchs, R. (2010). The structure of political discussion networks: A model for the analysis of online deliberation. *Journal of Information Technology*, *25*, 230–243.

Gómez, V., Kaltenbrunner, A., & López, V. (2008). Statistical analysis of the social network and discussion threads in Slashdot. In *Proceedings of the 17th international conference on world wide web* (pp. 645–654). ACM.

Greenwald, G. (2013a). NSA collecting phone records of millions of Verizon customers daily. *The Guardian*. Retrieved from http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order

Greenwald, G., MacAskill, E., & Poitras, L. (2013). Edward Snowden: The whistleblower behind the NSA surveillance revelations. *The Guardian*. Retrieved from http://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance

Guardian. (2009a). Community standards and participation guidelines. Retrieved from

http://www.theguardian.com/community-standards

Guardian. (2009b). Frequently asked questions about community on the Guardian website. Retrieved from http://www.theguardian.com/community-faqs

Guardian. (n.d.). Audience. Retrieved from http://advertising.theguardian.com/?tag=audience

Guess, A. (2008). Blogs and wikis and 3d, oh my! *Inside Higher Ed*. Retrieved from http://www.insidehighered.com/news/2008/05/09/blogs

Gumbrecht, M. (2004). Blogs as "protected space". In *WWW 2004 workshop on the weblogging ecosystem: Aggregation, analysis and dynamics* (Vol. 2004).

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT Press.

Hackett, R. A., & Zhao, Y. (1998). *Sustaining democracy?: Journalism and the politics of objectivity*. Broadview Press.

Halavais, A. M. C. (2001). *The Slashdot effect: Analysis of a large-scale public conversation on the world wide web* (PhD thesis). University of Washington.

Hall, S. (1978). The social production of news. In *Policing the crisis: mugging, the state, and law and order* (pp. 53–77). London: Macmillan.

Hermida, A., & Thurman, N. (2007). Comments please: How the British news media are struggling with user-generated content. *7th International Symposium on Online Journalism*.

Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. In *System sciences, 2004. proceedings of the 37th annual Hawaii international conference on* (pp. 11–pp). IEEE.

Hickman, L. A. (1990). *John Dewey's pragmatic technology*. Indiana University Press.

Hughes, E. (1984). *The sociological eye: Selected papers on work, self and the study of society*. New Brunswick, NJ: Transaction.

Jackson, H. J. (2001). *Marginalia: Readers writing in books*. New Haven, CT: Yale

University Press.

Jezebel. (2014, August). We have a rape gif problem and Gawker media won't do anything about it. *Jezebel*. Retrieved from http://jezebel.com/we-have-a-rape-gif-problem-and-gawker-media-wont-do-any-1619384265

Johnston, H. (2005). Comparative frame analysis. In H. Johnston & J. A. Noakes (Eds.), *Frames of protest: Social movements and the framing perspective* (pp. 237–260). Lanham, MD: Rowman & Littlefield.

Kaltenbrunner, A., Gomez, V., & Lopez, V. (2007). Description and prediction of Slashdot activity. In *Web conference, 2007. IA-WEB 2007. Latin American* (pp. 57–66). IEEE.

Kareklas, I., Muehling, D. D., & Weber, T. J. (2015). Reexamining health messages in the digital age: A fresh look at source credibility effects. *Journal of Advertising, 44*(2), 88-104. Retrieved from http://ssrn.com/abstract=2556998

Kehoe, A., & Gee, M. (2012). Reader comments as an aboutness indicator in online texts: Introducing the Birmingham blog corpus. *Studies in Variation, Contacts and Change in English*, *12*. Retrieved from http://www.helsinki.fi/varieng/series/volumes/12/kehoe_gee/

Kenney, K., Gorelik, A., & Mwangi, S. (2000). Interactive features of online newspapers. *First Monday*, *5*(1). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/720/629

Kerr, O. (2013a). The criminal charges against Aaron Swartz (part 1: the law). *The Volokh Conspiracy*. Retrieved from http://www.volokh.com/2013/01/14/aaron-swartz-charges/

Kerr, O. (2013b). The criminal charges against Aaron Swartz (part 2: prosecutorial discretion). *The Volokh Conspiracy*. Retrieved from http://www.volokh.com/2013/01/16/the-criminal-charges-against-aaron-swartz-part-2-prosecutorial-discretion/

Labarre, S. (2013). Why we're shutting off our comments. Retrieved from http://www.popsci.com/science/article/2013-09/why-were-shutting-our-comments

Laclau, E., & Mouffe, C. (1985). *Hegemony and socialist strategy*. London: Verso.

Lakoff, G. (2010). Why it matters how we frame the environment. *Environmental Communication*, *4*(1), 70–81.

Lampe, C. A. (2006). *Ratings use in an online discussion system: The Slashdot case* (PhD thesis). University of Michigan.

Lampe, C. A. C., Johnston, E., & Resnick, P. (2007). Follow the reader: Filtering comments on Slashdot. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1252–1263). New York: ACM.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford, U.K.: Oxford University Press.

Lau, A. (2015, September 9). Sue MacDonnell, Conservative riding association director, booted after 'inappropriate' Facebook posts. *Huffington Post*. http://www.huffingtonpost.ca/2015/09/09/sue-macdonell-conservative-facebook-quinte_n_8108178.html

Lee, E.-J. (2011). Perceived bias or biased perception? Effects of readers' comments, perceived public opinion and issue involvement on the perception of news slant. *Korean Journal of Journalism and Communication Studies*, 55(3), 179–198.

Lee, E.-J. (2012). That's Not the Way It Is: How User-Generated Comments on the News Affect Perceived Media Bias. *Journal of Computer-Mediated Communication*, *18*(1), 32-45.

Lee, T. B. (2013). Internet pioneer and information activist takes his own life. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/2013/01/internet-pioneer-and-information-activist-takes-his-own-life/

Lessig, L. (2013a). Prosecutor as bully. *Lessig v.2*. Retrieved from http://lessig.tumblr.com/post/40347463044/prosecutor-as-bully

Lessig, L. (2013b). A time for silence. *Lessig v.2*. Retrieved from http://lessig.tumblr.com/post/40845525507/a-time-for-silence

Lessig, L. (2008). *Remix: Making art and commerce thrive in the hybrid economy*. New York: Penguin.

Linebaugh, P., & Rediker, M. (2000). *The many-headed hydra: Sailors, slaves, commoners, and the hidden history of the revolutionary atlantic*. Beacon Press.

Loller, T. (2007). Newspapers debate online reader comments. *USA Today*. Retrieved from http://usatoday30.usatoday.com/tech/products/2007-04-26-393362505_x.htm

Lovell, B., & Lee, R. (2011). Perceptions about quality of care: Evidence from a Globe and Mail blog. *J Participat Med*, *3*, e21. Retrieved from http://www.jopm.org/evidence/research/2011/05/02/perceptions-about-quality-of-care-evidence-from-a-globe-and-mail-blog/

Mahoney, M. S. (2011). *The histories of computing*. (T. Haigh, Ed.). Cambridge, MA: Harvard University Press.

Malda, R. (2007). A brief history of Slashdot part 2, explosions. *Slashdot*. Retrieved from http://news.slashdot.org/story/07/10/10/1445216/a-brief-history-of-slashdot-part-2-explosions

Manosevitch, E., & Walker, D. (2009). Reader comments to online opinion journalism: A space of public deliberation. In *International symposium on online journalism* (Vol. 10, pp. 1–30).

Martinez, M. (2013). Internet prodigy, activist Aaron Swartz commits suicide. *CNN*. Retrieved from http://www.cnn.com/2013/01/12/us/new-york-reddit-founder-suicide

Mazzetti, M. (2013). Edward Snowden, ex-C.I.A. worker, says he disclosed U.S. surveillance. *New York Times*. Retrieved from http://www.nytimes.com/2013/06/10/us/former-cia-worker-says-he-leaked-surveillance-data.html

McKeon, D. (1997). Moderated newsgroups FAQ. Retrieved from http://pages.swcp.com/~dmckeon/mod-faq.html#Q2.3

Milberry, K. (2003). *Indymedia as a social movement?* (Master's thesis). University of Windsor.

Mishne, G., & Glance, N. (2006). Leave a reply: An analysis of weblog comments. In *Third annual workshop on the weblogging ecosystem*.

Mullin, J. (2013). NSA leaker ed snowden's life on ars technica. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/2013/06/nsa-leaker-ed-snowdens-life-on-ars-technica/

Munroe, R. (2008). Duty calls. Cartoon. Retrieved from https://xkcd.com/386/

Nardi, B. A., Schiano, D. J., & Gumbrecht, M. (2004). Blogging as social activity, or, would you let 900 million people read your diary? *Proceedings of the 2004 ACM conference on Computer supported cooperative work*.

Nielsen, C. E. (2012). Newspaper journalists support online comments. *Newspaper Research Journal*, *33*(1), 86.

Nielsen, C. E. (2014). Coproduction or cohabitation: Are anonymous online comments on newspaper websites shaping news content? *New Media & Society*, 16, 470–487. doi:10.1177/1461444813487958.

Noakes, J. A., & Johnston, H. (2005). Frames of protest: A road map to a perspective. (H. Johnston & J. A. Noakes, Eds.) *Frames of protest: Social movements and the framing perspective*, 1–29.

Oliver, P. E., & Johnston, H. (2005). What a good idea! Ideologies and frames in social movement research. In H. Johnston & J. A. Noakes (Eds.), *Frames of protest: Social movements and the framing perspective* (pp. 185–204). Lanham, MD: Rowman & Littlefield.

O'Brien, D. (2004). Notes on: Protecting your open discussion forum. Retrieved from http://www.oblomovka.com/entries/2004/07/29

Ó Baoill, A. (2000). Slashdot and the public sphere. *First Monday*, *5*(9).

Park, S., Ko, M., Kim, J., Liu, Y., & Song, J. (2011). The politics of comments: Predicting political orientation of news stories with commenters' sentiment patterns. In *Proceedings of the ACM 2011 conference on computer supported cooperative work* (pp. 113–122). ACM.

Peattie, L. (2001). Theorizing planning: Some comments on Flyvbjerg's rationality and power. *International Planning Studies*, *6*(3), 257–262.

Pinch, T. J., & Bijker, W. E. (2001). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In T. P. H. Wiebe E. Bijker & T. J. Pinch (Eds.), *The social construction of technological systems*. Cambridge, MA: MIT Press.

Poe, N. (2005). Poe's Law [Online forum comment]. Retrieved from

http://www.christianforums.com/t1962980-6/#post17606580

Polanyi, K. (2001). *The great transformation*. Boston: Beacon Press.

Poor, N. (2005). Mechanisms of an online public sphere: the website Slashdot. *Journal of Computer-Mediated Communication*, *10*(2).

Protess, B., & Silver-Greenberg, J. (2012). HSBC to pay $1.92 billion to settle charges of money laundering. *New York Times*. Retrieved from http://dealbook.nytimes.com/2012/12/10/hsbc-said-to-near-1-9-billion-settlement-over-money-laundering/

Public. (2013). In *Webster's Revised Unabridged Dictionary (1913 + 1828).* Retrieved from http://machaut.uchicago.edu/?resource=Webster%27s&word=public&use1913=on

Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., & Olmstead, K. (2010). Understanding the participatory news consumer. *Pew Internet and American Life Project*, *1*.

Remember Aaron Swartz. (2013). Remember Aaron Swartz. Retrieved from http://www.rememberaaronsw.com/

Ritholtz, Barry. (n.d.) The big picture. Retrieved from http://www.ritholtz.com/blog/

Robertson, A. (2014, August). Trolls drive Anita Sarkeesian out of her house to prove misogyny doesn't exist. *The Verge*. Retrieved from http://www.theverge.com/2014/8/27/6075179/anita-sarkeesian-says-she-was-driven-out-of-house-by-threats

Rogers, E. M. (1995). *Diffusion of innovations*. New York: Free Press.

Ruiz, C., Domingo, D., Micó, J. L., Díaz-Noci, J., Masip, P., & Meso, K. (2011). Public sphere 2.0? The democratic qualities of citizen debates in online newspapers. *The International Journal of Press/Politics.* doi:10.1177/1940161211415849.

samzenpus. (2013). NSA whistleblower outs himself. *Slashdot*. Retrieved from http://yro.slashdot.org/story/13/06/09/1958200/nsa-whistleblower-outs-himself

Sasso, B. (2013). Ex-NSA chief jokes about putting Edward Snowden on kill list. *The Hill*. Retrieved from http://thehill.com/policy/technology/326315-former-nsa-chief-jokes-about-putting-snowden-on-kill-list

Saunders, B. (2006). *Citizen media and investigative journalism* (Master's thesis). Queensland University of Technology, Brisbane, Australia.

Scheufele, D. A., Anderson, A. A., Brossard, D., Xenos, M. A., & Ladwig, P. (2013). Crude comments and concern: Online incivility's effect on risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*.

Schultz, T. (1999). Interactive options in online journalism: A content analysis of 100 U.S. newspapers. *Journal of Computer-Mediated Communication*, *5(1)*. doi:10.1111/j.1083-6101.1999.tb00331.x

Schwartz, J. (2013). Aaron Swartz, Internet activist, dies at 26. *New York Times*. Retrieved from http://www.nytimes.com/2013/01/13/technology/aaron-swartz-internet-activist-dies-at-26.html

Sipos, R., Ghosh, A., & Joachims, T. (2014). Was this review helpful to you? It depends! Context and voting patterns in online content. Retrieved from http://dx.doi.org/10.1145/2566486.2567998

Slashdot. (2013). Have a question for MIT's Aaron Swartz Review? Retrieved from http://yro.slashdot.org/story/13/01/25/1610202/have-questions-for-mits-aaron-swartz-review

Slashdot. (n.d.). Slashdot FAQ - comments and moderation. *Slashdot*. Retrieved from http://slashdot.org/faq/com-mod.shtml

Smith, M. (2013). NSA leaker comes forward, warns of agency's "existential threat". *CNN*. Retrieved from http://www.cnn.com/2013/06/09/politics/nsa-leak-identity/index.html

Snow, D. A., Rochford, J., E. Burke, Worden, S. K., & Benford, R. D. (1986). Frame alignment processes, micromobilization, and movement participation. In *American Sociological Review* (Vol. 51, pp. 464–481).

Söderberg, J. (2008). *Hacking capitalism: The free and open source software movement*. New York: Routledge.

Soulskill. (2013a). Aaron Swartz commits suicide. *Slashdot*. Retrieved from http://news.slashdot.org/story/13/01/12/1240255/aaron-swartz-commits-suicide

Soulskill. (2013b). Hacktivism: Civil disobedience or cyber crime? *Slashdot*. Retrieved

from http://yro.slashdot.org/story/13/01/18/1846206/hacktivism-civil-disobedience-or-cyber-crime

Soulskill. (2013c). Have questions for MIT's Aaron Swartz review? *Slashdot*. Retrieved from http://yro.slashdot.org/story/13/01/25/1610202/have-questions-for-mits-aaron-swartz-review

SteveTheNewbie. (2011, May 1). Re: Democracy. *Slashdot*. Retrieved from http://politics.slashdot.org/comments.pl?sid=2117934&cid=35991222

Stoller, M. (2013, January 14). Aaron Swartz's politics. *Naked Capitalism*. Retrieved from http://www.nakedcapitalism.com/2013/01/aaron-swartzs-politics.html

Swartz, A. (2006, September). Who writes wikipedia? *Raw thought*. Blog. Retrieved from http://www.aaronsw.com/weblog/whowriteswikipedia

Taibbi, M. (2012). Outrageous HSBC settlement proves the drug war is a joke. *Rolling Stone*. Retrieved from http://www.rollingstone.com/politics/blogs/taibblog/outrageous-hsbc-settlement-proves-the-drug-war-is-a-joke-20121213

Technica, A. (n.d.). About us. Retrieved from http://arstechnica.com/about-us/

Times, N. Y. (n.d.). Media kit. Retrieved from http://nytmediakit.com/online

timothy. (2008). Canada's proposed DMCA-style law draws fire. *Slashdot*. Retrieved from http://news.slashdot.org/story/08/06/12/1852215/canadas-proposed-dmca-style-law-draws-fire

Toepfl, F., & Piwoni, E. (2015). Public Spheres in Interaction: Comment Sections of News Websites as Counterpublic Spaces. *Journal of Communication*.

Tremonti, A. M. (2014, February 24). Should media sites stop feeding the trolls and abandon online comment sections? *CBC*. Retrieved from http://www.cbc.ca/thecurrent/episode/2014/02/24/should-media-sites-stop-feeding-the-trolls-and-abandon-online-comment-sections/

Tribble, E. B. (1993). *Margins and marginality: The printed page in early modern england*. Charlottesville, VA: University Press of Virginia.

Trice, M. (2011). How we comment on Web journalism: A case study in dialogue found in

news articles. In Drushel, B.E. & German, K. (Eds.), *The ethics of emerging media* (pp. 235-250). New York: Continuum, 2011.

Warner, M. (2002). *Publics and counterpublics*. New York: Zone Books.

Weber, P. (2014). Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments. *New Media & Society*, *16*(6), 941-957.

Weingarten, G. (2010). Gene Weingarten column mentions Lady Gaga. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2010/07/09/AR2010070904048.html

Westby, D. L. (2005). Strategic imperative, ideology, and frames. In H. Johnston & J. A. Noakes (Eds.), *Frames of protest: Social movements and the framing perspective* (pp. 217–236). Lanham, MD: Rowman & Littlefield.

Wright, S., & Street, J. (2007). Democracy, deliberation and design: The case of online discussion forums. *New Media and Society*, *5*(9), 849–869.

Xin, C., Glass, G., Feenberg, A., Bures, E., & Abrami, P. (2010). From active reading to active dialogue: An investigation of annotation-enhanced online discussion forum. In F. Pozzi & D. Persico (Eds.), (pp. 300–318). IGI Global. Retrieved from http://webmarginalia.net/wp-content/uploads/2013/12/CSCL_Marginalia_12Jan2010-4app.pdf

Zerilli, L. M. G. (2005). *Feminism and the abyss of freedom*. Chicago: University of Chicago Press.

Appendix A: Comparison to Same-Day Articles

The following are lists of news articles from around the same time as the articles I examined. They put the number of comments on the Swartz and Snowden articles in context, and illustrate the kinds of stories these sites cover. Each article title is preceded by a comment count in parentheses. Entries corresponding to study cases are italicized.

I have excluded articles for which comments were disabled. For some sites (*CNN*, the *Guardian*, the *New York Times*) there is a slim possibility that I have missed a few articles with comments. On these sites, comments are displayed using JavaScript. This tends to happen last when the page is loaded in the web browser. If the browser's connection is interrupted it might be displayed without comments. Similarly, if I do not wait long enough for the page to load I could conclude too soon that an article does not have comments enabled. I mention this less for fear that it affects my data than to alert other researchers to the risk.

Ars Technica

The *Ars Technica* story lists are based on Internet Archive snapshot's of the site's main page. Comment counts are as of May 2014. I excluded the six "Week's Top Stories," although three of them are included because they appeared elsewhere on the home page.

Swartz. This snapshot is from 21:03 on 12 January. It excludes any stories posted later, and any stories from before 11 January.

(378) False balance: Fox News demands a recount on US' warmest year
(265) Internet pioneer and information activist takes his own life
(168) Why Ultra HD won't be taking the world by storm
(139) RED has a 4K player for the coming wave of Ultra HD TVs
(127) Here's what you will look like in smart glasses
(84) Amazon AutoRip: How the labels held back progress for 14 years
(83) NASA rules out apocalypse in 2036, too
(81) Be productive: The Ultimate Smartphone Guide, part V
(80) But can it stream Crysis? Nvidia's new cloud gaming server explained
(71) How big does my project have to be to unit test?
(68) Apple blacklists Java on OS X to prevent latest "critical" exploits
(62) OWC is readying a 2TB, 3.5-inch form-factor workstation SSD for 2013
(53) "Buffy vs Edward" remix is back online, but no fallout for Lionsgate
(38) Ericsson gives 2,500 patents to maker of ancient wireless browser
(37) Critical Java vulnerability made possible by earlier incomplete patch
(29) California proposition to monitor sex offenders online gets put on ice
(21) Apple still nabbing things from Xerox—this time, a CFO
(21) Time to cash out: The last of CES
(20) Ask Ars: Does Facebook auto-delete content after a certain period of time?
(16) Aaron Swartz and me, over a loosely intertwined decade
(5) Android Apps of CES entertain your kids and enhance your conversations
(0) Leaving Las Vegas: 15 stories you might have missed this week

Snowden. This snapshot is from 14:35 on 9 June. The Internet Archive lacks any snapshots from later that day, or from 10 or 11 June. The home page snapshot from 12 June contains many following stories, but the original one is not listed, even though stories from as early as 8 June are present. Rather than focus on the fall-out of the story, I chose 9 June to give a more representative context. As 9 June was a Sunday, there are no stories from that day. I included stories from the previous two days instead. Though not present in the snapshot, I have included the Snowden story.

(459) Whistleblower who exposed NSA mass-surveillance revealed by The Guardian
(362) Obama defends digital spying: "I think we've struck the right balance"
(270) After burglaries, mystery car unlocking device has police stumped
(229) Xbox One, discs, and downloads: Better than feared, worse than hoped
(197) Google may not like it, but facial recognition is coming soon to Glass
(137) *Guardian* publishes *third* secret NSA document, on cyberwar
(107) Prenda Law's appeal bond rises to $237,584
(98) Chinese supercomputer destroys speed record and will get much faster
(93) Pirate Bay suggests Prenda did create "honeypot" for downloaders
(88) Behold, the world's most sophisticated Android trojan
(82) Is it a good idea to impose uniform code format for all developers?
(80) What we expect (and what we want) from Apple's WWDC keynote
(79) Bayes' theorem: Its triumphs and discontents
(67) Call connected through the NSA: Ars readers react
(62) Out with the old: the best Ultrabooks and convertibles from Computex
(43) Motorola and Cisco pay TiVo $490M to dodge East Texas patent trial
(42) Atomic bomb tests confirm formation of new brain cells
(41) Liveblog: Sony's pre-E3 2013 press conference
(37) Liveblog: Microsoft's pre-E3 2013 press conference
(35) Weird Science figures out why it lost its penis
(27) Under draft bill, EU wants to raise jail time for hackers, botnet operators
(25) Mad Catz will announce an Android-powered gaming console at E3
(23) Liveblog: Apple's WWDC 2013 keynote
(21) Airbus smart baggage tracks itself on plane trips
(19) Passengers can challenge gov't GPS tracking, court finds
(12) Liveblog: EA's pre-E3 2013 press conference
(12) Ars Technicast, Episode 28: WWDC rumors and what's missing from them
(7) Liveblog: Ubisoft's pre-E3 2013 press conference

CNN

The *CNN* article lists are taken from the latest (i.e. closest before midnight) Internet Archive snapshot of the *CNN* home page from that day. Videos without comment discussions are excluded. The *CNN* home page lists a very large number of stories, with recent and top stories at the top, along with collections of multiple stories about a special topic (e.g. gun violence, the NSA leaks, visiting the Democratic Republic of Congo on the days I examined). Beneath these are top stories broken down into specific categories: U.S., World, Politics, Tech, Business, etc. I excluded these topic-specific story lists from my data.

Swartz. The articles list is for 12 January 2013.

(16,759) Gun rights groups say Georgia home invasion proves their point
(7,108) Gabby and Mark: The new 'Bradys' of gun control
(4,609) Hobby Lobby finds way around $1.3-million-a-day Obamacare hit - for now
(4,493) Producer of popular gun-related videos found fatally shot
(4,267) Teacher talks armed student into giving up, police say
(3,271) Why Philly Will Never Be Newtown
(2,989) Internet Prodigy, 26, Commits Suicide
(1,733) Police: TV host Jimmy Savile exploited fame to abuse children on vast scale
(1,594) Judge agrees to delay plea for theater shooting suspect
(1,498) Had a flu shot? You'll be OK – maybe
(1,438) Florida tackling python problem with hunting contest
(1,272) Duchess of Cambridge's first official portrait unveiled
(1,091) 'American Taliban' wins right to group prayer in prison
(989) French bid to rescue hostage in Somalia fails, leading to soldier's death
(972) America flunks its health exam
(686) Security Clearance: Epic combat valor: Former soldier to receive Medal of Honor
(440) Flu shot myths addressed
(298) Flowers Foods to buy Wonder from Hostess
(295) Kobe Bryant, wife say they are calling off divorce
(229) 'Smart power': Army making cultural training a priority
(214) bleacher report: Ravens vs. Broncos: Live Score, Analysis for AFC Divisional Matchup
(181) political ticker: White House responds to Death Star petition: No
(176) Jack Lew: The signature that goes 'boing'!
(153) Britney Spears, Jason Trawick call off engagement
(149) Body of poisoned Illinois lotto winner to be exhumed
(149) 2013 Oscar nominations: Who got snubbed?
(118) Gunmen open fire on Italian consul's car in Benghazi
(117) eatocracy: You can trust a skinny chef! How the pros stay slim
(106) For many, a sense of purpose makes job stress worthwhile
(94) Surrealist photographer recreates his dreams in real life
(74) How the holidays spread the flu
(73) political ticker: Treasury Department rules out $1 trillion coin
(66) Amanpour: Karzai confident he can get U.S. troops immunity
(36) Anderson Cooper 360: Exposing Newtown conspiracy theory
(27) bleacher report: Duke vs. North Carolina State: Twitter Reaction, Postgame Recap and Analysis
(14) Anderson Cooper 360: Conspiracy theories in the wake of tragedy

Snowden. The article list is for 12 January 2013.

(13,974) NSA leaker comes forward, warns of agency's "existential threat"
(9,674) Santa Monica shooting victim dies, bringing toll to 5
(5,955) Arizona police: 4-year-old boy fatally shoots his father
(3,004) Texas actress first accuses husband, but she's arrested in ricin case
(2,750) Six months since Sandy Hook: Newtown residents find their voice
(2,666) Three questions for Clarence Thomas
(2,082) Data mining revelation opens political Pandora's box
(1,879) Internet laughs about being spied on

(1,008) Were the Wright brothers really first? Photo sparks flight fight
(918) Police: Missing Louisiana teacher's car found with body inside
(864) Suspect in deadly building collapse denied bail
(853) Privacy is not dead
(792) Report: Iran takes key step in nuclear reactor construction
(605) Man dies, 5 others rescued in extreme heat near Hoover Dam
(585) Holocaust artifacts bear witness
(558) Judge wants to know more about Paris Jackson incident
(550) Cowell gets egged during 'Britain's Got Talent' finale
(500) Mexico: As dangerous – and safe – as ever
(362) With no update on Mandela's condition, friend urges: 'It's time to let him go'
(361) Turkey's Erdogan to protesters: 'Even patience has an end'
(355) Mother who lost one son fights for other to get new lungs
(342) #Avgeeks: The new warriors on terror
(306) Thank you, Michael Douglas
(276) What to expect at Apple's WWDC
(273) Concerns about surveillance 'fanciful,' British official says
(217) Disappearances at Mexico City bar linked to gang
(203) bleacher report: Chauncey Billups Named NBA Teammate of the Year
(166) Travel Photo of the Day
(148) Report: North, South Korea hold talks
(138) Apparently This Matters: The worst time of day for work
(98) At 75, Judy Blume draws crowds with first film adaptation
(85) eatocracy: Into the heart of Congo
(77) Urban surfing: From Munich to China, daredevils ride inner-city rivers
(72) 10 things to know before visiting Democratic Republic of Congo
(69) Shirtless man with flare interrupts French Open final
(66) World's 50 best beach bars
(64) Record eighth French Open title for Nadal
(62) George Stroumboulopoulos: 'I like having good conversations with people'
(6) Mickelson comes up just short as English keeps nerve
(1) bleacher report: Pocono 400 2013 Results: Reaction, Leaders and Post-Race Analysis

The Guardian

The *Guardian* home page for 9 June also had 19 stories for which comments were not enabled.

(6,194) Edward Snowden: the whistleblower behind the NSA surveillance revelations
(2,935) NSA Prism program taps in to user data of Apple, Google and others
(2,511) NSA collecting phone records of millions of Verizon customers daily
(1,398) Boundless Informant: the NSA's secret tool to track global surveillance data
(1,259) Obama orders US to draw up overseas target list for cyber-attacks
(558) Erdoğan accuses EU members of hypocrisy over Turkey protests
(507) Turkey's protesters proclaimed as true heirs of nation's founding father
(415) Edward Snowden: more conscientious objector than common thief
(397) Climate change is happening but we can meet the challenge
(387) Recep Tayyip Erdoğan dismisses Turkey protesters as vandals

(368) There's a right way to deal with hecklers. Then there's Michelle Obama's…
(334) Edward Snowden identifies himself as source of NSA leaks - as it happened
(326) John Oliver: a very British coup
(297) Obama deflects criticism over NSA surveillance as Democrats sound alarm
(279) The 10 best British actors on American TV – in pictures
(274) The US knows force-feeding hunger strikers at Guantánamo is illegal
(264) Is liberalism dead? - video
(232) Iain Banks dies aged 59
(221) Guantánamo Bay hunger strike: quarter of inmates now being force-fed
(191) Should Google serve the state – or serve its customers?
(171) E3 2013: the biggest show on earth welcomes the next generation
(170) Edward Snowden's choice of Hong Kong as haven is a high-stakes gamble
(169) US surveillance has 'expanded' under Obama, says Bush's NSA director
(165) Rafael Nadal downs David Ferrer to win record eighth French Open crown
(158) In Lebanon, bravado about Syrian civil war is replaced by foreboding
(147) Nasa's Opportunity rover finds Martian water appropriate for the origin of life
(139) Fifth victim dies of wounds following Santa Monica shooting spree
(108) Champions Trophy 2013: New Zealand beat Sri Lanka by one wicket
(99) Gillian Anderson: The Fall girl who never bowed to Hollywood demands
(95) Nelson Mandela back in hospital with lung infection
(92) Ethan Hawke: playing all the angles
(92) NSA surveillance: lawmakers urge disclosure as Obama 'welcomes' debate
(75) James Franco: why I recreated Psycho
(69) Man Of Steel: eye to eye with Michael Shannon, the new General Zod
(67) Santa Monica shooting spree touched off by domestic violence, police say
(61) The National Security Agency: surveillance giant with eyes on America
(53) Barack Obama and Xi Jinping meet as cyber-scandals swirl
(48) NSA's Prism surveillance program: how it works and what it can do
(44) McCain and Feinstein pledge to close Guantánamo but sidestep hunger strike
(38) Bit-part actor charged over plot to frame husband for ricin letters
(37) China v the US: how the superpowers compare
(36) UN drone investigator expecting 'dramatic' decrease in US strikes
(34) Karzai demands return of all Afghans held prisoner by the UK in Helmand
(32) Meet the woman who risked jail in Egypt to fight against NGOs crackdown
(32) US assisting Middle Eastern allies against cyber threats from Iran
(27) US army suspends two-star general for allegedly failing to investigate assault
(26) US senators McCain and Feinstein visit Guantánamo amid prison hunger strike
(18) Cicadageddon part three: Staten Island swarmed by emerging insects
(18) Fragile peace in San Salvador as youth gangs trade weapons for jobs and hope
(12) Boston Bruins sweep past Pittsburgh Penguins and into Stanley Cup finals
(9) Ian Madigan kicks Ireland to victory over US Eagles in Texas
(7) Kane takes Blackhawks to Stanley Cup finals with double OT win over Kings
(6) Obama keen to hit the links but expect Xi to stay in the clubhouse
(2) 2013 Tony awards: Broadway kicks off its Kinky Boots and celebrates the show
(2) Margaret Atwood: doyenne of digital-savvy authors
(1) The Norm Chronicles by Michael Blastland and David Spiegelhalter – digested read

The New York Times

The *New York Times* maintains an online archive of news articles. I did a search for the relevant date for the category of "articles," which includes wedding and death announcements but excludes blog posts. I have excluded wedding and death announcements from my article counts. The lists below include only articles with associated comment discussions.

Swartz. The *New York Times* archive for January 12 contains 127 news articles (including wedding and death announcements). The *Times* editors enabled commenting on 11 of these.

(683) Obama Will Seek Citizenship Path in One Fast Push
(444) Aaron Swartz, Internet Activist, Dies at 26
(344) Hawks on Iraq Prepare for War Again, Against Hagel
(314) We Offer More Than Ankles, Gentlemen (op-ed about Obama and women)
(276) Darwin Was Wrong About Dating
(258) Over 50, and Under No Illusions
(245) Misguided Social Security "Reform"
(176) Gifted, Talented and Separated
(156) Democrats Behaving Badly (D senator Harry Reid)
(121) The Obama Synthesis (op-ed about Obama foreign policy)
(98) Collaborate vs. Collaborate

Snowden. The archive for June 9 contains 131 articles, plus 56 wedding and death announcements. The *Times* editors enabled commenting on 10 of these.

(2,182) Edward Snowden, ex-C.I.A Worker, Says He Disclosed U.S. Surveillance
(730) The Big Shrug (Krugman)
(703) Grouping Students by Ability Regains Favor in Classroom
(405) Senate Digs In for Long Battle Over Immigration Bill
(389) What's Next for Social Security?
(347) Affirmative Reaction
(65) When Artworks Crash: Restorers Face Digital Test
(31) Japanese Suitor for Rodriguez Got No Response
(3) Skewered Again, That Erotic Page Turner
(1) In Soccer's U.S. Open Cup, a Twisted Loan Deal

Slashdot

Slashdot maintains an archive of news stories by date, listing article titles and comment counts.

Swartz. The article list is for 12 January 2013.

(589) Aaron Swartz Commits Suicide
(372) Ask Slashdot: How To Stay Fit In the Office?
(244) Should Microsoft Switch To WebKit?
(226) The Android Lag Fix That Really Wasn't
(215) Amazon AutoRip — 14 Years Late

(210) Australian Spy Agency Seeks Permission To Hack Third-Party Computers
(168) Geothermal Power Advances
(161) Norway Tax Auditors Want To Open Source Cash Registers To Combat Fraud
(121) Smartphones: Life's Remote Control
(118) Who Controls Vert.x: Red Hat, VMware, Neither?
(106) Astronomers Discover a Group of Quasars 4 Billion Light Years Across
(191) This Isn't the Petition Response You're Looking For
(93) Vietnam Admits Deploying Bloggers
(76) Alleged ZeuS Botmaster Arrested For Stealing $100M From US Banks
(67) DARPA Wants To Seed the Ocean With Delayed-Action Robot Pods

Snowden. The article list is for 9 June 2013.

(860) NSA WhistleBlower Outs Himself
(480) Ask Slashdot: What To Do When Another Dev Steals Your Work and Adds Their Name?
(476) Scientists Explain Why Chairman of House Committee On Science Is Wrong
(404) British Foreign Secretary on Surveillance Worries: "Law Abiding Citizens Have Nothing To Fear"
(385) NSA Surveillance Heat Map: NSA Lied To Congress
(334) Linus Torvalds Promises Profanity Over Linux 3.10-rc5
(297) Steubenville Hacker Faces Longer Prison Sentence Than the Rapists
(266) Decommissioning San Onofre Nuclear Plant May Take Decades
(206) Chemists Build App That Could Identify Cheap Replacements For Luxury Wines
(141) Iain Banks Dies of Cancer At 59
(126) Supermarkets: High-Tech Hotbeds
(96) Asteroid Passes (Just) 65,000 Miles From Earth
(51) "Anti-Gravity" 3D Printer Sculpts Shapes On Any Surface
(42) Phenomenon Discovered In Ultracold Atoms Brings Us a Step Closer To Atomtronics
(40) NASA's "Opportunity" Rover Finds New Evidence For Once-Habitable Mars

The Volokh Conspiracy

The list is of articles from 13 to 16 June 2013. Starting with the second *Volokh* article (on 16 January), I worked backward collecting comment counts for stories up to and including the day before the first *Volokh* article (which was on 14 January). Two articles did not allow comments; I have included them in the list with a blank comment count.

(807) The Criminal Charges Against Aaron Swartz (Part 2: Prosecutorial Discretion)
(649) The Criminal Charges Against Aaron Swartz (Part 1: The Law)
(329) What Can Be Done about the Debt Ceiling?
(320) Treasury Rejects Platinum Coin Ploy
(200) Here Comes "Mandate Plus"
(183) Activities for Gun Appreciation Day
(165) Ronald Reagan's AR-15
(113) Who Needs a Budget?
(103) 15-Year Prison Sentence in Egypt for Converting from Islam

(93) Does Congress have the Power to Enforce Treaties? Part I
(86) There Is No Textual Foundation For The Claim That Treaties Can Increase The Power of Congress
(68) Does Congress Have the Power to Enforce Treaties? Part II
(63) The Supreme Court and Partisanship
(59) Judge Blocks Plaintiff's Attempt to Sell Indigent Defendant's Appeal Rights
(51) Gun Crime Against 12-to-17-Year-Olds Supposedly Fell 95% from 1994 to 2010
(37) Introducing Guest-Blogger Prof. Rick Pildes of NYU, to Debate Whether a Treaty Can Increase the Legislative Power of Congress
(31) Utopian for Beginners
(31) Moot Courts for Faculty Members Who Are About to Have Oral Arguments
(28) Koontz Oral Argument
(26) Justice Sotomayor on 60 Minutes
(18) Different Word Processors, Different Word Counts
(9) Freedom in The Bruce
(9) Congratulations to the Balkinization Folks on the Blog's 10th Anniversary!
(8) Treaties, the Law of Nations, and Foreign Commerce
(8) Koontz v. St. Johns River Water Management District
(6) The Effect of Privatization on the Public and Private Prison Lobbies — part 3
(4) The Framers Gave Congress a Robust List of Powers; They Did Not Provide That These Legislative Powers Can Be Increased By Treaty
(3) *Chevron* Revisited in *City of Arlington v. FCC*
(3) Free Online Constitutional Law Treatise, from the Library of Congress
(0) Talk on Kelo and Post-Kelo Eminent Domain Reform at South Texas College of Law
(0) Deadline for Georgetown Center for the Constitution Fellowship is February 1st.
(—) Law and Robots Conference Call for Papers, and a Link to a Video From HRW's Tom Malinowski Which, Though I am Not Persuaded, Will Always Treasure
(—) An "Ethical Turing Test"? More on Comparing Self-Driving Vehicles and Autonomous Weapon Systems

Appendix B: Data Collection

Although I wrote my own analysis software, I collected comment data manually. Here I explain the reasons for that choice, and how I went about data collection.

HTML pages are easy to scrape, crawl or save directly from a web browser. This is the case for comments on *Ars Technica* and *Slashdot*, where comments can be displayed as simple HTML pages. Saving these was a matter of making sure all comments were present (setting the rating threshold to display all on *Slashdot*, and logging in on *Ars* so that comments hidden due to down-voting would be visible), then using the browser's Save Page facility to save each page of comments as an HTML file.

Unlike static HTML, dynamic content generated by Javascript presents problems. The other sites in my study all made use of Javascript to load comments. They present only a few comments or threads by default; the reader must then click links or buttons to load more. (Some systems show no comments at all until the reader scrolls down to that part of the page.) When these links are clicked, Javascript on the page manipulates the DOM (Document Object Model): it sends requests to the server and inserts HTML containing additional comment text.

Unfortunately, the browser Save facility only stores the static HTML file fetched from the web server when the page is first viewed, excluding any comments loaded dynamically since then. Saving the complete page, with images and scripts, is of no help either; it only saves statically-linked content, not dynamically-generated content. I tried other browsers, but they do the same as Firefox. To obtain all the comments, I saw two main possibilities.

The first is to request the comment data directly using the comment API (in many cases is hosted by a third party service provider like Disqus). This is likely to entail submitting a unique session ID and (for third-party services) an API key along with requests to the API. Furthermore, the service may reject requests originating from any domain other than the one hosting the comments. The requests must therefore be made by Javascript from *within* the page containing the contents, using ID values that are valid only for that particular session. This entails a) obtaining the keys, then passing them on to a separate application (a tedious manual process), b) using a scraper that simulates a browser session (not just an API interaction), thereby obtaining its own keys, or c) modifying the page containing the comments within the browser. This last option is probably most practical, but it is complex and must be done carefully to avoid exposing the researcher's computer to security risks. While this more or less amounts to automating requests the user could make manually by clicking on “more” buttons etc., it must be done in a way that does not cause trouble for the news site or service provider. (Corrupting comment data should not be possible, as the add-on would be running with the same API permissions as the comment page itself—otherwise it would be vulnerable to attack—but high-bandwidth requests could place a load on the server.)

Given the open-ended complexity of this approach, I opted for the second possibility: to save the page DOM *after* all the comments have been loaded. This entails manually clicking on all the “more” buttons until all comments are visible (this took over ten minutes for the CNN articles), then taking a snapshot of the DOM. (Automating the process of clicking would entail many of the same difficulties as calling the comment

API.) To do this, I used the Save DOM capability of the DOM Inspector add-on for Firefox. Compared to the API approach, this is slow and error-prone but is also straightforward and guaranteed to work: what you see is what you get.

A third option would be to contact the news organization or third-party comment service and get the data directly from them. In the past, *Slashdot* has provided researchers with data. I never even considered asking. Assuming I could even get them to agree, I wanted to do the research on my terms.

I should note that different users may see different views of the comment data. I found this on *Ars Technica*, where down-voted comments are only visible to logged-in users. It may be that other systems also present different views to different people.

Appendix C: Data Parsing & Analysis Software

Here I describe the software I wrote to help me parse and analyze my data. The software is a web application based on the Django framework for Python. I wrote the software in tandem with my research, adding features and fixing bugs as necessary.

Data Parsing & Normalization

To view and manipulate the comments in a consistent way, I had to convert them to a standard format, stripping out the various bits of scripting and styling present on each site.

As the comment HTML format on each site is different, I had to write a different parsing script for each. This extracted comment data and metadata from the saved HTML, generating two outputs: a new normalized HTML file, and corresponding database records for the discussion and for each comment within it. I could have dispensed with the HTML file and stored all comment data in the database and generated the HTML on the fly. This might have been preferable, but when I began with over 600 *Slashdot* discussions I felt that the HTML provided a reliable archive format which I could work with on its own, and from which I could reconstruct the database if something went wrong. (Of course I also kept the original saved DOM data.)

Every site provided different metadata. For example: some had down votes, some did not; *Slashdot* had comment titles, but no other site did; and so forth. The metadata fields I stored for each comment, if applicable, included:

- title (only *Slashdot* comments feature a title)
- original comment ID (most sites assign each comment a unique ID, useful for debugging)
- posting time
- author name
- author user ID (most sites assign users unique IDs)
- author description (e.g. on *Ars*)
- up-votes
- down-votes
- aggregate rating (usually up-votes minus down)
- a flag indicating that this comment was hidden by moderators
- a flag indicating whether this comment was deleted by moderators
- a flag indicating that this comment was a reader favorite
- a flag indicating that this comment was an editor pick
- pointers into the HTML file indicating the location of the comment

The database also stored several derived fields (e.g. number of replies, an ordinal indicating posting order). Additional fields were present in the normalized HTML file, but I did not create columns in the database for them (e.g. *Slashdot*'s rating descriptor or comment modification times on *Ars*). I could see these while perusing comments, but my analysis software could not filter based on them.

One of the most important tasks provided by the parser was data validation. In some cases, comments had fields missing that seemed to be essential; these pointed to misunderstandings I had about the comment format. More importantly, the parser checked that the number of comments found in the file matched the count reported by the site. This ensured that I actually did get all the comments (or let me know that some were missing). I caught several programming and data collection errors this way (including a bug on the *Slashdot* site, and the need to log-in to *Ars*).

Data Filtering and Exploration

The main analysis application was built on top of a note-taking application I wrote years ago for keeping track of research notes and documents. This software already allowed me to create and search for snippets of text, and included a highlighting and annotation mechanism that I adapted for coding my data (I developed Marginalia, the annotation technology, for use in education; see http://webmarginalia.net/).

The software displays each comment discussion as a single HTML document, beginning with the primary article followed by comments sorted in chronological order within each thread (oldest to newest). Filtering mechanisms based on *Slashdot*'s interface alter which comments are visible at time: each comment can be shown in full, collapsed down to a single line, or hidden altogether. Expanding or collapsing comments can be done individually, by clicking on a comment's title, or in bulk, by typing in queries that select comments matching specified criteria.

Running along side the document is a margin where notes can be added. A note is created by highlighting a passage of text, then clicking in the margin and typing a margin note. This combination of quote (highlighted passage) and note stays with the document, but can also be searched for and listed independently of the document itself. Each document can have multiple independent margins, of which only one is shown at a time. These can be used to reduce clutter (keeping different kinds of annotations in different margins) or for trying out different coding schemes.

Coding is done by entering Twitter-style hashtags in margin notes. The software tracks these hashtags, associating them both with the note of which they are part and with the annotated comment. Hashtags are hierarchical, using a dot notation. For example, #if is my code for Institutional Failure; it is broken down into two sub-tags: #if.s for strong instances, and #if.w for weak ones. So far as the software is concerned, every instance of #if.s or #if.w is also an instance of #if. This is useful when querying; when the software tabulates code counts, it rolls up sub-codes into their parents. There is no limit to code hierarchy (there could be #if.w.a and #if.w.b, for instance.)

The key to filtering the data is a query language. At the top of each discussion is a text box into which the researcher can type a query, similar to performing a Google search. Comments matching the query are shown in full, while those not matching the query are collapsed or hidden. Comments can still be expanded and collapsed manually, so it is possible to search for certain types of comments, then explore others nearby (e.g. parents and replies) by clicking to expand them.

The Query Language

At first, the query language was for full text searches for notes and quotes. The following query matches quotes or notes containing the words democratic or republican:

```
qts "democrat", "republican"
```

When analyzing comments, I found I usually queried for tags. The following query finds comments coded #if or #uj ("chs" means chapters, i.e. comments):

```
chs #if, #uj
```

This queries for comments coded #if but not #uj:

```
chs #if (not #uj)
```

This finds comments written by a user named "geof":

```
chs author:geof
```

Comments with between 10 and 1000 votes up:

```
chs votesup:10..1000
```

There is a function for taking a random sample:

```
chs sample n=10 seed=1
```

This selects a pseudorandom sample of ten comments. So long as the same seed value is specified, it will always return the same sample. The seed is salted with the discussion URL, so that the same query on a different discussion will return a different sample. At several points I saved queries like this in my notes so that I could repeat them if I needed to go back and look at the same subset of data. The sample is stable as n changes. The following query retrieves the same ten comments, but adds five more at random:

```
chs sample n=15 seed=1
```

Of course these terms can be combined. The following query takes the top hundred replies by votes up, then selects ten of them at random:

```
chs sample n=10 seed=1 top n=100 field=votesup
isreply:yes
```

The query language was initially inspired by Google searches for ease of use, but ended up approaching a LISP-like functional language. Were the above query written with LISP-style parentheses, it would look like this:

```
(chs (sample n=10 seed=1 (top n=100 field=votesup
isreply:yes)))
```

Queries can also specify documents (discussions, “docs”) and individual annotations (“qts”). For example, one could search for all discussions where the #if code has been used in an annotation somewhere:

```
docs qts #if
```

Documents can also be tagged. This fetches my Aaron Swartz discussions:

```
docs #aaronsw
```

While this finds articles and discussions from *Slashdot*:

```
docs author:Slashdot
```

This finds all annotations on comments coded Institutional Failure in *Slashdot* discussions about Aaron Swartz (i.e., it includes annotations that do not include #if so long as there is an #if annotation somewhere on the same comment):

```
qts chs #if docs #aaronsw author:Slashdot
```

I found it useful to look at *Slashdot* comments with ratings of 4 or 5, and their parents and ancestor comments regardless of rating:

```
chs ancestors score:4..5
```

Queries can be restricted to notes in a particular named margin. Here I find comments coded #m (male), but only in the margin for coding comment author sex:

```
chs #m qts sheet:sex
```

All of this helped me to slice and dice the data as I explored it. For example, when counting male and female user names in the *New York Times* discussion, I created a new margin for this purpose. I took a random sample of comments and annotated each identifiable name with #m or #f. I then queried for all coded comments to see how many I had. If it was too few, I grew the sample until I had enough identifiably male or female names in each. Finally, I looked at the number of comments tagged with each code to see the sex distribution of names.

I also used queries during qualitative analysis. For example, I frequently used a “q” sub-code when I was in doubt about whether a code should apply, e.g. coding #if.q. I would then go back and query for chs #if.q to review borderline cases.

When working up a description of a code like Institutional Failure, I would query for annotations coded #if, in all documents. I could then read through all annotated passages to ensure my definition was consistent and that the coding was correct. If I needed to look at the context, the application provided a link back to the whole discussion, with the browser scrolled to that particular annotation. Similarly, I reviewed the results of queries like these when I wanted to find good examples during my write-up.

Statistics & Tables

The software can report some basic statistical information about discussions, including graphs for the following (along with measures like mean, median, and percentile distributions):

- distribution of up-votes, down-votes, aggregate ratings, and total votes (up plus down)
- distribution of comments per user (and top 20 most prolific users)
- distribution of reply counts
- distribution of comment word counts (added later, so only calculated for some discussions)
- distribution of posting times
- scatter plot of ratings vs word counts
- scatter plot of ratings vs posting time

These can be generated for a whole discussion, or for a subset selected by running a query.

I also implemented summary tables for codes occurring in documents, for how where pairs of codes co-occur, for usernames common to multiple discussions, and so forth. Again, queries are used to select the data for tabulation.